Hermann Flaschka

Principles of Analysis

University of Arizona
Math 527
© 1995

Contents

Index	i ÷ vi
Chapter 1. Distance and size	I-1
Section 1. Metric spaces	I-2
Subsection 1.1 Examples of metric spaces	I-4
Subsection 1.2 Equivalence of metrics	I-21
Subsection 1.3 Hölder and Minkowsky inequalities	I-30
Subsection 1.4 Lebesgue spaces	I-43
Section 2. Normed linear spaces	I-56
Subsection 2.1 Linear spaces	I-57
Subsection 2.2 Norms	I-60
Subsection 2.3 Application: norms of matrices	I-63
Subsection 2.4 Application: differential equations	I-84
Section 3. Topology	I-106
Subsection 3.1 Metric space topology	I-109
Subsection 3.2 Continuity	I-119
Subsection 3.3 Hausdorff spaces	I-131
Subsection 3.4 How to define topologies	I-136
Subsection 3.5 Vocabulary of topology	I-144
Subsection 3.6 Application: brownian motion	I-151
Subsection 3.7 Application: weak topology	I-157
Section 4. Application: theory of Disttributions	I-171
Subsection 4.1 Examples of distributions	I-177
Subsection 4.2 Test function spaces	I-198
Subsection 4.3 Generalized functions	I-217
Section 5. The size of sets	I-227
Subsection 5.1 Small sets, large sets	I-228
Subsection 5.2 Measure theory	I-239
Subsection 5.3 Application: Hausdorff measure	I-252
Subsection 5.4 Application: Brownian motion	I-264

Chapter 2. Convergence and compactness II-281
Section 1. Convergence and compactness II-282
Subsection 1.1 Cauchy sequences II-284
Subsection 1.2 Completions II-299
Subsection 1.3 Compactness II-303
Subsection 1.4 Uniformity II-314
Section 2. The Lebesgue integral II-334
Subsection 2.1 Completion of $C([0,1])$ II-335
Subsection 2.2 On defining integrals II-338
Subsection 2.3 Measurable functions II-347
Subsection 2.4 The Lebesgue integral II-353
Subsection 2.5 Convergence theorems II-361
Subsection 2.6 The Lebesgue spaces II-373
Subsection 2.7 Fubini's theorem II-387
Section 3. Arzelà-Ascoli theorem II-394
Subsection 3.1 Solutions of ordinary differential equations II-398
Subsection 3.2 The length of curves II-403

absolutely convergent improper Riemann integral, 360.
accumulation point. 146. 148.
almost never, 270.
almost surely, 270.
anisotropic heat equation. 106.
arithmetic mean, 42.
Arzelá-Ascoli Theorem, (II.3) 396.

Baire category theorem, 229, 296.
Baire space, 230.
Banach space, 380.
base (for a topology), 138.
-s, equivalent, 139.
base three expansion. 114.
base two expansion. 171. 271. 272.
basis (of a linear space), 59.
algebraic –, 59.
beaded string, 85.
Bernoulli trials, 265, 272.
Bessel's inequality, 169.
Bolzano-Weierstrass Theorem, 306.
Borel-Cantelli Lemma. 268.
Borel measure, 246, 254.
– set, 246.
– sigma-algebra. 246, 267, 351, 381.
– zero-one law, 270.
boundary condition. 98, 99.
boundary point, 150.
bounded metric space. 303.
bounded subset of metric space, 304.
bounded variation (function), 347.
Brownian motion. 151. 249, 264. 350. 381.
bump function, 203. 219.

$C^\infty(\mathbf{R}^n)$, 199.
$C_0^\infty(\mathbf{R})'$, 217.
$C_0^\infty(\mathbf{R}^n)$. 201.
$C(M)$, 320.
$C(M)$ is complete. 297. 320.
$C([a.b])$. 44.

Cantor diagonal process, 16. 114. 397.
Cantor set. 113. 114, 149. 230. 232. 247. 260.
generalized –, 247
Cauchy-Schwarz inequality. 33. 41. 49.
Cauchy Principal Value, 180. 211. 214.
Cauchy sequence, 284, 375.
Cauchy test for series, 20.
Cauchy's formula (complex analysis), 187.
chaos. 18.
Chapman-Kolmogorov-Smoluchowski equation, 277.
characteristic function, 335. 353.
Chebyshev's Inequality, 378, 385, 386.
closed set, 144.
closure, 145.
point of –, 146.
coarser metric, 30.
– topology, 158.
cocountable topology, 124.
coin toss. 18, 140, 171, 248, 265. 341. 386.
compact, 201.
– metric space, 304.
– set, 304. 308.
– subset. 309.
– support. 201.
– topological space, 308.
comparison of L^1 and L^2 convergence, 51.
comparison test for series, 370.
complete, 287.
– measure space, 376.
– metric space, 287.
completeness of $C^1([0,1])$, 291, 293.
– of l^p, 290.
– of $L^p(X)$, 377.
completion. 302.
condition number, 65, 80.

conditional convergence, 356.
conjugate exponents, 32, 39.
connected, 128, 298.
continuity (on metric space), 120.
continuity (on topological space), 120, 127.
continuity and convergence, 125.
  – and sequences, 124.
  – in the mean, 381.
contour integration, 179.
contraction, 294.
Contraction Mapping Theorem, 294, 298.
convergence, conditional, 356.
convergence in $C_0^\infty(\mathbf{R}^n)$, 212.
  – in measure, 372.
  – in metric space, 25.
  – of a sequence, 284.
  – of distributions, 223.
  – of functions in sense of distributions, 223.
  – of test functions, 210.
  pointwise –, 318.
  uniform –, 318.
convolution, 333, 391.
  Fourier transform of, 393.
  L^p-estimate for a –, 392.
countable additivity, 250.
cylinder set, 267.

$\mathcal{D}$ (test function space), 201.
$\mathcal{D}'$ (distributions on $\mathcal{D}$), 223.
δ', 221, 224.
δ'', 225.
d'Alembert's formula, 95.
Dedekind cut, 299.
delta distribution on a sphere, see under *Dirac delta*.
delta measure, 251.
dense subset, 228.
die, 265.
dimension (of a linear space), 58.
Dirac delta, 211, 221, 369, 173, 174, 186, 191, 369.
  Fourier transform of –, 382.
  – on a sphere, (I.4.1.57) 192, 224.
  – on $\mathbf{R}^2$, 186.
Dirichlet function, 343, 348, 350, 353, 374.
discrete metric, 115, 149.
discriminant of a polynomial, 237.
dispersion, 164.
distribution, 175, 185, 217, 221.
  change of variables in, 195.
Dominated Convergence Theorem, 366.
dot product, 33.
drift, 275.

ϵ-ball, 109.
ϵ-net, 307, 313.
eigenvalues, 71, 74.
elliptic functions, 18.
energy norm, see under *norm*.
equicontinuous, 395.
equivalence class, 301, 336, 375.
equivalence of metrics, 22.
equivalence relation, 301.
essential supremum, 374.
Euclidean norm, 61.
Euler's constant, 177, 370.
Euler's method for ode's, 399.
evaluation functions, 152.
event, 248.
expectation, 384.
extended real numbers, 349.

F_σ set, 154.
$F_{\sigma\delta}$ set, 154.
Fatou's Lemma, 363.
finer topology, 158.
finite additivity, 250.
finite intersection property, 309.
first category set, 228.
first countable, 125, 134, 150.
Fourier series, 9, 94.
  – Transform, 368.
  – – of a convolution, 393.
Fresnel integral, 179.

Fubini's Theorem, 390.
Fundamental Theorem of Calculus, 103.

G_δ set, 154, 297, 337.
$G_{\delta\sigma}$ set, 154.
Gamma function, 177, 180, 262, 333, 370.
generalized Cantor set, 247.
generalized functions, (I.4.3) 220.
generic, 233.
geodesic, 403.
geometric mean, 42.
Green's formula, 100, 105.

Hamiltonian differential equation, 351.
harmonic oscillator, 96.
Hausdorff dimension, 258.
Hausdorff measure, 254.
Hausdorff outer measure, 254.
Hausdorff space, (I.3.3) 131, 311, 313, 314.
heat equation, 99, 171, 186, 196, 219, 221, 323, 333, 392.
  energy decay in the –, 101.
  anisotropic –, 106.
Heine-Borel Theorem, 306.
homeomorphism, 128.
Hölder's inequality, (I.1.3) 32, 48.
  for functions of several variables, 54.
  for three functions, 55.
  for vector functions, 49.

Improper integral, 330.
improper Riemann integral, 177, 179, 333, 356, 359, 360.
  – depending on a parameter, 330.
inclusion between l^p-spaces, 14, 48.
inclusion between L^p_0-spaces, 50.
independence (of events), 265.
inf, 83.
infinitely often (event), 268.
inhomogenous wave equation, 96.
initial condition, 98, 99.
initial/boundary value problem, 92.
integrable, 361.
integration by parts (multivariable), 103.
interchange of limit and integral, 328.
interior, 148, 228.
interior point, 148, 228.
Intermediate Value Theorem, 298.
inverse of a function, 126.
isometry, 303.

jump function, 346, 371.

kinetic energy, 88.
Kolmogorov extension theorem, 268, 278.
Korteweg - deVries equation, 164, 170.

$l^p(\mathbf{R},\mathbf{N})$, 9.
$l^p(\mathbf{R},\mathbf{Z})$, 10.
$L^p(X), 1 \le p \le \infty$, (II.2.6) 374.
Lagrange multipliers, 70.
Laplace Equation, 99.
  energy norm for, 101.
  operator, 97.
least upper bound, 81, 112.
Lebesgue integral, II.2.4, 345, 353, 372.
  translation invariance of –, 359.
  when – is undefined, 354, 357, 358, 359.
Lebesgue measure, 244.
  – zero, 231.
Lebesgue spaces, (I.1.4, II.2)
length of a curve, 404.
liminf, 362, 406.
liminf of a sequence of functions, 362.
limsup, 364.
linear function on $l^2(\mathbf{R},\mathbf{N})$, 158.
  – on $L^2_0([a,b])$, 162.
linear space, 57.

linearly dependent, 58.
linearly independent, 58.
Liouville's theorem , 352.
local base, 117, 138.
locally compact, 313, 381.
lower semicontinuous, 405.
L^p spaces, 45, 373.
lub, 81.

max, 81.
meager set, 228, 297.
measurable function, (II.2.3) 347, 349.
measurable set, 241.
measurable space, 241.
measure, 242.
measure space, 242.
  complete, 376.
  sigma finite, 389.
measure zero, 231.
measure-preserving transformation, 351.
metric, 3.
  Euclidean, 4.
metric space, (I.1) 3.
  bounded, 303.
  separable 150, 308.
  discrete, 3, 149.
metric topology, 110.
metrics, equivalent, (I.1.2) 22, 116.
  – on $\mathbf{R}^n$, 26.
  – on sequence spaces, 29.
  – on L_0^p, 50.
min, 81.
Minkowski's Inequality, (I.3.1) 31. 32, 48, 49, 54, 377.
  for vector functions, 49.
modulus of continuity, 317.
Monotone Convergence Theorem, 364.

natural matrix norm, 67.
neighborhood, 144.
Neumann boundary value problem, 105.
Newton-Raphson method, 286, 295.
nondifferentiable functions, 278, 324-328.
norm, (I.2.2) 32, 60.
  energy –, for o.d.e.'s 90.
  – –, for p.d.e.'s, 101, 104, 106.
  – on $l^p(\mathbf{R}, \mathbf{N})$, 61.
  – on $L_0^p([a, b])$, 62.
  – on $L^p(X)$, 374, 377.
  – on Sobolev space, 62.
  natural matrix, (I.2.3) 40 (Ex. 1.3.33), 67.
normal curve, 274.
  – probability measure, 251.
normed linear space, 60.
Nova, 260.
nowhere dense set, 228.

one-point compactification of $\mathbf{R}$, 129, 314.
open covering, 308.
open set, 108.

$\mathrm{PV}\frac{1}{x}$, 186, 214, 216.
$\mathrm{PV}\frac{1}{x^2}$, 186.
$\mathrm{PV}\frac{1}{x^3}$, 186.
$p_{m,k}(\varphi)$ (seminorm on C_0^∞), 208.
parametrized curve, 403.
partition, 338.
piecewise continuous, 44.
Plemelj's formula, 187.
point of closure, 146.
point vortex, 196.
pointwise convergence, 318.
potential energy, 88.
probability measure, 250.
product measure, 389.
product topology, 143, 314.
projective space, 142.

$q_{m,k}(\varphi)$ (seminorm on Schwartz space), 216.

Rademacher function, 171.
random variable, 348, 384.
rationals are countable, 16.

regular Borel measure, 381.
regularization, 177.
relative error, 64.
relative topology, 309.
residual set, 228, 326.
Riemann integrable, 339.
  limits of sequences may not be –, 350.
Riemann integral, 316, 338, 339, 355.
  lower and upper, 339.
Riemann zeta function, 321.
Riemann-Lebesgue lemma, 166, 382, 383.
Riesz-Fischer Theorem, 377.
root mean square, 48.

$\mathcal{S}(\mathbf{R}^n)$, 200.
saltus function, 346.
sample path, 153.
sample point, 248.
sample space, 248.
Sard's Theorem, 234.
Schwartz space, 200.
second category set, 228.
second countable, 150.
semialgebra, 268.
seminorm, 209.
separable, 395.
separable metric space, 150, 308.
sequence in a topological space, 123.
sequence spaces, 8.
series, double, (I.1.1.39) 17.
sigma algebra, 241.
sigma finite measure space, 389.
sigma-algebra, 348.
  – generated by a collection of sets, 247.
simple function, 353.
smooth, 201.
Sobolev Space, (I.1.4.23) 52.
  discrete –, 15.
space of distributions, 223.
sphere, volume of, (I.5.3.25) 261.
spherical polar coordinates, 54.
stable processes, 277.
standard deviation, 385.
stationary fluid flow, 196.
strong convergence, 161, 221.
subbase, 138.
subsequence, 303.
sup, 81.
support, 201.

tempered distribution, 223.
term-by-term differentiation, 323.
term-by-term integration, 328.
test function spaces, 185.
test function, 175, 181.
Tihonov's Theorem, 314.
Tonelli's Theorem, 390.
topological space, 107.
topology, 108.
  coarser –, 158.
  – defined by functions, 141.
  finer –, 158.
  – on $C^\infty(\mathbf{R})$, 118, 209.
  – on $C^\infty(\mathbf{R}^n)$, 210.
  – on $C_0^\infty(\mathbf{R}^n)$, 214.
  – on $\mathcal{S}(\mathbf{R}^n)$, 216.
transition layer, 207.
triangle inequality, 2, 4, 30, 31, 77.
  – for Sobolev spaces, 53.

uniform convergence, 318.
uniform continuity, 314.
upper bound, 81.
upper semicontinuous, 405.

variance, 385.
vector space, 57.
volume of ball, see under *sphere*.
vortex lines, 196.
vortex sheet, 198.

wave equation, 91, 92, 98, 196.
  anisotropic –, 103.
weak convergence, 161, 221, 223.

weak topology, (I.3.7) 139, 158.
162.
Hausdorff property and –. 170.
Weierstrass test (for series), 319,
321, 323.
wheel, 37.
Wiener measure. 272, 381.
work, 88.

$(x + i0)^{-1}$, (I.4.1.36) 186.
$(x - i0)^{-1}$, 191.

Young's inequality, 36, 41.

Zariski generic set, (II.5.1.31), 235.
Zariski topology, (I.3.3.13) 134.
149, 158, 235. 313, 314.

CHAPTER I

Distance and Size

To determine when two numbers $r, s \in \mathbf{R}$ are almost equal, we place a ruler along the ever-popular number line and ask: is the distance $|r - s|$ between those numbers fairly small? To gauge the accuracy of the numerical solution of a system of equations, we try to estimate the "size of the error"; the error, if the unknowns are $x_1, \ldots, x_n$, is a measure of the differences

$$x_1^* - x_1^{**}, \ldots, x_n^* - x_n^{**}$$

where $x_1^* \ldots, x_n^*$ is the true solution and $x_1^{**} \ldots x_n^{**}$ is the computed solution. If those equations are linear, we might ask: when are two systems defined by matrices $\mathbf{A}, \tilde{\mathbf{A}}$, e.g.,

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \quad \tilde{\mathbf{A}}\mathbf{x} = \mathbf{b},$$

nearly the same? When the matrices are "close to each other," you say, but how do you measure the distance between two matrices $\mathbf{A}$ and $\tilde{\mathbf{A}}$? It is not enough merely to invent a definition, like "the largest difference $|A_{ij} - \tilde{A}_{ij}|$", one must be able to draw *useful conclusions*: "if two matrices are sufficiently close to each other, then the solutions of the respective systems are also close." In a different vein, one sometimes needs to quantify the difference between a signal sent on a transmission line and the signal received: how big is the "loss of information"? That requires a definition of the "distance" between two functions,

$$F_{\text{sent}}(t) - F_{\text{received}}(t), \ -\infty < t < \infty.$$

These are but a few illustrations of a fundamental problem: the measurement of size and distance. This is a recurring theme in analysis. Without it, one cannot begin to speak about other, seemingly more relevant, topics such as continuity, approximation, or convergence.

This chapter deals with many different concepts of size and distance. No one definition can cover all situations and give useful information in every instance. It will be necessary to understand the types of definitions that are available, and to know which is likely to be useful in a certain setting.

In Section 1, *metric spaces* are introduced. They encode the essentials of the intuitive notion of *distance*, and allow one to measure differences between complicated objects such as functions, sequences, sets, and random events. When one considers objects that can be added and scaled (they are elements of a *vector*

space), there may be a way to quantify "size" as well as "distance." In Section 2, we discuss *norms*, which are a generalization of *absolute value*. Norms assign lengths to vectors — and those "vectors" may be fairly exotic things, such as functions, matrices, and the like. Section 3 deals with general topological spaces. The notion of distance is almost lost: one no longer knows the numerical value of the distance between two objects, one only knows —in some vague sense— whether or not the objects are close to each other. This caricature of the solid old concept of "distance" is exactly the right notion to help one to understand the space of random motions of a microscopic particle (*Brownian motion*). It is difficult to appreciate the need for such ideas if no applications are encountered. Therefore, in Section 4 we anticipate some of the material to be described in detail in later chapters, show how continuity and convergence fit into the picture, and introduce elements of the theory of distributions. In Section 5, the emphasis changes: we no longer ask about the size of individual objects, but about the size of whole sets. First, there is a rough distinction between big sets and small sets. Most numbers are not rational, most quadratic equations have distinct roots — these and similar statements are made precise by concepts such as *Baire category* and *measure zero sets*. The quantitative theory of the size of sets, *measure theory*, is one of the staples of modern analysis; a short summary is given, and the need for a theory is illustrated by a complicated example: the probabilistic theory of Brownian motion.

By the end of this chapter, it will be clear to the reader that the mere definition of a logically consistent concept of distance or size can be a complicated business. Later on, we will see what one does with those definitions.

Section 1 Metric spaces

The concept of *metric space* is an axiomatization of the distance $|r-s|$ between real numbers. It was invented by Maurice Fréchet in a 1906 paper [EDM II, p. 1014]. The following properties of $|r-s|$ are assumed to be familiar:

$$|r-s| \geq 0, \text{and} = 0 \quad \text{iff } r = s, \tag{1.0.1}$$

$$|r-s| = |s-r|, \tag{1.0.2}$$

$$|r-t| \leq |r-s| + |s-t|. \tag{1.0.3}$$

Requirement (1.0.1) is sometimes referred to as *nondegeneracy*, (1.0.2) is called *symmetry*, and (1.0.3) is the *triangle inequality*. That name derives from a property of vectors in the plane: the sum of the lengths of two sides of a triangle is never less than the length of the third side (see below).

It is true, but not evident, that those three properties capture all the familiar properties of "distance" on the number line. The difficult part of Fréchet's work was to see past the exceedingly rich structure of the real number line (the order relations $<$ and $>$, rationals, irrationals, length, *etc.*) to identify those abstract features that bear on the concept of "distance".

1.0.4 Definition. *A metric space is a pair* (M, d)*, where* M *is a set and* $d : M \times M \to [0, \infty)$*, called a metric on* M*, is a function satisfying*

$$d(x, y) = 0. \text{ iff } x = y \tag{1.0.5}$$

$$d(x, y) = d(y, x), \text{ for all } x, y \in M \tag{1.0.6}$$

$$d(x, z) \leq d(x, y) + d(y, z), \text{ for all } x, y, z \epsilon M. \blacksquare \tag{1.0.7}$$

1.0.8 Remark. The function $d(x, y)$, which is intended to resemble $|r - s|$, measures *the distance between two points of* M. It *does not follow* that a single point of M should itself have a "size". In other words, there is not yet a counterpart of the absolute value $|r|$ of a single number.

This is perhaps the most penetrating insight in the definition of metric space: one may be able to talk about distances between objects without having an intrinsic measure of the size of a single object. Often, one has both, but not always: that will be discussed later. ∎

• **1.0.9 Exercise.** Check that the real numbers **R** become a metric space in two different ways (at least):

$$\text{first, } d(r, s) \stackrel{\text{def}}{=} |r - s|$$

$$\text{or, } d(r, s) \stackrel{\text{def}}{=} 100 \times |r - s|.$$

The second definition amounts to changing the unit of measurement from meters to centimeters. This should not make a great difference, and in Subsection 1.2 we study the concept of *equivalence* of different metrics. ∎

• **1.0.10 Exercise.** Let M be a set. Define

$$d(x, y) = \begin{cases} 1, & \text{if } x \neq y, \\ 0, & \text{if } x = y. \end{cases}$$

Show that this is a metric on M. (It is called the *discrete* metric, because two distinct points always stay quite far from each other. The term *discrete*, however, does not refer to "propriety" or "keeping one's distance", it is the antonym of *continuous*). ∎

• **1.0.11 Exercise.** Let (M, d) be a metric space. Show that for all $x, y, z \in M$, one has $|d(x, y) - d(x, z)| \leq d(y, z)$. Draw a picture to illustrate the meaning of this inequality. (Hint: write the triangle inequality twice, with the symbols x, y, and z in different orders, and subtract one inequality from the other.) ∎

Subsection 1.1 Examples of Metric Spaces

We now look at several important metric spaces (M, d). In most cases, we will not prove that the examples satisfy the axioms required of metric spaces — the prototype of those arguments is left for Subsection 1.3.

1.1.1 Definition. *The Euclidean metric on* $\mathbf{R}^n$ *is*

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2} \tag{1.1.2}$$

where $\mathbf{x} = (x_1 \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n)$. ∎

This is surely the most familiar metric. In $\mathbf{R}^2$, for example, it gives as the distance between $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$:

$$\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}.$$

The triangle inequality,

$$d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$$

now has an obvious geometric interpretation:

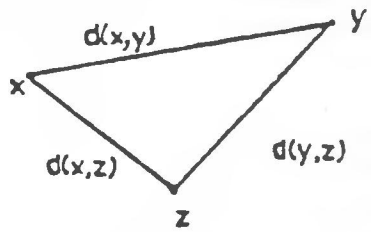


Figure 1. *The triangle inequality*

You may not find it particularly easy to *prove* the triangle inequality even in this simple setting. We give the proof later, after presenting many examples that point the way to a comprehensive setting.

It turns out that, for any real number $p \geq 1$, the following formula also defines a metric on $\mathbf{R}^n$:

$$d_p(\mathbf{x}, \mathbf{y}) = \sqrt[p]{\sum_{j=1}^{n} |x_j - y_j|^p}. \tag{1.1.3}$$

Obviously, $d_p(\mathbf{x}, \mathbf{y}) \geq 0$, and $d_p(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$, and equally obviously, $d_p(\mathbf{x}, \mathbf{y}) = d_p(\mathbf{y}, \mathbf{x})$. Not so obviously, $d_p(\mathbf{x}, \mathbf{z}) \leq d_p(\mathbf{x}, \mathbf{y}) + d_p(\mathbf{y}, \mathbf{z})$ — again, this is discussed later.

1.1.4 Definition. *For $p \geq 1$, define the l^p metric [1] on $\mathbf{R}^n$ to be*

$$d_p(\mathbf{x},\mathbf{y}) \overset{\text{notation}}{=} \|\mathbf{x}-\mathbf{y}\|_p \overset{\text{def}}{=} \sqrt[p]{\sum_{j=1}^{n} |x_j - y_j|^p}.$$

*For $p = 2$, this is called the **Euclidean metric**.* ∎

One begins to understand these metrics by picturing the appropriate "unit circle" [2] in l^p, centered at the origin of $\mathbf{R}^2$. Figures 2,3 and 4 show these spheres for $p = 1, 2, 10$.

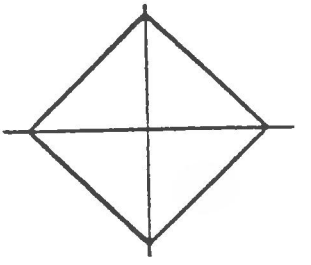


Figure 2. $\{\mathbf{x} \in \mathbf{R}^2 \mid \|\mathbf{x}-0\|_1\} = \{(x_1,x_2) \mid |x_1| + |x_2| = 1\}$

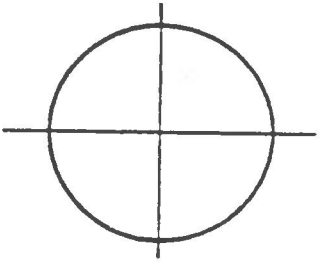


Figure 3. $\{\mathbf{x} \in \mathbf{R}^2 \mid \|\mathbf{x}-0\|_2 = 1\} = \{(x_1,x_2) \mid x_1^2 + x_2^2 = 1\}$

---

[1] In the symbol "l^p", "l" stands for "Lebesgue." This metric plays an important role in Lebesgue measure theory, see Subsection II.2.6.

[2] These sets are honest circles only for the l^2 metric, but they are all referred to as "circles." This should cause no confusion, as long as you remember that we are not doing plane geometry now. In higher dimensions, those sets are called "spheres". The set of points on and inside a sphere is called a "disk" in two dimensions, and a "ball" in two or more dimensions.

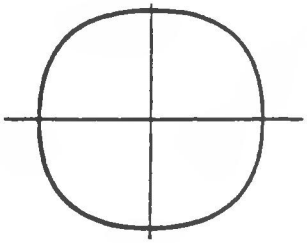


Figure 4. $\{\mathbf{x} \in \mathbf{R}^2 \,|\, \|\mathbf{x} - 0\|_{10} = 1\} = \{(x_1, x_2) \,|\, x_1^{10} + x_2^{10} = 1\}$

As $p \to \infty$, the "circle" becomes a square (see Exercise 1.1.11), and we are led to

1.1.5 Definition. *The* l^∞ *metric on* $\mathbf{R}^n$ *is*

$$d_\infty(\mathbf{x}, \mathbf{y}) \overset{\text{notation}}{=} \|\mathbf{x} - \mathbf{y}\|_\infty \overset{\text{def}}{=} \max\{|x_1 - y_1|, \cdots, |x_n - y_n|\}. \blacksquare$$

The set $\{\mathbf{x} \in \mathbf{R}^2 | \|\mathbf{x} - 0\|_\infty = 1\} = \{(x_1, x_2) | \max\{|x_1|, |x_2|\} = 1\}$ is the l^∞ unit "circle" in $\mathbf{R}^2$, drawn in Figure 5.

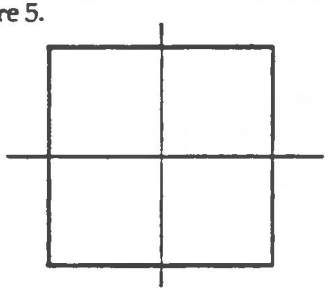


Figure 5. *The* l^∞ *unit circle*

1.1.6 Example. The standard Euclidean distance of a point $\mathbf{x} : (x_1, x_2)$ from the origin is equal to the radius of the circle (centered at the origin) that passes through $\mathbf{x}$. Analogously, one can find the l^p distance between $\mathbf{x}$ and $(0, 0)$ from a simple sketch. For example, the l^1 distance between $(0, 0)$ and $(1, 2)$ is 3 (Figure 6). Draw a picture from which you can read off the l^∞ distance between these two points. $\blacksquare$

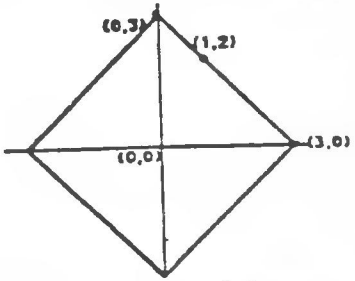


Figure 6. *Measuring* l^1 *distance*

1.1.7 Remark. Why does one impose the restriction $p \geq 1$ in the definition of l^p metric? Here is an example. Let $p = 1/2$, and consider

$$\begin{aligned} d(\mathbf{x},\mathbf{y}) &= (|x_1 - y_1|^{1/2} + |x_2 - y_2|^{1/2})^{1/\frac{1}{2}} \\ &= |x_1 - y_1| + 2\sqrt{|x_1 - y_1||x_2 - y_2|} + |x_2 - y_2|. \end{aligned}$$

If this is a metric, we must have $d(\mathbf{x},\mathbf{y}) \leq d(\mathbf{x},0) + d(0,\mathbf{y})$ or

$$\begin{aligned} |x_1 - y_1| + 2\sqrt{|x_1 - y_1| \cdot |x_2 - y_2|} + |x_2 - y_2| \leq |x_1| + 2\sqrt{|x_1| \cdot |x_2|} + |x_2| \\ + |y_1| + 2\sqrt{|y_1| \cdot |y_2|} + |y_2|. \end{aligned}$$

Make the special choice $x_2 = 0, y_1 = 0$: then we should have

$$|x_1| + 2\sqrt{|x_1||y_2|} + |y_2| \leq |x_1| + |y_2|$$

or

$$2\sqrt{|x_1||y_2|} \leq 0.$$

which is obviously incorrect if $x_1 y_2 \neq 0$.

In other words, the proposed metric does not satisfy the triangle inequality for the special pairs $\mathbf{x} = (x_1, 0)$, $\mathbf{y} = (0, y_2)$. So it is in fact *not* a metric. Similar difficulties exclude all other $p \in (0,1)$ from the definition of l^p metric. See Exercises (1.1.13), (1.1.14), and (1.1.59). ∎

• **1.1.8 Exercise.** Prove the triangle inequality for the l^1 metric on $\mathbf{R}^2$; this special case is elementary. ∎

• **1.1.9 Exercise.** Find the l^p distance between $(0,0)$ and $(1,1)$ as function of p. ∎

• **1.1.10 Exercise.** What is the largest r for which the l^2 circle $\{\mathbf{x} \in \mathbf{R}^2 \mid \|\mathbf{x} - 0\|_2 = r\}$ fits into the l^1 unit circle? What is the "radius" of the largest l^1 circle that will fit into the l^2 unit circle? ∎

1.1.11 Exercise. Prove that for $\mathbf{x}, \mathbf{y} \in \mathbf{R}^2$, $\lim_{p\to\infty} \|\mathbf{x} - \mathbf{y}\|_p = \|\mathbf{x} - \mathbf{y}\|_\infty$. ∎

• **1.1.12 Exercise.** Find the shortest distance, in the l^1 metric on $\mathbf{R}^2$, from the origin to the line $x_1 + x_2 = 2$. In the l^∞ metric. In the l^2 metric. Is the shortest distance in the l^p metric a monotone function of p? ∎

1.1.13 Exercise. Prove that $\|\mathbf{x} - \mathbf{y}\| \stackrel{\text{def}}{=} \left[|x_1 - y_1|^{1/3} + |x_2 - y_2|^{1/3} \right]^{1/\frac{1}{3}}$ is not a metric on $\mathbf{R}^2$. ∎

1.1.14 Exercise. Take the points $\mathbf{x} = (a,0), \mathbf{y} = (0,a), a \neq 0$, as in Remark 1.1.7. Let d be one of the l^p metrics for $p \geq 1$, or let it be the corresponding non-metric for $0 < p < 1$. Show that $d(\mathbf{x},\mathbf{y}) < d(\mathbf{x},0) + d(0,\mathbf{y})$ if $p > 1$, that $d(\mathbf{x},\mathbf{y}) = d(\mathbf{x},0) + d(0,\mathbf{y})$ if $p = 1$, and that $d(\mathbf{x},\mathbf{y}) > d(\mathbf{x},0) + d(0,\mathbf{y})$ if $p < 1$. Draw the circles of radius a centered on the origin, on $\mathbf{x}$, and on $\mathbf{y}$ in each of the three cases. Notice how they "bend" the wrong way when $p < 1$. ∎

The l^p metrics in $\mathbf{R}^n$ have analogs in n-dimensional complex vector spaces. This is fairly straightforward.

1.1.15 Definition. *C^n is the set of n-tuples of complex numbers. It is a complex n-dimensional vector space. For $\mathbf{z} = (z_1, \ldots, z_n)$ and $\mathbf{w} = (w_1, \ldots, w_n)$ in C^n, define the l^p metrics on C by*

$$\|\mathbf{z} - \mathbf{w}\|_p = \sqrt[p]{\sum_{j=1}^{n} |z_j - w_j|^p}, \quad 1 \le p < \infty. \tag{1.1.16}$$

$$\|\mathbf{z} - \mathbf{w}\|_\infty = \max\{|z_1 - w_1|, \ldots, |z_n - w_n|\}.$$

Here $|z_j - w_j|$ denotes the modulus of the complex number $z_j - w_j$. ∎

• **1.1.17 Exercise.** Sketch the unit circle for the l^1 metric on C. ∎

1.1.18 Exercise. C^2 can be identified in a natural way with $\mathbf{R}^4$: if $z_j = x_j + i\,y_j$, let $(z_1, z_2) \in C^2$ correspond to $(x_1, y_1, x_2, y_2) \in \mathbf{R}^4$. Does the unit sphere in the l^2 metric on C^2 correspond to the unit sphere in the l^2 metric on $\mathbf{R}^4$? Do the l^1 unit spheres correspond to each other? ∎

We get powerful generalizations by allowing vectors with infinitely many components. The resulting *sequence spaces* are important in many applications, and they also provide examples of some of the popular infinite-dimensional spaces — Hilbert spaces, Banach spaces, and the like. (Most of those topics are taken up in later chapters).

Notation. N denotes the set of natural numbers, $\{0, 1, 2, 3, \ldots\}$. Z is the set of integers, $\{\ldots, -2, -1, 0, 1, 2, \ldots\}$. Define:

$\mathbf{R}^{\mathbf{N}}$ = set of sequences $(r_0, r_1, r_2, \ldots)$ of real numbers,
$C^{\mathbf{N}}$ = set of sequences $(z_0, z_1, z_2, \ldots)$ of complex numbers,
$\mathbf{R}^{\mathbf{Z}}$ = set of sequences $(\ldots, r_{-2}, r_{-1}, r_0, r_1, r_2, \ldots)$ of real numbers,
$C^{\mathbf{Z}}$ = set of sequences $(\ldots, z_{-2}, z_{-1}, z_0, z_1, z_2, \ldots)$ of complex numbers.

Compare the notations: $\mathbf{R}^3$ consists of 3-tuples of real numbers; $\mathbf{R}^{\mathbf{N}}$ consists of N-tuples of real numbers. (Why would $\mathbf{R}^{\mathbf{R}}$ be a natural notation for the set of all real-valued functions on R?) ∎

We will introduce l^p-type metrics ($1 \le p \le \infty$) on certain subsets of these large sets. First, however, here is one example to illustrate why such definitions might be useful.

1.1.19 Example. In many problems, it helps to have a complicated function represented as a sum (perhaps infinite) of simple functions. You may know about Fourier series (which will be described in Chapter 3): many functions on $[0, 2\pi]$ can be written as

$$f(t) = \sum_{n=0}^{\infty} a_n \cos nt.$$

If, likewise, $g(t) = \sum_{n=0}^{\infty} b_n \cos nt$, then

$$f(t) - g(t) = \sum_{n=0}^{\infty} (a_n - b_n) \cos nt$$

and somehow the difference $f - g$ should be measured by the sequence

$$a_0 - b_0, a_1 - b_1, a_2 - b_2, \ldots.$$

So, one wants to define a "distance" $d(\mathbf{a}, \mathbf{b})$ between sequences $\mathbf{a} = (a_0, a_1, a_2, \ldots)$ and $\mathbf{b} = (b_0, b_1, b_2, \ldots)$, and — more important – to relate this to a distance defined directly in terms of f and g. In that way, one can compare functions (think of them as "signals") in the "time domain", i.e. $f(t)$ vs. $g(t)$, as well as in the so-called "frequency domain", a vs. b, by using the Fourier coefficients. ∎

In all the examples of sequence spaces, the axioms

$$d(\mathbf{x}, \mathbf{y}) \geq 0, \text{ and } = 0 \quad \text{iff } \mathbf{x} = \mathbf{y} \tag{0.1.5}$$

$$d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}) \tag{0.1.6}$$

will be easy to check, and you should always check them. The triangle inequality,

$$d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) \tag{0.1.7}$$

is *almost never* obvious. It will be proved for many examples all at once, in Subsection 1.3.

1.1.20 Definition. *Fix a $p \in [1, \infty)$. Let $l^p(\mathbf{R}, \mathbf{N})$ be the set of sequences in $\mathbf{R}^{\mathbf{N}}$ for which*

$$\sum_{j=0}^{\infty} |r_j|^p < \infty. \tag{1.1.21}$$

Let $l^\infty(\mathbf{R}, \mathbf{N})$ be the set of sequences in $\mathbf{R}^{\mathbf{N}}$ for which

$$\sup_{0 \leq j < \infty} |r_j| < \infty \tag{1.1.22}$$

(meaning that $(r_0, r_1, \ldots)$ is a bounded sequence). (See the Appendix to Section 2.3 for a summary of properties of sup, inf, and related matters). Likewise, for $p \in [1, \infty)$, let $l^p(\mathbf{R}, \mathbf{Z})$ be the set of sequences in $\mathbf{R}^{\mathbf{Z}}$ for which

$$\sum_{j=-\infty}^{\infty} |r_j|^p < \infty,$$

and let $l^\infty(\mathbf{R}, \mathbf{Z})$ be the set of sequences in $\mathbf{R}^{\mathbf{Z}}$ for which

$$\sup_{-\infty<j<\infty} |r_j| < \infty.$$

Define $l^p(\mathbf{C}, \mathbf{N})$, $l^p(\mathbf{C}, \mathbf{Z})$, $(1 \le p \le \infty)$ in a similar way for complex sequences. ∎

• **1.1.23 Exercise.** Let $x_j = 1/(j+1)$ for $j = 0, 1, 2, \ldots$. Show that this sequence belongs to $l^p(\mathbf{R}, \mathbf{N})$ if, and only if, $1 < p \le \infty$. ∎

• **1.1.24 Exercise.** The series $\sum_{j=0}^{\infty} (-1)^j/(j+1)$ converges conditionally, but not absolutely, and therefore the sequence $x_j = (-1)^j/(j+1)$ does not belong to $l^1(\mathbf{R}, \mathbf{N})$. ∎

• **1.1.25 Exercise.** Let $z_j = 1/(1 + \sqrt{-1}\,j)$. To which spaces $l^p(\mathbf{C}, \mathbf{Z})$ does this sequence belong? ∎

• **1.1.26 Exercise.** Find a sequence that belongs to $l^p(\mathbf{R}, \mathbf{Z})$ only for $p = \infty$. ∎

1.1.27 Proposition. *The sets listed in Definition (1.1.20) are all metric spaces, for the following analogues of the l^p metrics on $\mathbf{R}^n$ and $\mathbf{C}^n$:*

On $l^p(\mathbf{R}, \mathbf{N})$, $1 \le p < \infty$

$$\|\mathbf{x} - \mathbf{y}\|_p \stackrel{\text{def}}{=} \sqrt[p]{\sum_{j=0}^{\infty} |x_j - y_j|^p}$$

where $\mathbf{x} = (x_0, x_1, x_2, \ldots)$, $\mathbf{y} = (y_0, y_1, y_2, \ldots)$, and on $l^\infty(\mathbf{R}, \mathbf{N})$,

$$\|\mathbf{x} - \mathbf{y}\|_\infty \stackrel{\text{def}}{=} \sup_{0 \le j < \infty} |x_j - y_j|.$$

The sequences belonging to $l^p(\mathbf{R}, \mathbf{Z})$ are infinite in both directions and those in $l^p(\mathbf{C}, \mathbf{N})$ and $l^p(\mathbf{C}, \mathbf{Z})$ have complex entries. The modifications in the definitions of the metrics will be evident. ∎

1.1.28 Remarks

A. At the present introductory level, there is no real difference between the spaces $l^p(\mathbf{R}, \mathbf{N})$, $l^p(\mathbf{R}, \mathbf{Z})$, $l^p(\mathbf{C}, \mathbf{N})$, and $l^p(\mathbf{C}, \mathbf{Z})$. Therefore (as in the last proposition), I will usually state a result for $l^p(\mathbf{R}, \mathbf{N})$, and let you make the straightforward translation to the other cases. In the more advanced theory, there are indeed differences between real and complex vector spaces — one sees this already in finite-dimensional linear algebra.

B. The subset of $l^p(\mathbf{R}, \mathbf{N})$ consisting of sequences of the form

$$(x_0, x_1, \ldots, x_{n-1}, 0, 0, \ldots),$$

(*i.e.* $x_j = 0$ for $j \geq n$) may be identified with $\mathbf{R}^n$. This subset belongs to all $l^p(\mathbf{R}, \mathbf{N}), 1 \leq p \leq \infty$, and on this subset, the l^p metric just defined on the sequence spaces is precisely the l^p metric in Definitions (1.1.4) and (1.1.5). Therefore, once we verify the metric space axioms for $l^p(\mathbf{R}, \mathbf{N})$, we will also have verified them for $\mathbf{R}^n$. In most other respects, the sequence spaces $l^p(\mathbf{R}, \mathbf{N})$ are much subtler generalizations of the metric spaces $(\mathbf{R}^n, l^p)$: the convergence requirement (1.1.21) and the boundedness condition (1.1.22) introduce significant complications.

C. In everyday practice, the sets

$$l^1, \quad l^2, \quad \text{and} \quad l^\infty$$

are the most important. The corresponding metrics,

$$\|\mathbf{x} - \mathbf{y}\|_1 = \sum_{j=0}^{\infty} |x_j - y_j|$$

$$\|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{j=0}^{\infty} |x_j - y_j|^2}$$

$$\|\mathbf{x} - \mathbf{y}\|_\infty = \sup_{0 \leq j < \infty} |x_j - y_j|$$

are the most pleasant to behold. (The middle one is just Euclidean distance). Exotic ones like

$$\|\mathbf{x} - \mathbf{y}\|_{3/2} = \left[\sum_{j=0}^{\infty} |x_j - y_j|^{3/2}\right]^{2/3}$$

actually come up in some nonlinear problems, at a more advanced level. Still, many results hold for all l^p's, and it costs nothing to keep the semblance of generality. ∎

Proposition 1.1.27 is quite difficult to prove, and its implications will be understood only gradually. For the time being, I will be content with short discussions of a few basic properties of l^p spaces.

1.1.29 Remark. *On the verification of the metric space axioms.* What has to be proved? Let's take $l^p(\mathbf{R}, \mathbf{N})$, with $1 \le p < \infty$, as an example.

a) (0.1.5): $\sum_{j=0}^{\infty} |x_j - y_j|^p = 0$ if, and only if, $x_j = y_j$ for every j. Easy (do it!).

b) (0.1.6): $\sum_{j=0}^{\infty} |x_j - y_j|^p = \sum_{j=0}^{\infty} |y_j - x_j|^p$. That's even easier.

c) (0.1.7):

$$\sqrt[p]{\sum_{j=0}^{\infty} |x_j - z_j|^p} \le \sqrt[p]{\sum_{j=0}^{\infty} |x_j - y_j|^p} + \sqrt[p]{\sum_{j=0}^{\infty} |y_j - z_j|^p}.$$

This one, as has been advertised, is quite difficult.

d) There is another step that must not be overlooked. A typical element $\mathbf{x} \in l^p(\mathbf{R}, \mathbf{N})$ is a sequence satisfying $\sum_{j=0}^{\infty} |x_j|^p < \infty$. The l^p distance between two such sequences $\mathbf{x}, \mathbf{y}$ was defined to be

$$\sum_{j=0}^{\infty} |x_j - y_j|^p. \tag{1.1.30}$$

Could it be that this latter series diverges? If so, the definition of l^p distance would make no sense, because $d(\mathbf{x}, \mathbf{y})$ might be ∞ for some pairs $\mathbf{x}, \mathbf{y}$.

We must therefore prove, before trying to verify any of the axioms a), b), or c), that *the series (1.1.30) always converges.*

This is easy in two cases:

$$\text{If } p = 1, \text{ then } \|\mathbf{x} - \mathbf{y}\|_1 \stackrel{\text{def}}{=} \sum_{j=0}^{\infty} |x_j - y_j| \le \sum_{j=0}^{\infty} |x_j| + |y_j| = \sum_{j=0}^{\infty} |x_j| + \sum_{j=0}^{\infty} |y_j|,$$

and both series on the right converge for $\mathbf{x}, \mathbf{y} \in l^1(\mathbf{R}, \mathbf{N})$. Hence the first series also converges.

If $p = \infty$, then

$$\sup_{0 \le j < \infty} |x_j - y_j| \le \sup_{0 \le j < \infty} |x_j| + |y_j| \le \sup_{0 \le j < \infty} |x_j| + \sup_{0 \le j < \infty} |y_j|$$

and both suprema are finite since $\mathbf{x}, \mathbf{y} \in l^{\infty}(\mathbf{R}, \mathbf{N})$.

For $1 < p < \infty$, the matter is more tricky: how does convergence of

$$\sum_{j=0}^{\infty} |x_j - y_j|^p$$

follow from convergence of $\sum_{j=0}^{\infty} |x_j|^p$ and $\sum_{j=0}^{\infty} |y_j|^p$?

Answer: Let r, s be real numbers. We know that

$$|r - s| \leq |r| + |s|,$$

and so

$$|r - s| \leq 2 \times \max\{|r|, |s|\},$$

whence

$$\begin{aligned} |r - s|^p &\leq 2^p \times \max\{|r|^p, |s|^p\} \\ &\leq 2^p(|r|^p + |s|^p). \end{aligned}$$

Therefore

$$\sum_{j=0}^{\infty} |x_j - y_j|^p \leq \sum_{j=0}^{\infty} 2^p(|x_j|^p + |y_j|^p)$$

and the series on the right converges if $x, y \in l^p(\mathbf{R}, \mathbf{N})$. Hence, the series on the left also converges.

The argument adapts in an obvious way to the other cases ($l^p(\mathbf{R}, \mathbf{Z})$, *etc.*). So at least the purported metric is finite on $l^p(\mathbf{R}, \mathbf{N})$ and on the other spaces defined in (1.1.20). ∎

Example. ***Relations between $\mathbf{l^p}$ spaces.*** The sets $l^p(\mathbf{R}, \mathbf{N})$ are *distinct* subsets of $\mathbf{R}^{\mathbf{N}}$. For example, the sequence $(1, 1, 1, \ldots)$ belongs to $l^\infty(\mathbf{R}, \mathbf{N})$ because

$$\sup_{0 \leq j < \infty} 1 < \infty,$$

but not to any other $l^p(\mathbf{R}, \mathbf{N})$, since

$$\sum_{j=0}^{\infty} 1^p = \infty, \quad 1 \leq p < \infty$$

(cf. Exercise 1.1.26). At the opposite extreme, the harmonic sequence

$$(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots)$$

does not belong to $l^1(\mathbf{R}, \mathbf{N})$ because $\sum_{j=0}^{\infty} |\frac{1}{j+1}| = \infty$, but it *does* belong to $l^p(\mathbf{R}, \mathbf{N})$ for $1 < p \leq \infty$:

$$\sum_{j=0}^{\infty} |\frac{1}{j+1}|^p < \infty, \quad p > 1$$

and

$$\sup_{1 \leq j < \infty} |\frac{1}{j+1}| < \infty$$

(cf. Exercise 1.1.23).

The general result is:

1.1.31 Proposition. *Fix p and q satisfying $1 \le p < q \le \infty$. Then: a) $l^p(\mathbf{R}, \mathbf{N}) \subset l^q(\mathbf{R}, \mathbf{N})$; b) the inclusion is strict: there are sequences in l^q which do not belong to l^p.* ∎

Proof. Part a) requires us to show that

$$\text{whenever } \sum_{j=0}^{\infty} |x_j|^p < \infty \quad (1 \le p < \infty) \tag{1.1.32}$$

$$\text{then } \sum_{j=0}^{\infty} |x_j|^q < \infty \quad \text{for } q > p.$$

First suppose that $q = \infty$. A sequence $\mathbf{x} = (x_0, x_1, \ldots)$ for which the series (1.1.32) converges is certainly bounded,

$$\sup_{0 \le j < \infty} |x_j| < \infty,$$

and therefore

$$l^p(\mathbf{R}, \mathbf{N}) \subset l^\infty(\mathbf{R}, \mathbf{N}).$$

If $q < \infty$, argue as follows:

$$\begin{aligned} \sum_{j=0}^{\infty} |x_j|^q &= \sum_{j=0}^{\infty} |x_j|^{q-p} |x_j|^p \\ &\le \sum_{j=0}^{\infty} \Big(\sup_{0 \le j < \infty} |x_j|^{q-p} \Big) \times |x_j|^p \\ &= \Big(\sup_{0 \le j < \infty} |x_j|^{q-p} \Big) \sum_{j=0}^{\infty} |x_j|^p < \infty. \end{aligned}$$

Part b) is a little easier. We need only *one* sequence that is in $l^q(\mathbf{R}, \mathbf{N})$ but not in $l^p(\mathbf{R}, \mathbf{N})$, and the standard candidate in this kind of problem is a variant of the series $\sum j^{-\alpha}$, because we know exactly for which α it converges. The idea here is to find an exponent α so that the sequence $(1, (\frac{1}{2})^\alpha, (\frac{1}{3})^\alpha, \ldots)$ belongs to $l^q(\mathbf{R}, \mathbf{N})$:

$$\sum_{j=0}^{\infty} \Big(\frac{1}{j+1} \Big)^{\alpha q} < \infty$$

but not to $l^p(\mathbf{R}, \mathbf{N})$ (remember, $p < q$),

$$\sum_{j=0}^{\infty} \Big(\frac{1}{j+1} \Big)^{\alpha p} = \infty.$$

So, we want $\alpha p \leq 1$ and $\alpha q > 1$ or

$$\frac{1}{q} < \alpha \leq \frac{1}{p}.$$

Since $p < q \Longrightarrow 1/q < 1/p$, such an α exists — e.g., $\alpha = 1/p$. ∎

• **1.1.33 Exercise.** Acquire the habit of checking general arguments by substituting numbers for symbols. With $p = 2, q = 3$, say,

$$\sum_{j=0}^{\infty} |\frac{1}{j+1}|^3 = \sum_{j=0}^{\infty} |\frac{1}{j+1}| \cdot |\frac{1}{j+1}|^2 \leq \sum_{j=0}^{\infty} |\frac{1}{j+1}|^2$$

so the fact that $(\frac{1}{j+1}) \in l^2$ implies that $(\frac{1}{j+1}) \in l^3$. Adapt this argument to show that $(\frac{1}{j+1}) \in l^2$ implies that this sequence belongs to l^{20}. ∎

1.1.34 Exercise. Let $1 \leq p < \infty$. Is there a sequence $\mathbf{x} \in \mathbf{R}^{\mathbf{N}}$ which belongs to $l^r(\mathbf{R}, \mathbf{N})$ for all $r \in (p, \infty)$ and no other r? For all $r \in [p, \infty)$ and no other r? ∎

1.1.35 Example. A refinement of the sets $l^p(\mathbf{R}, \mathbf{N})$ is afforded by these subsets of $\mathbf{R}^{\mathbf{N}}$:

$$l^{p,k}(\mathbf{R}, \mathbf{N}) \stackrel{\text{def}}{=} \{\mathbf{x} \in l^p(\mathbf{R}, \mathbf{N}) \,|\, \sum_{j=0}^{\infty} (1 + j^p + j^{2p} + \ldots + j^{kp}) |x_j|^p < \infty\}$$

$(1 \leq p < \infty, \quad k = 0, 1, 2, \ldots)$. The metric is

$$\|\mathbf{x} - \mathbf{y}\|_{p,k} \stackrel{\text{def}}{=} \sqrt[p]{\sum_{j=0}^{\infty} (1 + j^p + j^{2p} + \ldots + j^{kp}) |x_j - y_j|^p}. \tag{1.1.36}$$

A brief explanation of why such funny sets arise:

If $f(t) = \sum_{j=0}^{\infty} a_j \cos jt$, then one hopes that

$$f'(t) = -\sum_{j=0}^{\infty} j a_j \sin jt$$

$$f''(t) = -\sum_{j=0}^{\infty} j^2 a_j \cos jt$$

etc. To compare $f(t)$ and $g(t)$, as well as their derivatives up to order k, one therefore needs to compare, *simultaneously*, the sequences

$$(j^r a_j) \text{ and } (j^r b_j) \quad \text{for } 0 \leq r \leq k.$$

The metrics $\|\cdot - \cdot\|_{p,k}$ are, in this respect, more sensitive as measure of distance. For instance, $\sum_{j=0}^{\infty} (j+1)^{-1}$ belongs to $l^2(\mathbf{R}, \mathbf{N})$ but not to $l^{2,1}(\mathbf{R}, \mathbf{N})$. These sets play an important role in certain applications; they are called *Sobolev spaces*. We will meet them on occasion. ∎

• **1.1.37 Exercise.** Prove that $l^{p,k}(\mathbf{R}, \mathbf{N}) \subset l^p(\mathbf{R}, \mathbf{N})$ for all integers $k \geq 0$. ∎

• **1.1.38 Exercise.** Find a sequence that belongs to $l^{2.1}$ but not to $l^{2.2}$. ∎

1.1.39 SUPPLEMENT: Multiple series

Spaces like $l^p(\mathbf{R}, \mathbf{N})$ can be generalized by changing the index set N, e.g. to $\mathbf{N}^2 \stackrel{\text{def}}{=} \{(n_1, n_2) | \, n_1, n_2 \in \mathbf{N}\}$, or, more generally, to $\mathbf{N}^k$. For instance, $l^p(\mathbf{R}, \mathbf{N}^2)$ consists of sequences $(x_{i,j})$, $i, j \in \mathbf{N}$, such that

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |x_{i,j}|^p < \infty. \tag{1.1.40}$$

On an abstract level, this new space is really the same as $l^p(\mathbf{R}, \mathbf{N})$, because the indexing sets N and $\mathbf{N}^2$ are both countably infinite and can be put in one-to-one correspondence. To so this, arrange the pairs $(n_1, n_2) \in \mathbf{N}^2$ in a rectangular array, as follows:

$$\begin{array}{lllll} (0,0) & (0,1) & (0,2) & \ldots \\ (1,0) & (1,1) & (1,2) & \ldots \\ (2,0) & (2,1) & (2,2) & \ldots \\ \vdots & \vdots & \vdots & \end{array}$$

The pairs are then identified with the natural numbers by following the dotted curve: $(0,0) \mapsto 0, (0,1) \mapsto 1, (1,0) \mapsto 2$, etc.[3]

A double series such as the one in (1.1.39) can thereby be rewritten as simple series.[4] This helps in one respect: once the triangle inequality is proved for $l^p(\mathbf{R}, \mathbf{N})$, it will hold automatically for $l^p(\mathbf{R}, \mathbf{N}^2)$ or, more generally, for $l^p(\mathbf{R}, \mathbf{N}^k)$ for any $k = 1, 2, 3, \ldots$.

On the other hand, the renumbering process is complicated enough to be useless in the study of specific examples. Consider, for example, the series

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{1}{(i+1)^2 + (j+1)^2}.$$

Does this converge? To what l^p spaces does the sequence $(x_{i,j}) = (1/[(i+1)^2 + (j+1)^2])$ belong? You are invited to rewrite the double series as a single series — not much will come of the attempt, I suspect.

---

[3] This is known as *Cantor's diagonal process*; it is also used to show that the rational numbers are a countable set. In that version, the pair (n, m), $n, m > 0$ is replaced by the rational number n/m. The fractions in which m and n have a common factor are crossed out so that they are not counted more than once, and the remaining ones are put into a one-to-one correspondence with the natural numbers.

[4] The following result from advanced calculus is used: an absolutely convergent series can be rearranged at will, and the rearranged series have the same limit. See II.2.5.50 and II.2.7.14.

Let us show, to illustrate some of the subtleties, that

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{(i^2+j^2)^2}<\infty.$$

(The dummy indices were changed, because the latter form of the series is easier to type and to look at). To this end, we study the partial sum

$$S_M \stackrel{\text{def}}{=} \sum_{i=1}^{M}\sum_{j=1}^{M}\frac{1}{(i^2+j^2)^2}.$$

We divide S_M into three pieces: $i=j$, $i>j$, and $i<j$:

$$S_M=\sum_{k=1}^{M}\frac{1}{(2k^2)^2}+\sum_{i=1}^{M}\sum_{j=1}^{i-1}\frac{1}{(i^2+j^2)^2}+\sum_{j=1}^{M}\sum_{i=1}^{j-1}\frac{1}{(i^2+j^2)^2}. \tag{1.1.41}$$

The first sum is the partial sum of a convergent series. The object is to dominate each of the two double sums by partial sums of convergent series.

Look at the first double sum. Since

$$\frac{1}{(i^2+j^2)^2}\le\frac{1}{(i^2)^2},$$

we have

$$\sum_{j=1}^{i-1}\frac{1}{(i^2+j^2)^2}\le\sum_{j=1}^{i-1}\frac{1}{(i^2)^2}=\frac{(i-1)}{(i^2)^2}<\frac{i}{(i^2)^2}=\frac{1}{i^3},$$

whence

$$\sum_{i=1}^{M}\sum_{j=1}^{i-1}\frac{1}{(i^2+j^2)^2}\le\sum_{i=1}^{M}\frac{1}{i^3}.$$

A similar estimate holds for the second sum in (1.1.41). Hence,

$$S_M<\sum_{k=1}^{M}\frac{1}{2^2k^4}+\sum_{i=1}^{M}\frac{1}{i^3}+\sum_{j=1}^{M}\frac{1}{j^3}.$$

Each term on the right side is the partial sum of a convergent series. Hence, the partial sums S_M are bounded. More generally, the partial sums $S_{MN}=\sum_{i=1}^{M}\sum_{j=1}^{N}$ are bounded by $S_{\max\{M,N\}}$, which we just found to be bounded. It follows that the original double series converges.

• **1.1.42 Exercise.** Adapt the technique in this example to prove that

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{1}{[(i+1)^2+(j+1)^2]^{\alpha}}$$

converges for $\alpha > 1$. Prove that the double series does not converge when $\alpha = 1$. ∎

1.1.43 Exercise*. Let ω_1, ω_2 be complex numbers satisfying $\Im(\omega_1/\omega_2) > 0$, where $\Im$ denotes the imaginary part of a complex number. For what values of α does

$$\sum_{m=-\infty}^{\infty}{\sum_{n=-\infty}^{\infty}}'\frac{1}{(m\omega_1+n\omega_2)^{\alpha}}$$

converge? The $'$ indicates that the summation indices do not take on the value $(m,n) = (0,0)$. (This series arises in the theory of Weierstrass elliptic functions). ∎

1.1.44 Exercise. Suppose that $1 \le p < q$. Is it true that $\mathcal{P}(\mathbf{R},\mathbf{N}^2) \subset l^q(\mathbf{R},\mathbf{N}^2)$? (Compare (1.1.31)). ∎

1.1.45 Example. *Sequences of 0's and 1's.* Lest the reader think that all metrics must be defined on n-tuples of real or complex numbers, we present as our final example a sequence space with a new wrinkle: the sequences are made up of 0's and 1's only.

Let M be the set of all sequences $\underline{\epsilon} = (\cdots, \epsilon_{-2}, \epsilon_{-1}, \epsilon_0, \epsilon_1, \epsilon_2, \cdots)$, infinite in both directions, where each entry ϵ_i is either 0 or 1.

Such sets have come to play a big role in the study of differential equations and "chaotic behavior". Think of the following experiment: from the beginning of time to the end of time, a coin is flipped every second. At time $t = i$, record "0" if the coin comes up "tails," and record "1" if it comes up "heads." The life history of one sequence of coin tosses is then one element of the set M.

We now show that a metric is defined on M by

$$d(\underline{\epsilon},\underline{\epsilon}') = \sum_{j=-\infty}^{\infty}\frac{|\epsilon_j - \epsilon'_j|}{2^{|j|}}. \tag{1.1.46}$$

Note that since

$$|\epsilon_j - \epsilon'_j| \le \begin{cases} |0-0| \\ |1-0| \\ |0-1| \\ |1-1| \end{cases} \le 1,$$

the series (1.1.46) is dominated by $\sum_{j=-\infty}^{\infty} 1/2^{|j|} < \infty$, and therefore converges.

Obviously, $d(\underline{\epsilon},\underline{\epsilon}') \ge 0$, and $= 0$ iff $\underline{\epsilon} = \underline{\epsilon}'$, and equally obviously, $d(\underline{\epsilon},\underline{\epsilon}') = d(\underline{\epsilon}',\underline{\epsilon})$. Even the triangle inequality is easy in this example:

$$d(\underline{\epsilon},\underline{\epsilon}'') = \sum_{j=-\infty}^{\infty}\frac{|\epsilon_j - \epsilon''_j|}{2^{|j|}} \le \sum_{j=-\infty}^{\infty}\frac{|\epsilon_j - \epsilon'_j| + |\epsilon'_j - \epsilon''_j|}{2^{|j|}}$$

$$= d(\underline{\epsilon},\underline{\epsilon}') + d(\underline{\epsilon}',\underline{\epsilon}'').$$

All we need is the triangle inequality $|\epsilon_j - \epsilon''_j| \le |\epsilon_j - \epsilon'_j| + |\epsilon'_j - \epsilon''_j|$ for the numbers $0, 1$, and that is trivial.

So we have a metric. Now ask a more interesting question: when are two sequences $\underline{\epsilon}, \underline{\epsilon}'$ "close"?

The first observation is that two sequences in this metric space are never *very* far apart. Indeed,

$$\sum_{j=-\infty}^{\infty} \frac{|\epsilon_j - \epsilon'_j|}{2^{|j|}} \le \sum_{j=-\infty}^{\infty} \frac{1}{2^{|j|}} = 3,$$

so the distance between two sequences cannot exceed 3.

1.1.47 Exercise. What is the largest possible distance between two sequences? ∎

A specific example: if $\epsilon_0 = 1$ and $\epsilon'_0 = 0$, then the distance between $\underline{\epsilon}$ and $\underline{\epsilon}'$ is at least 1 (because of the term $|\epsilon_0 - \epsilon'_0|/2^0$).

It is easy to prove that for $d(\underline{\epsilon}, \underline{\epsilon}')$ to be $< 1/2^N$, the sequences $\underline{\epsilon}, \underline{\epsilon}'$ must agree *at least* for $|j| \le N$.

In the language of coin tosses: two coin-toss sequences are close if they produce heads and tails in the same order in the *recent past and near future*, $(-N \le j \le N)$. What happened long ago (or will happen much later) is not so important.

This is perhaps not the measure of distance you would have thought of yourself, and it is instructive to imagine other potential metrics.

How about $d(\underline{\epsilon}, \underline{\epsilon}') = \sup_{-\infty < j < \infty} |\epsilon_j - \epsilon'_j|$?

Well, if $\underline{\epsilon}$ and $\underline{\epsilon}'$ differ at all, $|\epsilon_j - \epsilon'_j| = 1$ for at least one j, and two different sequences are always at distance 1. This *is* a metric, but obviously not a very discriminating one. (It is the *discrete* metric introduced, for an arbitrary set M, in (1.0.10)).

Perhaps one should count the average deviation between $\underline{\epsilon}$ and $\underline{\epsilon}'$ without favoring the present, which corresponds to $j \sim 0$:

$$d(\underline{\epsilon}, \underline{\epsilon}') = \lim_{n \to \infty} \frac{1}{2n+1} \sum_{j=-n}^{n} |\epsilon_j - \epsilon'_j|. \tag{1.1.48}$$

Notice that $\sum_{j=-n}^{n} |\epsilon_j - \epsilon'_j|$ = number of times ϵ_j and ϵ'_j differ between $-n$ and n, so this "metric" measures a limiting value of the number of differences per flip of the coin.

Is this a metric at all? No, *it is not*, because if $\underline{\epsilon}, \underline{\epsilon}'$ differ in exactly one entry (say, at $j = 0$), the limit in (1.1.48) will be zero, and we would have

$$d(\underline{\epsilon}, \underline{\epsilon}') = 0 \text{ even though } \underline{\epsilon} \ne \underline{\epsilon}'.$$

So this proposal is no good, either, and we ask:

Why should a metric, (1.1.46), which places so much emphasis on the present time ($-N \le j \le N$), and ignores huge differences in the past or future, be useful?

Perhaps because in random processes, like a coin toss, the interest is often in *predicting* what will happen next. For that purpose, the recent past is more important than the remote past — which presumably has less influence on the future developments. For a process as random as a coin toss, this explanation is a bit misleading, but it will have to suffice.

In any case, this type of question: "does the metric really measure important differences?" is basic for people who try to quantify the distance between objects of interest in a given problem. ∎

• **1.1.49 Exercise.** Let M be the set of sequences (ϵ_j), $j \in \mathbf{Z}$, in which each ϵ_j is 0, 1 or 2. Define a metric analogous to the one in the preceding example, and verify that the axioms for a metric space are satisfied. ∎

Miscellaneous Exercises

• **1.1.50 Exercise.** In the axioms for a metric space, (1.0.5,6,7), omit the condition $d(x, y) \ge 0$ from (1.0.5). Show that it follows from the remaining axioms. (Hint: set $z = x$ in (1.0.7).) ∎

1.1.51 Exercise. The following convergence test, called the *Cauchy test*, is useful for checking "borderline" series.

Let $\{a_n\}$ be a non-increasing sequence of positive numbers. Then $\sum_n a_n$ converges if, and only if,

$$\sum_k 2^k a_{2^k}$$

converges.

Prove this. (Hint:

$$a_1 + (a_2 + a_3) + (a_4 + a_5 + a_6 + a_7) + \ldots \le a_1 + 2a_2 + 4a_4 + \ldots.$$

and so on.)[1] ∎

1.1.52 Exercise. Use the criterion from the preceding exercise to test for convergence of the following series:

$$\sum_n \frac{1}{n}, \quad \sum_n \frac{1}{n \ln n}, \quad \sum_n \frac{1}{n(\ln n)^2}. \ ∎$$

---

[1] I have seen this wonderful convergence test in only one place: the classic text by K. Knopp, *Theory and Applications of Infinite Series*. You really should leaf through it. One of the highlights of my early mathematics education was working through Knopp's derivation of all the important properties of trigonometric functions from their power series expansions. While I'm making propaganda: Knopp's five slender volumes on *Complex Function Theory*, available as Dover reprints, are still my favorite text on that subject.

• **1.1.53 Exercise.** Let M be the set of real 2×2 matrices. For two matrices

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \mathbf{A}' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix},$$

define

$$d(\mathbf{A}, \mathbf{A}') = \max\{|a - a'| + |b - b'|, |c - c'| + |d - d'|\}.$$

Show that this is a metric on M. ∎

1.1.54 Exercise. Let M be the surface of a sphere of unit radius, and define the distance between two points P, P' to be the arclength of the shorter of the two segements of the great circle that passes through P and P'. Show that the axioms for a metric space are satisfied. ∎

1.1.55 Exercise. Define a function $f : \mathbf{R}^2 \to \mathbf{R}$ by $f(x_1, x_2) = x_1^2 + x_2^2$. Show that f is continuous in the l^1 metric on $\mathbf{R}^2$. (Hint: show that

$$\|\mathbf{x}^{(n)} - \mathbf{x}\|_1 \to 0 \Rightarrow f(\mathbf{x}^{(n)}) \to f(\mathbf{x}).) \blacksquare$$

• **1.1.56 Exercise.** Find the distance of the point $(1, -1)$ from the line $x_1 + x_2 = 1$ in the l^p metric on $\mathbf{R}^2$, as function of p. ∎

• **1.1.57 Exercise.** Find the shortest distance between the origin and the plane $x_1 + x_2 + x_3 = 1$ in the l^1, l^2, l^∞ metrics on $\mathbf{R}^3$. ∎

1.1.58 Exercise. In $\mathbf{R}^n$: find the shortest distance between the origin and the plane $x_1 + \ldots + x_n = 1$, in the l^1, l^2 and l^∞ metrics. ∎

1.1.59 Exercise. Let $0 < p < 1$. Show that

$$d_p(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \sum_{j=1}^{n} |x_j - y_j|^p$$

defines a metric on $\mathbf{R}^n$. ∎

Subsection 1.2 Equivalence of Metrics

In the preceding subsection, we saw that it is possible to equip one set (such as $\mathbf{R}^n$) with infinitely many *apparently* different metrics, the l^p metrics for $1 \leq p \leq \infty$. One question will have occurred to every reader: does the choice of metric ever matter? This is a sensible question, which deserves a better answer than I can give at this moment: sometimes the choice of metric makes a difference, at other times it doesn't. It makes a difference, for example, when one tries to estimate the growth rate of the solution of a differential equation, and chooses a metric adapted to the equation in order to make computations easier. It often does not make a difference when one investigates convergence of sequences: the numerical value of the distance from the limit may be unimportant, and what matters is only that the distance tends to zero. Here is a simple example.

1.2.1 Example. In $\mathbf{R}^2$, consider the points $\mathbf{x}^{(n)}$ defined by

$$\mathbf{x}^{(n)} = (\frac{1}{n}, \frac{1}{n}).$$

Writing $0 = (0,0)$, we have

$$\|\mathbf{x}^{(n)} - 0\|_p = \begin{cases} 2^{1/p}/n, & \text{if } 1 \leq p < \infty, \\ 1/n, & \text{if } p = \infty. \end{cases}$$

For fixed n, these distances obviously change with p. No matter what p is chosen, however, one has

$$\lim_{n \to \infty} \|\mathbf{x}^{(n)} - 0\|_p = 0.$$

The points $\mathbf{x}^{(n)}$ approach the origin 0, no matter which l^p metric is used to measure the distance from the origin. ∎

In this subsection, we study the *equivalence of metrics*. Let there be a set M with two metrics, d and d'. *Equivalence* will mean the following: whenever metric d says that a sequence of points $x_n \in M$ has limit $x \in M$, meaning that the distance $d(x_n, x)$ tends to zero as $n \to \infty$, then metric d' agrees: $d'(x_n, x) \to 0$ as $n \to \infty$; conversely, whenever $d'(x_n, x) \to 0$ as $n \to \infty$, then also $d(x_n, x) \to 0$. Think of measuring distance on M through distorting lenses — we get different values for $d(x, y)$ and $d'(x, y)$ as we look at M with different lenses, but we all agree on the limits of sequences of points.

The official definition, of course, is less picturesque.

1.2.2 Definition. *Two metrics d and d' on a set M are said to be* **equivalent** *if, for every $x \in M$ and $\epsilon > 0$, there exist δ and δ' (in general depending on x and ϵ) such that*

$$d(x, y) < \delta \implies d'(x, y) < \epsilon$$

and

$$d'(x, y) < \delta' \implies d(x, y) < \epsilon. \text{ ∎}$$

In words: x and y will be as close as you please (within ϵ) in the metric d' if you just make them close enough (within δ) in the metric d, and they will be as close as you please (within ϵ) in d if you just make them close enough (within δ') in d'.

Let us try to capture this definition by a (schematic) picture. Suppose the "balls" $\{y \in M \mid d(x, y) < \eta\}$, having center at x and "radius" η, are drawn as disks, while the "balls" $\{y \in M \mid d'(x, y) < \eta\}$ are drawn as diamonds (Figure 7).

Figure 7. *Balls in different metrics*

(The shapes are derived from the l^2 and l^1 metrics on $\mathbf{R}^2$, but they are merely meant to suggest more abstract relationships.) Now give an $\epsilon > 0$. The first requirement in the definition,

$$d(x,y) < \delta \quad \Longrightarrow \quad d'(x,y) < \epsilon,$$

says that there is a δ such that the diamond, center at x, with preassigned small radius ϵ, will contain the disk, center at x, radius δ. (Left part of Figure 8.)

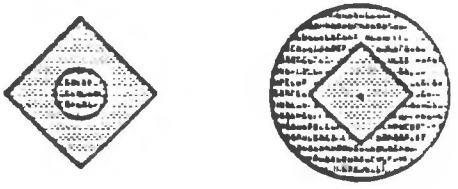


Figure 8. *Equivalence of metrics*

The second line in the Definition,

$$d'(x,y) < \delta' \quad \Longrightarrow \quad d(x,y) < \epsilon,$$

says that the disk, center at x, with preassigned small radius ϵ contains a diamond, center at x, having a sufficiently small δ' as radius. The right part of Figure 8 shows the inclusions.

We will also want to picture metrics that are *not* equivalent. To this end, we replace the diamond by a stranger shape (in Figure 9), which is still intended to suggest the d'-ball $\{y \in M \mid d'(x,y) < \eta\}$. The "center" x is drawn at the boundary of the "ball"; this is merely a pictorial trick that keeps us from fitting a (circular) d-ball centered at x, of arbitrarily small radius, into the d'-ball. We therefore have a picture for the *negation* of one condition in the definition of equivalence:

There is an x and an $\epsilon > 0$ such that for every $\delta > 0$ one can find a y with $d(x,y) < \delta$ but $d'(x,y) \geq \epsilon$.

Our picture negates only *half* of the definition of equivalence. Given $\epsilon > 0$ again, it does seem to be possible (Figure 10) to fit a sufficiently small d'-ball into the (circular) d-ball of radius ϵ. This illustrates: given $\epsilon > 0$, there is a δ', such that $d'(x,y) < \delta' \Longrightarrow d(x,y) < \epsilon$.

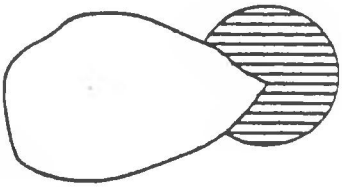


Figure 9. *How one metric can mess up equivalence*

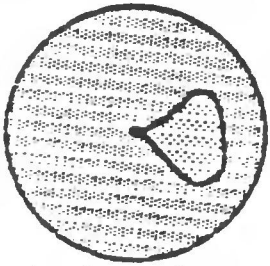


Figure 10. *How the other metric can try to be equivalent*

Should we want to illustrate negation of *both* parts of the definition of equivalence, we could draw the "center" x of the d-ball on the circumference of the disk. Try it.

I now leave the pictures, temporarily, and prove a little theorem about convergence. Note that the hypothesis uses only half of the definition of equivalence.

1.2.3 Proposition. *Suppose that*

(1.2.4)

for every $x \in M$ and $\epsilon > 0$ there is a $\delta > 0$ such that $d(x,y) < \delta \Longrightarrow d'(x,y) < \epsilon$.

Then it is impossible to find a sequence $\{x_n\}$, *an* $x \in M$, *and an* $a > 0$ *such that*

$$d(x_n, x) \to 0 \quad \text{as} \quad n \to \infty$$

while

$$d'(x_n, x) \geq a > 0 \quad \text{for infinitely many} \quad n.$$

(If x_n *gets close to* x *in* d, *it must get close to* x *in* d' *as well).*

Proof. Suppose the contrary: such $\{x_n\}, x$ and a exist. We show that this assumption leads to a contradiction. Pick an ϵ satisfying $0 < \epsilon < a$. Then, by (1.2.4), there is a $\delta > 0$ so that

$$d(y,x) < \delta$$

implies

$$d'(y,x) < \epsilon < a.$$

In particular, since $d(x_n, x) \to 0$, we will eventually have, for all sufficiently large n,

$$d(x_n, x) < \delta$$

and then

$$d'(x_n, x) < \epsilon < a,$$

and so $d'(x_n, x) \geq a$ for infinitely many x_n is impossible. ∎

There is a more standard way to state the Proposition; it involves the definition of *convergence* in a metric space.

1.2.5 Definition. *Let (M, d) be a metric space. Let $x_n \in M$ for $n \geq 1$, and let $x \in M$. One says that the sequence $\{x_n\}$ converges to x if, given any $\epsilon > 0$, there is an integer N_ϵ (depending on ϵ) so that*

$$d(x_n, x) < \epsilon \quad \text{for } n > N_\epsilon.$$

(For emphasis, we sometimes say: x_n converges to x in the metric d, even though this is implicit in the notation (M, d).)

Proposition 1.2.3 says that if condition (1.2.4) holds, then a sequence convergent in the metric d also converges in the metric d' (to the same limit). If d and d' are equivalent, then the proposition is true with d and d' interchanged, and we have the following corollary:

1.2.6 Corollary. *Let d and d' be equivalent metrics on the set M. Then a sequence x_n converges to x in the metric d if, and only if, it converges to x in the metric d'.*

It is worth another picture to illustrate the definition of convergence and Corollary (1.2.6)[2]. We suppose that x_n converges to x in the metric d. Definition 1.2.5 expresses this as: draw any d-ball with center x; then from some point on $(n > N_\epsilon)$ all x_n will lie within this ball. That is the left part of the Figure 11. The middle part shows the desired conclusion: every d'-ball with center at x will contain all x_n from some (other) point on.

---

[2] As a matter of general "proof strategy," one should *always* start with pictures that make the statement to be verified seem obvious. The official steps may not be clear yet, and may in fact be nontrivial, but an intuitive understanding is always preferable to a line-by-line verification of someone else's argument.

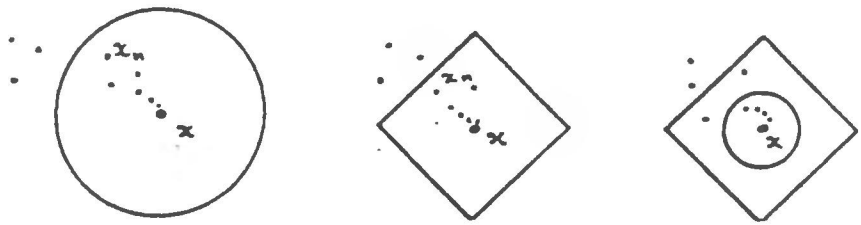


Figure 11. *Convergence*

By the definition of equivalence of d and d', our given diamond contains a disk. But that disk contains all x_n from some point on, and therefore the diamond also contains all x_n from some point on (see the right part of Figure 11). End of proof by picture. Now is a good time to translate this sketch into an official argument involving ϵ's and δ's!

• **1.2.7 Exercise.** Write a complete proof of Corollary (1.2.6) based on the pictorial argument just given. ∎

• **1.2.8 Exercise.** Refer to Proposition 1.2.3. Suppose that the condition (1.2.4) does *not* hold. Draw pictures that illustrate how one can have x_n converging to x in the metric d but not in the metric d'. ∎

So far, I have dealt only in generalities. It is time to look in some detail at the l^p metrics. The fundamental property of these metrics on $\mathbf{R}^n$ is their mutual equivalence. Hence, sequences that converge in the usual Euclidean metric converge in every one of the other metrics. All standard theorems about continuity, convergence, *etc.*, need only the Euclidean metric. That is why one does not meet the stranger l^p metrics in calculus courses. I have introduced them partly as preparation for infinite-dimensional problems, and partly because they are quite useful in numerical analysis.

1.2.9 Proposition. *For $1 \leq p \leq \infty$, the l^p metrics on $\mathbf{R}^n$ are mutually equivalent.* ∎

Proof. There are several tricks that help to streamline the argument.

First, we need only show that every l^p metric, $1 \leq p < \infty$, is equivalent to the l^∞ metric. Then it follows that any given l^p metric is equivalent to any given l^q metric. Indeed, an l^p-ball contains an l^∞-ball, because those two are equivalent metrics, and the l^∞-ball in turn contains an l^q-ball, because those two are also equivalent. (And vice versa: every l^q-ball contains an l^p-ball, with the same center of course.)[3]

[3] If you are familiar with *equivalence relations*, you will notice that this is just an application of the rule: $a \sim c$ and $b \sim c$ implies $a \sim b$. Equivalence of metrics is, indeed, an equivalence relation. See II.1.2.4.

Second, we may take all the balls to be centered at the origin 0, because translation to another center y will not change the radii or the fact that one "ball" is contained in the other.

The distances of a point x from 0 in the l^∞ and l^p metrics are then given by

$$d_\infty(\mathbf{x}, 0) = \max_{1 \le j \le n} |x_j| \quad \text{and} \quad d_p(\mathbf{x}, 0) = \sqrt[p]{\sum_{j=1}^{n} |x_j|^p}.$$

I will prove the following two relations between these quantities:

$$\max_{1 \le j \le n} |x_j| \le \sqrt[p]{\sum_{j=1}^{n} |x_j|^p} \tag{1.2.10}$$

and

$$\sqrt[p]{\sum_{j=1}^{n} |x_j|^p} \le n^{1/p} \max_{1 \le j \le n} |x_j|. \tag{1.2.11}$$

Equivalence is a consequence, in the following way. Let $\epsilon > 0$ be given. If one has $d_p(\mathbf{x}, 0) < \epsilon$, it follows from (1.2.10) that $d_\infty(\mathbf{x}, 0) < \epsilon$ as well. Hence, with the choice $\delta = \epsilon$,

$$d_p(\mathbf{x}, 0) < \delta \implies d_\infty(\mathbf{x}, 0) < \epsilon.$$

In the other direction, if

$$d_\infty(\mathbf{x}, 0) < \frac{\epsilon}{n^{1/p}},$$

then it follows from (1.2.11) that

$$d_p(\mathbf{x}, 0) < n^{1/p} \cdot \left(\frac{\epsilon}{n^{1/p}} \right) = \epsilon,$$

so that, with $\delta' = \epsilon / n^{1/p}$,

$$d_\infty(\mathbf{x}, 0) < \delta' \implies d_p(\mathbf{x}, 0) < \epsilon.$$

We must still prove (1.2.10) and (1.2.11). Let $|x_k| = \max_{1 \le j \le n} |x_j|$ (there may be more than one such k; that's OK). Then

$$|x_k|^p \le \sum_{j=1}^{n} |x_j|^p,$$

and by taking roots we find

$$\max_{1\le j\le n} |x_j| = |x_k| \le \sqrt[p]{\sum_{j=1}^{n} |x_j|^p}.$$

which is (1.2.10). Note that we used the monotonicity of the function $\sqrt[p]{\ }$.

With $|x_k|$ as before, we have

$$\sum_{j=1} |x_j|^p \le \sum_{j=1}^{n} |x_k|^p = n\,|x_k|^p.$$

Take the root again, to get

$$\sqrt[p]{\sum_{j=1}^{n} |x_j|^p} \le n^{1/p}|x_k| = n^{1/p} \max_{1\le j\le n} |x_j|,$$

which is (1.2.11). The proof is finished. ∎

1.2.12 Example. Equivalence of the l^p metrics on $\mathbf{R}$ has some down-to-earth and obvious consequences. Look again at $\mathbf{R}^2$. The l^2 metric on $\mathbf{R}^2$ is the familiar Euclidean distance. The l^∞ distance between x and y is $\max\{|x_1 - y_1|, |x_2 - y_2|\}$. Suppose that $\mathbf{x}^{(n)} \to \mathbf{x}$ in the Euclidean metric. Equivalence of the two metrics means that $\mathbf{x}^{(n)}$ also converges to x in l^∞, i. e. that the l^∞ distances $\|\mathbf{x}^{(n)} - \mathbf{x}\|_\infty$ approach zero. This, in turn, says just that

$$\max\{|x_1^{(n)} - x_1|, |x_2^{(n)} - x_2|\} \to 0.$$

In words: the two entries of $\mathbf{x}^{(n)}$ get close to the corresponding entries of x as $n \to \infty$, which is certainly believable. From our rarefied perspective, this result follows from, and is an illustration of, the equivalence of metrics. ∎

1.2.13 Example. The l^p metrics on $\mathbf{R}^n$ are equivalent, but there is a subtler point. Look again at inequality (1.2.11) in the proof of Proposition 1.2.9, from which we deduced that

$$d_\infty(\mathbf{x}, 0) < \frac{\epsilon}{n^{1/p}} \quad \Longrightarrow \quad d_p(\mathbf{x}, 0) < \epsilon. \tag{1.2.14}$$

The radius of the l^∞-ball that fits into the l^p-ball (of fixed radius ϵ) decreases as the dimension n increases. In the limit $n \to \infty$, the radius becomes zero, leading one to guess that somehow the "limiting metrics" on $l^\infty(\mathbf{R}, \mathbf{N})$ and $l^p(\mathbf{R}, \mathbf{N})$ are no longer equivalent. This is true, but one must be careful: the spaces $l^\infty(\mathbf{R}, \mathbf{N})$

and $l^p(\mathbf{R}, \mathbf{N})$ are different sets, and the definition of equivalence (which required the metrics to be defined on a single set) does not apply.

Recall from (1.1.31) that $l^1(\mathbf{R}, \mathbf{N}) \subset l^p(\mathbf{R}, \mathbf{N}) \subset l^\infty(\mathbf{R}, \mathbf{N})$ for all $1 < p < \infty$. Set $M = l^1(\mathbf{R}, \mathbf{N})$. Every sequence $\mathbf{x} \in M$ belongs to all other $l^p(\mathbf{R}, \mathbf{N})$. We can then equip M with different l^p metrics, and compare them as we did in $\mathbf{R}^n$.

What is the result? The proof of Proposition 1.2.9 still shows that every ball in $l^\infty(\mathbf{R}, \mathbf{N})$ contains an $l^p(\mathbf{R}, \mathbf{N})$-ball (with the same center). (What does this say about convergence?) The proof fails to show the converse. This does *not* mean that $l^p(\mathbf{R}, \mathbf{N})$-balls do not contain $l^\infty(\mathbf{R}, \mathbf{N})$-balls, or that the metrics are not equivalent: it only shows that the particular proof strategy fails.

To prove that the metrics are really not equivalent, one must *exhibit* an $l^p(\mathbf{R}, \mathbf{N})$-ball that does not contain any concentric $l^\infty(\mathbf{R}, \mathbf{N})$-ball, no matter how small. I claim that the $l^p(\mathbf{R}, \mathbf{N})$ unit ball, centered at the origin $0 = (0, 0, 0, \ldots)$, has this property. Pick any $\delta > 0$, set $\eta = \delta/2$, and let N be an integer $> 1/\eta^p$. Set

$$\mathbf{z} = (\underbrace{\eta, \ldots, \eta}_{N \text{ times}}, 0, 0, 0, \ldots).$$

Then $\|\mathbf{z} - 0\|_\infty = \eta < \delta$, while $\|\mathbf{z} - 0\|_p^p = N \times \eta^p > 1$. Thus, there can be no δ such that, with $\epsilon = 1$,

$$d_\infty(\mathbf{x}, 0) < \delta \quad \Longrightarrow \quad d_p(\mathbf{x}, 0) < \epsilon$$

holds for all $\mathbf{x}$.

The situation here is precisely as sketched in Figure 9. Balls in d_p fit into balls in d_∞, but not vice versa. Intuition is hard to come by in infinite dimensions, and my picture of the center of the "ball" on the circumference may be as revealing as a drawing can be. ∎

Remark. Equivalence of metrics will reappear in a perhaps more convincing form in Section 3, in connection with topological spaces. In topology, one does not try to assign a number to the "distance between points." One only knows (roughly speaking) when two points are "close" to each other. Equivalence then means, quite naturally, that two apparently different concepts of "closeness" are in fact identical. ∎

• **1.2.15 Exercise.** Let d be the l^1-metric on $\mathbf{R}^2$ and let d' be the l^2 metric on $\mathbf{R}^2$. We know, from Proposition 1.2.9, that they are equivalent, and that therefore the two conditions in Definition 1.2.2 are satisfied. Given ϵ, find explicit values for δ and δ'. How large can δ and δ' be (in terms of ϵ) before the conditions are violated? ∎

1.2.16 Exercise. Fix $\epsilon > 0$. Find the largest δ such that (for $\mathbf{x} \in \mathbf{R}^n$) $\|\mathbf{x} - 0\|_2 < \delta$ guarantees that $\|\mathbf{x} - 0\|_1 < \epsilon$. This δ will depend on n(as well as on ϵ). What happens as $n \to \infty$? What is the largest δ' such that $\|\mathbf{x} - 0\|_1 < \delta'$ guarantees that $\|\mathbf{x} - 0\|_2 < \epsilon$? What happens as $n \to \infty$? In the limit $n \to \infty$, the *sets* $l^1(\mathbf{R}, \mathbf{N})$ and $l^2(\mathbf{R}, \mathbf{N})$ are certainly different. But, are the metrics equivalent when they are used on sequences that belong to $l^1(\mathbf{R}, \mathbf{N})$? ∎

• **1.2.17 Exercise.** Find, if possible, a sequence $x^{(n)}$ in $l^1(\mathbf{R}, \mathbf{N})$ that converges to 0 in the l^1 metric but not in the l^∞ metric. Find, if possible, a sequence $x^{(n)}$ in $l^1(\mathbf{R}, \mathbf{N})$ that converges to 0 in the l^∞ metric but not in the l^1 metric. Review the relation of your example(s) to the concept of equivalence of metrics. ∎

• **1.2.18 Exercise.** Same as the preceding exercise, but replace l^∞ by l^2. ∎

• **1.2.19 Exercise.** The following statement should seem reasonable: $\|x^{(n)} - 0\|_2 \to 0$ (in $\mathbf{R}^2$, as $n \to \infty$), if and only if $|x_1^{(n)}| + |x_2^{(n)}| \to 0$ as $n \to \infty$. Explain what this has to do with equivalence of l^1 and l^2 metrics. ∎

• **1.2.20 Exercise.** (A review question, to be done without looking at the text.) Let d and d' be two metrics on a set M. We say that d' is *coarser than* d if every ball $\{y \mid d'(x, y) < \epsilon\}$ contains a ball $\{y \mid d(x, y) < \delta\}$. (It may no longer be the case that every d ball contains a d' ball). Figure out which of the following two statements is then true:

$$\text{if } d(x_n, x) \to 0 \text{ then } d'(x_n, x) \to 0,$$
$$\text{if } d'(x_n, x) \to 0 \text{ then } d(x_n, x) \to 0,$$

and prove it. ∎

1.2.21 Exercise. Show that the l^2 metric on $\mathbf{R}^2$ is coarser than the discrete metric (see Exercise 1.0.10). ∎

1.2.22 Exercise. Which metric on $l^1(\mathbf{R}, \mathbf{N})$ is coarser: the l^1 metric or the l^p metric with $1 < p \leq \infty$? ∎

• **1.2.23 Exercise.** (Refer to Definition 1.2.5 of convergence.) Give an ϵ-δ definition of the following statement: the sequence x_n *does not* converge to the point x. ∎

Subsection 1.3 Hölder and Minkowski Inequalities

We have now seen many examples of metric spaces, and in all of them, the first two requirements (1.0.5), (1.0.6) on $d(\cdot, \cdot)$ were obviously satisfied. The triangle inequality (1.0.7) is, as I have said before, much more difficult.

The goal in this subsection is to prove Proposition 1.1.27, whose main assertion is that what we have been calling "the l^p metric on $l^p(\mathbf{R}, \mathbf{N})$" satisfies the triangle inequality and therefore deserves the name "metric". The proof, which takes several pages, is divided into two lemmas of independent importance; each of them warrants separate discussion. First, some general comments.

1.3.1 Remarks

A. It was noted in Remark 1.1.28.B that $\mathbf{R}^n$ is a subset of $l^p(\mathbf{R}, \mathbf{N})$. The triangle inequality for the (hoped-for) l^p metric on $\mathbf{R}^n$ follows from the triangle inequality for $l^p(\mathbf{R}, \mathbf{N})$.

B. It will become evident that all parts of the proof of the triangle inequality adapt to $l^p(\mathbf{R}, \mathbf{Z})$, $l^p(\mathbf{C}, \mathbf{N})$, and $l^p(\mathbf{C}, \mathbf{Z})$ with almost no changes.

C. The triangle inequality is easy for $p = 1.\infty$. The proof for these values of p is presented first. It serves as a warmup, but more important, the general argument does not work well with l^∞, because the formula for that metric does not fit the pattern that holds for $1 < p < \infty$. We will see presently that l^1 is also somewhat exceptional.

First, let $p = \infty$:

$$\|\mathbf{x} - \mathbf{z}\|_\infty \stackrel{\text{def}}{=} \sup_{j\geq 0} |x_j - z_j| \leq \sup_{j\geq 0}(|x_j - y_j| + |y_j - z_j|) \leq$$
$$\sup_{j\geq 0} |x_j - y_j| + \sup_{j\geq 0} |y_j - z_j| = \|\mathbf{x} - \mathbf{y}\|_\infty + \|\mathbf{y} - \mathbf{z}\|_\infty.$$

Next, let $p = 1$:

$$\begin{aligned}\|\mathbf{x} - \mathbf{z}\|_1 &\stackrel{\text{def}}{=} \sum_{j=0}^{\infty} |x_j - z_j| \\ &\leq \sum_{j=0}^{\infty} (|x_j - y_j| + |y_j - z_j|) \\ &= \sum_{j=0}^{\infty} |x_j - y_j| + \sum_{j=0}^{\infty} |y_j - z_j| \stackrel{\text{def}}{=} \|\mathbf{x} - \mathbf{y}\|_1 + \|\mathbf{y} - \mathbf{z}\|_1.\end{aligned}$$

Hence, we need only prove the triangle inequality for $1 < p < \infty$. ∎

To obtain the triangle inequality, it will be sufficient (and convenient) to prove a related statement which is simpler to write down and work with.

1.3.2 Minkowski's inequality[4]. *Let* $\mathbf{x}, \mathbf{y} \in l^p(\mathbf{R}, \mathbf{N})$, *with* $1 \leq p < \infty$. *Then*

$$\sqrt[p]{\sum_{j=0}^{\infty} |x_j + y_j|^p} \leq \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} + \sqrt[p]{\sum_{j=0}^{\infty} |y_j|^p}. \quad (1.3.3)$$ ∎

Once this is established, we get the usual triangle inequality:

1.3.4 Corollary. *The triangle inequality holds for the* l^p *metrics,* $1 \leq p \leq \infty$. ∎

Proof of Corollary. The case $p = \infty$ has already been done in Remark 1.3.1.C. For all other cases, simply replace x_j by $x_j - y_j$ and y_j by $y_j - z_j$ in (1.3.3) to get the usual triangle inequality. ∎

---

[4] After Hermann Minkowski, German mathematician, 1864-1909.

1.3.5 Remark. Minkowski's inequality is at least as important as the triangle inequality, as will be seen in the next section. The quantity

$$\sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} \tag{1.3.6}$$

will there be interpreted as a measure of length of the vector $\mathbf{x} = (x_0, x_1, x_2, \ldots)$. It generalizes the concept of vector length one meets in calculus,

$$|\mathbf{v}| = \sqrt{v_1^2 + v_2^2 + v_3^2}, \tag{1.3.7}$$

and is just as basic for linear algebra and vector calculus in infinite dimensions as the familiar definition is to computations in $\mathbf{R}^2$ and $\mathbf{R}^3$. One uses the symbol $\|\mathbf{x}\|_p$ as abbreviation for (1.3.6), and calls it the *p-norm of* $\mathbf{x}$. (The standard vector length is the 2-norm of $\mathbf{v}$.) Minkowski's inequality can now be written very compactly:

$$\|\mathbf{x} + \mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p. \tag{1.3.8}$$

Again, there is an extension to $p = \infty$ — one simply defines

$$\|\mathbf{x}\|_\infty = \sup_{0 \leq j < \infty} |x_j|.$$

Minkowski's inequality for $p = \infty$ is proved just as in Remark 1.3.1.C. ∎

The proof of Minkowski's inequality is based on another important inequality; I will state it, discuss it, and then prove it.

1.3.9 Hölder's inequality[5]. *Let* $\mathbf{x} \in l^p(\mathbf{R}, \mathbf{N})$ *and* $\mathbf{y} \in l^q(\mathbf{R}, \mathbf{N})$, *and require that* $1/p + 1/q = 1$ $(1 < p, q < \infty)$. *Such* p, q *are said to be* ***conjugate****. Then*

$$\sum_{j=0}^{\infty} |x_j y_j| \leq \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} \sqrt[q]{\sum_{j=0}^{\infty} |y_j|^q}. \tag{1.3.10}$$

The two sides are equal only if the sequences $\{|x_j|\}$ *and* $\{|y_j|\}$ *are proportional, or if at least one of* $\mathbf{x}$ *and* $\mathbf{y}$ *is zero.*

A special case of Hölder's inequality has a special name:

[5] After Otto Hölder, German analyst, 1859-1937.

1.3.11 Terminology. *Hölder's inequality for $p = q = 2$,*

$$\sum_{j=0}^{\infty} |x_j y_j| \le \sqrt{\sum_{j=0}^{\infty} |x_j|^2} \sqrt{\sum_{j=0}^{\infty} |y_j|^2}$$

is called the Cauchy - Schwarz inequality. ∎

I will use the name "Cauchy-Schwarz inequality" when I feel like it. When I don't feel like typing the longer name, I will use "Hölder's inequality", even though $p = q = 2$. The only accepted rule seems to be to use "Cauchy-Schwarz" when one is dealing exclusively with l^2 spaces.

Hölder's inequality is probably not familiar from vector analysis (or anything else). Therefore, I want to spend time on special cases and applications, to make it less intimidating, before embarking on the proof.

1.3.12 Examples

A. Let x and y be as in the statement of the inequality. *Temporarily,* introduce the notation

$$\mathbf{x} \cdot \mathbf{y} = \sum_{j=0}^{\infty} x_j y_j;$$

it is supposed to remind you of the dot product of vectors. Recall also the norms $\| \cdot \|_p$ defined in (1.3.6). Since

$$\left| \sum_{j=0}^{\infty} x_j y_j \right| \le \sum_{j=0}^{\infty} |x_j y_j|,$$

Hölder's inequality implies

$$|\mathbf{x} \cdot \mathbf{y}| \le \|\mathbf{x}\|_p \, \|\mathbf{y}\|_q. \tag{1.3.13}$$

Now set $p = q = 2$ in (1.3.10), and specialize to $\mathbf{R}^2$ by taking $x_j = y_j = 0$ for $j \ge 2$. Then (1.3.13) becomes a two-dimensional version of the Cauchy-Schwarz inequality (1.3.11),

$$|x_0 y_0 + x_1 y_1| \le \|x\|_2 \, \|y\|_2 \tag{1.3.14}$$

where $\|x\|_2$ denotes the usual Euclidean length. Remember from vector calculus that (in $\mathbf{R}^2$)

$$\mathbf{x} \cdot \mathbf{y} = \|x\|_2 \, \|y\|_2 \cos\theta. \tag{1.3.15}$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$. Since $|\cos\theta| \leq 1$, (1.3.14) follows from the standard identity (1.3.15). From this point of view, Hölder's inequality is a generalization of the dot-product formula to higher dimensions and more general p's and q's.

B. Let $\mathbf{x} \in \mathbf{R}^n$, and let $\mathbf{y} \in \mathbf{R}^n$ be the vector $(1, 1, \ldots, 1)$. By the Cauchy-Schwarz inequality,

$$\sum_{j=1}^{n} |x_j| \leq \sqrt{\sum_{j=1}^{n} 1} \sqrt{\sum_{j=1}^{n} |x_j|^2}.$$

We can write this in terms of the l^1- and l^2-distances of $\mathbf{x}$ from the origin 0:

$$d_1(\mathbf{x}, 0) \leq \sqrt{n}\, d_2(\mathbf{x}, 0). \tag{1.3.16}$$

It follows from (1.3.16) that if $d_2(\mathbf{x}, 0) < \epsilon/\sqrt{n}$, then $d_1(\mathbf{x}, 0) < \epsilon$. This is one half of the estimate needed to prove equivalence of the l^1 and l^2 metrics (compare Proposition 1.2.9 and Exercise 1.2.16); it is obtained here by a simple application of (a special case of) Hölder's inequality.

C. Note that Hölder's inequality is also true for $p = 1$, $q = \infty$, but the $\sqrt{\ }$ is replaced, as usual, by a sup:

$$\sum_{j=0}^{\infty} |x_j y_j| \leq \sum_{j=0}^{\infty} (\sup_{j \geq 0} |y_j|)\, |x_j| = (\sup_{j \geq 0} |y_j|) \cdot \sum_{j=0}^{\infty} |x_j| = \|\mathbf{y}\|_\infty \|\mathbf{x}\|_1.$$

D. Hölder's inequality is often used to prove convergence of sequences. Here is a sample (other examples are given as exercises).

Claim. If $\mathbf{x} = (x_0, x_1, \ldots) \in l^2(\mathbf{R}, \mathbf{N})$, then the series

$$\sum_{j=0}^{\infty} \frac{x_j}{j+1}$$

converges.

To prove this, define $\mathbf{y} \in l^2(\mathbf{R}, \mathbf{N})$ by $y_j = 1/(j+1)$. Then by Hölder's inequality,

$$\sum_{j=0}^{\infty} \frac{|x_j|}{j+1} \leq \sqrt{\sum_{j=0}^{\infty} |x_j|^2} \sqrt{\sum_{j=0}^{\infty} \left(\frac{1}{j+1}\right)^2} < \infty.$$

One could show convergence in other ways, but none are more efficient. Actually, the hypothesis on $\mathbf{x}$ can be made to look weaker:

Claim. If $\sum_{j=0}^{\infty} x_j$ converges absolutely, then the series

$$\sum_{j=0}^{\infty} \frac{x_j}{j+1}$$

also converges absolutely.

This follows from the preceding claim, because if $\sum_{j=0}^{\infty} x_j$ converges absolutely, it belongs to $l^1(\mathbf{R}, \mathbf{N})$, which is a subset of $l^2(\mathbf{R}, \mathbf{N})$ by (1.1.31).

E. Example D can be generalized as follows. Hölder's inequality (or Cauchy-Schwarz, if you prefer) states that if $\mathbf{x}, \mathbf{y} \in l^2(\mathbf{R}, \mathbf{N})$, then the sequence $(x_j y_j)$ is in l^1 (it is *absolutely summable*). ∎

Proof of Hölder's inequality. The inequality is restated for your convenience:

$$\sum_{j=0}^{\infty} |x_j y_j| \leq \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} \sqrt[q]{\sum_{j=0}^{\infty} |y_j|^q}. \tag{1.3.17}$$

This assertion will be proved in detail, but the identification of the cases in which "$\leq$" becomes "$=$" is left for Exercise 1.3.41.

Reduction: I claim that it suffices to prove inequality (1.3.17) in a special case: Suppose that two sequences $\mathbf{x}' \in l^p(\mathbf{R}, \mathbf{N})$, $\mathbf{y}' \in l^q(\mathbf{R}, \mathbf{N})$ satisfy

$$\sum_{j=0}^{\infty} |x'_j|^p = 1. \quad \sum_{j=0}^{\infty} |y'_j|^q = 1. \tag{1.3.18}$$

For them, Hölder's inequality (1.3.17) takes the form

$$\sum_{j=0}^{\infty} |x'_j y'_j| \leq 1. \tag{1.3.19}$$

Now assume that (1.3.19) is true whenever the restrictions (1.3.18) hold. I show that the general inequality (1.3.17) is then necessarily true as well.

Let the sequences $(x_0, x_1, \ldots) \in l^p(\mathbf{R}, \mathbf{N})$ and $(y_0, y_1, \ldots) \in l^q(\mathbf{R}, \mathbf{N})$ be given. Set

$$\alpha = \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p}, \quad \beta = \sqrt[q]{\sum_{j=0}^{\infty} |y_j|^q}$$

and define $x'_j = x_j/\alpha$. $y'_j = y_j/\beta$. Then

$$\sum_{j=0}^{\infty} |x'_j|^p = \sum_{j=0}^{\infty} \frac{|x_j|^p}{\alpha^p} = \frac{1}{\alpha^p} \sum_{j=0}^{\infty} |x_j|^p = 1,$$

the last equality holding by definition of α. Likewise, $\sum_{j=0}^{\infty} |y'_j|^q = 1$. Thus, assumption (1.3.18) is satisfied, and according to (1.3.19) —assumed to be true— we now have $\sum_{j=0}^{\infty} |x'_j y'_j| \leq 1$. This, when translated back to the x_j and y_j, gives

$$\sum_{j=0}^{\infty} |x'_j y'_j| = \sum_{j=0}^{\infty} \left|\frac{x_j}{\alpha} \cdot \frac{y_j}{\beta}\right| = \frac{\sum_{j=0}^{\infty} |x_j y_j|}{\alpha\beta} \leq 1.$$

or

$$\sum_{j=0}^{\infty} |x_j y_j| \leq \alpha\beta = \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} \sqrt[q]{\sum_{j=0}^{\infty} |y_j|^q}.$$

This is Hölder's inequality (1.3.17). To prove it in all generality, then, we need only verify the special case (1.3.19) subject to the restriction (1.3.18). This is done next.

Let us drop the primes for simplicity of notation. We will prove that

$$\sum_{j=0}^{\infty} |x_j y_j| \leq 1$$

provided that

$$\sum_{j=0}^{\infty} |x_j|^p = \sum_{j=0}^{\infty} |y_j|^q = 1.$$

Again, there is a clever insight:[6]

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad , \quad a, b > 0. \tag{1.3.20}$$

Assuming (1.3.20) to be true, we find

$$\sum_{j=0}^{\infty} |x_j y_j| \leq \sum_{j=0}^{\infty} \frac{|x_j|^p}{p} + \frac{|y_j|^q}{q} = \frac{1}{p} \underbrace{\sum_{j=0}^{\infty} |x_j|^p}_{=1} + \frac{1}{q} \underbrace{\sum_{j=0}^{\infty} |y_j|^q}_{=1} = \frac{1}{p} + \frac{1}{q} = 1,$$

which is the inequality we needed to prove.

It remains to derive (1.3.20), which involves only positive real numbers. To this end, we consider the curve $t = s^{p-1}$, and certain areas shown in Figure 12. The area A_1 is $\int_0^a s^{p-1} ds = a^p/p$; the area A_2 is $\int_0^b t^{q-1} dt = b^q/q$. The relation $1/p + 1/q = 1$ is used here.

---

[6] A version of *Young's inequality*, after W. H. Young (1863-1942). See Exercises 1.3.38, 1.3.39, 1.3.42 for more information.

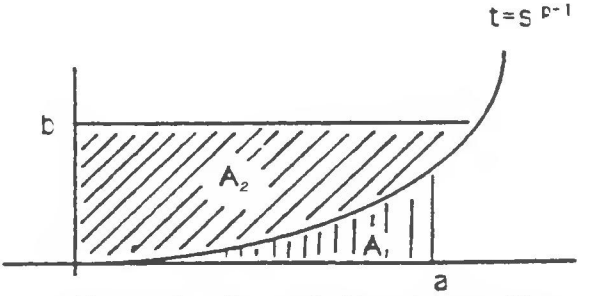


Figure 12. *Setting for Young's inequality*

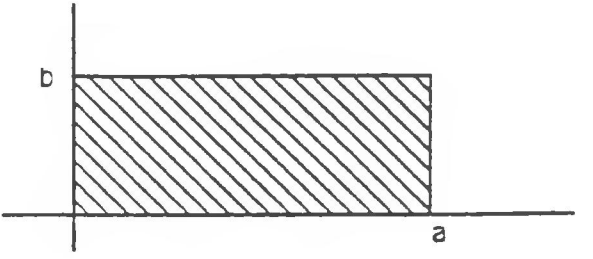


Figure 13. *Punch line of Young's inequality*

Since $A_1 + A_2 \geq ab$ = area of the rectangle in Figure 13, we get

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

as desired. *This concludes the proof of Hölder's inequality.*

Good Advice: When reading proofs like the one just presented you should *always* ask: how could I have found the argument myself? Often, the answer will be: no way. To know how to use $1/p + 1/q = 1$, or to realize that $ab \leq a^p/p + b^q/q$, is not easy — those insights come after weeks or months of experimentation (if ever). It is all the more important, then, to understand the basic strategies and inequalities and make them part of your mathematical tool kit. We can't all invent the wheel, but we can at least learn how to use it. ∎

Proof of Minkowski's inequality. For two numbers a, b, we have the identity

$$\begin{aligned}(|a| + |b|)^p &= (|a| + |b|)^{p-1}(|a| + |b|) \\ &= (|a| + |b|)^{p-1}|a| + (|a| + |b|)^{p-1}|b|.\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{j=0}^{\infty}(|x_j| + |y_j|)^p \leq &\sum_{j=0}^{\infty}(|x_j| + |y_j|)^{p-1}|x_j| \\ &+ \sum_{j=0}^{\infty}(|x_j| + |y_j|)^{p-1}|y_j|.\end{aligned} \tag{1.3.21}$$

We will estimate the two series on the right side by a combination of Hölder's inequality and the argument in Remark 1.1.29.

Since $\mathbf{x}, \mathbf{y} \in l^p(\mathbf{R}, \mathbf{N})$,

$$\sum_{j=0}^{\infty} |x_j|^p < \infty, \quad \text{and} \quad \sum_{j=0}^{\infty} |y_j|^p < \infty.$$

Next, because

$$|x_j| + |y_j| \le 2 \times \max\{|x_j|, |y_j|\},$$

we have

$$(|x_j| + |y_j|)^p \le 2^p \times \max\{|x_j|^p, |y_j|^p\} \le 2^p \times (|x_j|^p + |y_j|^p).$$

Therefore,

$$\sum_{j=0}^{\infty} (|x_j| + |y_j|)^p \le 2^p \sum_{j=0}^{\infty} (|x_j|^p + |y_j|^p) < \infty. \tag{1.3.22}$$

The relation $1/p + 1/q = 1$ can be written in the form $(p-1)q = p$, and so (1.3.22) implies that

$$\sum_{j=0}^{\infty} \left[(|x_j| + |y_j|)^{p-1} \right]^q < \infty,$$

i.e., the sequence $(|x_j| + |y_j|)^{p-1}$ belongs to $l^q(\mathbf{R}, \mathbf{N})$.

We now apply Hölder's inequality to the sequences $(|x_j| + |y_j|)^{p-1}$ in l^q and $|x_j|$ in l^p:

$$\begin{aligned} \sum_{j=0}^{\infty} (|x_j| + |y_j|)^{p-1} |x_j| &\le \sqrt[q]{\sum_{j=0}^{\infty} [(|x_j| + |y_j|)^{p-1}]^q} \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} \\ &= \sqrt[q]{\sum_{j=0}^{\infty} (|x_j| + |y_j|)^p} \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p}. \end{aligned} \tag{1.3.23}$$

The two series on the right side converge, hence so does the series on the left. This concludes the examination of the first term on the right side in (1.3.21). The second term on the right side of (1.3.21) is estimated in the same way, with x_j and y_j interchanged. We add the resulting inequality to (1.3.23), and then find from (1.3.21) that

$$\sum_{j=0}^{\infty} (|x_j| + |y_j|)^p \le \sqrt[q]{\sum_{j=0}^{\infty} (|x_j| + |y_j|)^p} \left[\sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} + \sqrt[p]{\sum_{j=0}^{\infty} |y_j|^p} \right]. \tag{1.3.24}$$

Call the sum on the left Σ. The inequality (1.3.24) has the form

$$\Sigma \le \Sigma^{1/q} \left[\sqrt[p]{\sum |x_j|^p} + \sqrt[p]{\sum |y_j|^p} \right].$$

Now divide both sides by Σ. On the left, we have

$$\frac{\Sigma}{\Sigma^{1/q}} = \Sigma^{1-(1/q)} = \Sigma^{1/p},$$

and so

$$\sqrt[p]{\sum_{j=0}^{\infty} (|x_j| + |y_j|)^p} \le \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} + \sqrt[p]{\sum_{j=0}^{\infty} |y_j|^p}.$$

The proof of Minkowski's inequality is complete, and with it, the proof of the triangle inequality. ∎

Remark. The proof may seem to require unfair trickery. Not so. The interplay between conjugate p and q (i.e., p and q satisfying $1/p + 1/q = 1$) is basic in analysis, and it took many years for people to realize that *it is* basic. Nowadays, one knows to look for these $p - q$ relations when dealing with $l^p(\mathbf{R}, \mathbf{N})$, but that does not make the ideas any more elementary. ∎

• **1.3.25 Exercise.** State Hölder's and Minkowski's inequalities for $l^p(\mathbf{R}, \mathbf{Z})$, $l^p(\mathbf{C}, \mathbf{N})$, and for $l^p(\mathbf{C}, \mathbf{Z})$. ∎

• **1.3.26 Exercise.** Let $1 \le p \le 2$, and pick $\mathbf{x} \in l^p(\mathbf{R}, \mathbf{N})$. Prove that the sequence $(x_0^2, x_1^2, \ldots)$ belongs to $l^1(\mathbf{R}, \mathbf{N})$. Is the conclusion still true if $2 < p < \infty$? ∎

• **1.3.27 Exercise.** Let $\mathbf{x} \in l^q(\mathbf{R}, \mathbf{N})$. Show that the sequence $(|x_0|^{q-1}, |x_1|^{q-1}, \ldots)$ belongs to $l^p(\mathbf{R}, \mathbf{N})$, where $1/p + 1/q = 1$. (Be sure to try a specific value of q, say $q = 4$.) ∎

1.3.28 Exercise. Let $\mathbf{x} = (x_j) \in l^2(\mathbf{R}, \mathbf{N})$. Show that the sequence $(x_j/(j+1)^\alpha)$ lies in $l^1(\mathbf{R}, \mathbf{N})$ for $\alpha > 1/2$, and find an example of a sequence (x_j) for which $(x_j/\sqrt{j+1})$ does not lie in $l^1(\mathbf{R}, \mathbf{N})$. ∎

• **1.3.29 Exercise.** Let $x_j > 0$, $j = 1, \ldots, n$. Show that

$$\left(\sum_{j=1}^{n} x_j\right)\left(\sum_{j=1}^{n} \frac{1}{x_j}\right) \ge n^2. \text{ ∎}$$

1.3.30 Exercise. Let $\mathbf{A} = (a_{ij})$ be a real $n \times n$ matrix. Let $\mathbf{x} \in \mathbf{R}^n$, and let $\mathbf{y} = \mathbf{Ax}$. Show that

(*) $$\|\mathbf{y}\|_2 \le C \|\mathbf{x}\|_2,$$

where $\|\cdot\|_2$ is the usual Euclidean vector length, and C is the number

$$\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}.$$

(Hint: write

$$y_i = \sum_{j=1}^{n} a_{ij}x_j;$$

use the Cauchy-Schwarz inequality, square both sides, and sum over i.) ∎

1.3.31 Exercise. Inequality (*) in the preceding exercise can be used to prove that matrix multiplication is continuous: if $\mathbf{x}^{(n)} \to \mathbf{x}$ in the usual Euclidean sense, then $\mathbf{A}\mathbf{x}^{(n)} \to \mathbf{A}\mathbf{x}$. Give the details. ∎

1.3.32 Exercise. Inequality (*) in Exercise 1.3.30 remains true even if $\mathbf{x}$ belongs to $l^2(\mathbf{R}, \mathbf{N})$, as long as the *infinite* matrix $\mathbf{A}$ satisfies

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}|a_{ij}|^2 < \infty.$$

If this is true, then (*) shows that $\mathbf{y}$ also belongs to $l^2(\mathbf{R}, \mathbf{N})$, and that multiplication by this infinite matrix is again continuous. Write out the details. (Infinite matrices first appeared in mathematics in Hilbert's study of integral equations.) ∎

1.3.33 Exercise*. Exercise 1.3.30 can be generalized. Let $\mathbf{A} = (a_{ij})$ be a real $n \times n$ matrix. Let $\mathbf{x} \in \mathbf{R}^n$, and let $\mathbf{y} = \mathbf{A}\mathbf{x}$. Set

$$M = \sup_{1\le i\le n}\sum_{j=1}^{n}|a_{ij}|$$

and

$$N = \sup_{1\le j\le n}\sum_{i=1}^{n}|a_{ij}|.$$

(M and N are called the *maximum row sum of* $\mathbf{A}$ and the *maximum column sum of* $\mathbf{A}$, respectively.) Let $1/p + 1/q = 1, 1 < p, q < \infty$. Show that

$$\|\mathbf{y}\|_q \le M^{1/p}N^{1/q}\|\mathbf{x}\|_q.$$

Here $\|\mathbf{x}\|_q$ is the "l^q-length of $\mathbf{x}$",

$$\|\mathbf{x}\|_q \stackrel{\text{def}}{=} \sqrt[q]{\sum_{j=1}^{n}|x_j|^q}.$$

(Hint: write

$$|y_i| \le \sum_{j=1}^{n}|a_{ij}|^{1/p}|a_{ij}|^{1/q}|x_j|,$$

and use Hölder's inequality.) ∎

• **1.3.34 Exercise.** Let $\mathbf{x}, \mathbf{y}$ be two unit vectors in $\mathbf{R}^n$. Show that their dot product is always ≤ 1 in absolute value. Try to show this *without* using Hölder's inequality! ∎

• **1.3.35 Exercise.** Let $p = q = 2$. The proof of Hölder's inequality given above becomes easier to look at. Work through it with those special values of p and q. In particular, give an elementary proof of $ab \leq a^2/2 + b^2/2$. ∎

1.3.36 Exercise. Here is a still simpler proof of Hölder's inequality for $p = q = 2$ (the Cauchy - Schwarz inequality, see (1.3.11)). It is outlined for $\mathbf{R}^n$, but the idea works in more general settings.

Let $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ be fixed. For $\lambda \in \mathbf{R}$,

$$\sum_{j=1}^{n} |\lambda x_j + y_j|^2 = \lambda^2 \sum_{j=1}^{n} |x_j|^2 + 2\lambda \sum_{j=1}^{n} x_j y_j + \sum_{j=1}^{n} |y_j|^2.$$

Since the expression on the left is obviously ≥ 0, so is the expression on the right. Now, if $a\lambda^2 + b\lambda + c \geq 0$ for all real λ, then the quadratic equation $a\lambda^2 + b\lambda + c = 0$ cannot have distinct real roots. Hence $b^2 - 4ac \leq 0$. This inequality, expressed in terms of $\mathbf{x}$ and $\mathbf{y}$, is the Cauchy - Schwarz inequality. Fill in the details. ∎

1.3.37 Exercise. What happens in the Cauchy-Schwarz inequality (1.3.11) when $\mathbf{x} = \mathbf{y}$? Examine the proof in the preceding exercise, and show that the $\leq$ sign in the inequality becomes an $=$ sign if, and only if, the sequences $\{|x_j|\}$ and $\{|y_j|\}$ are proportional or if at least one of them is $\mathbf{0}$. ∎

1.3.38 Exercise. Fill in the following outline for a different proof of Young's inequality (1.3.20),

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}, \quad a, b > 0, \quad \frac{1}{p} + \frac{1}{q} = 1.$$

First, it suffices to prove that for $a, b > 0$ and $0 < \eta < 1$,

$$a^\eta b^{1-\eta} \leq \eta a + (1 - \eta) b.$$

To prove the last inequality, replace a by ab, cancel b, and use elementary calculus to show that

$$a^\eta - \eta a - (1 - \eta) \leq 0, \quad \text{for } a > 0,\ 0 < \eta < 1.$$

This proof of Young's inequality also shows that $ab = a^p/p + b^q/q$ if, and only if, $a = b$. ∎

1.3.39 Exercise. Here is a generalization of Young's inequality (for some people, it is *the* Young inequality). Let f be continuous and strictly increasing on $[0, \infty)$, with $f(0) = 0$. Let f^{-1} be the inverse function. Then for $a, b > 0$,

$$ab \leq \int_0^a f(t)\, dt + \int_0^b f^{-1}(t)\, dt, \tag{1.3.40}$$

and equality holds if, and only if, $b = f(a)$.

For $f(t) = t^{p-1}$, (1.3.40) becomes Young's inequality, version (1.3.20). The proof given in the text follows this path. Adapt that pictorial proof to explain why (1.3.40) should be true. ∎

1.3.41 Exercise. Use the last line in Exercise 1.3.38 to show: if neither x nor y is zero, then Hölder's inequality becomes an equality only if the sequences $\{|x_j|^p\}$ and $\{|y_j|^q\}$ are proportional. ∎

1.3.42 Exercise*. Let $a_1, \ldots, a_n$ be positive numbers. The usual average

$$\frac{a_1 + \ldots + a_n}{n}$$

is also called the *arithmetic mean* of the a_j, and

$$\sqrt[n]{a_1 \cdots a_n}$$

is called their *geometric mean*. The "theorem of the arithmetic and geometric means" says:

(1.3.43) $$\sqrt[n]{a_1 \cdots a_n} \leq \frac{a_1 + \ldots + a_n}{n}.$$

with equality holding only if $a_1 = \ldots = a_n$.

(The geometric mean of a_1 and a_2 is the length of the side of a square whose area = the area of the rectangle with sides a_1 and a_2).

When $\eta = 1/2$, the inequality

$$a^{\eta} b^{1-\eta} \leq \eta a + (1 - \eta) b$$

from Exercise 1.3.38 reduces to (1.3.43) with $n = 2$. Prove the following generalization of the theorem of the arithmetic and geometric means:

Let the positive numbers η_j satisfy $\eta_1 + \ldots + \eta_n = 1$. Then

$$a_1^{\eta_1} \cdots a_n^{\eta_n} \leq \eta_1 a_1 + \ldots + \eta_n a_n,$$

with equality holding only if $a_1 = \ldots = a_n$. ∎

Subsection 1.4 Lebesgue Spaces

So far in this chapter, we have seen the gradual generalization of some elementary properties of vectors in $\mathbf{R}^2$. The notion of distance, the triangle inequality, and the dot product were extended, first to real and complex n-vectors, then to sequences, which may be thought of as vectors with countably many components. Now comes one more step: everything we have done so far is extended to vectors with a "continuum" of components. Those are just *functions*, and we arrive at a class of *function spaces* which play a big role in many areas of analysis.

It becomes more difficult, however, to give a complete and consistent description of the admissible functions and the metrics on sets of functions. The space $l^p(\mathbf{R}, \mathbf{N})$ was harder to understand than just $\mathbf{R}^n$ with its l^p metric, because issues involving convergence of series could not be overlooked. Thus, there were *limiting processes* that play no role in the definition of finite - dimensional vector spaces. When one replaces sequences by functions, another, and still more delicate, limiting process enters the picture: the series are replaced by integrals, and in addition to worrying about whether an integral converges one must first ask whether it even exists, in the sense of the Riemann sums from calculus or from a more general point of view.

The satisfactory resolution of the technical problems that arise requires the theory of measure and integration. In the present subsection, all measure - theoretic subtleties are ignored. Later in this chapter (Section 5), some elements of measure theory are presented, but nothing is said there about integration. In Chapter 2, a sketchy description of Lebesgue measure and integration will be provided, and the examples we introduce in the present subsection will not really be finished until then.

Throughout this subsection, the use of integrals will be kept at the level of introductory calculus. The reader need not remember any of the official definitions of integrals; experience in computing some of the standard integrals in one or more dimensions will suffice.

The transition from sequences to functions can be made plausible by appropriate change of notation. This does not help one to prove many results, but it may make them less intimidating.

Let us write a sequence $(x_0, x_1, x_2, \ldots) \in \mathbf{R}^{\mathbf{N}}$ in the following form:

$$(f(0), f(1), f(2), \ldots).$$

This is purely a matter of notation, $x_j \mapsto f(j)$. Let us also call the index of summation "t" rather than "j". An l^p metric then looks like

$$\|f - g\|_p = \sqrt[p]{\sum_{t=0}^{\infty} |f(t) - g(t)|^p}.$$

Now let the index t take on values not over $\{0, 1, 2, \ldots\}$ but over some interval $[a, b]$, $-\infty < a < b < \infty$. The sum becomes an integral (formally), and we have, for *functions*

$$f, g : [a, b] \to \mathbf{R}$$

the metric

$$\|f - g\|_p = \sqrt[p]{\int_a^b |f(t) - g(t)|^p \, dt}.$$

It turns out that a sensible concept of "integral" for completely arbitrary functions cannot be devised. We therefore restrict attention to the class of functions that is familiar from calculus.

1.4.1 Definition. *For $-\infty < a < b < \infty$, let $C([a, b])$ be the set of continuous functions on $[a, b]$. For $f, g \in C([a, b])$, set*

$$\|f - g\|_p = \sqrt[p]{\int_a^b |f(t) - g(t)|^p \, dt} \quad , \quad 1 \le p < \infty$$

$$\|f - g\|_\infty = \max_{a \le t \le b} |f(t) - g(t)|.$$

The definition of $\|\cdot - \cdot\|_\infty$ assumes the following result from advanced calculus: a continuous function on a closed bounded interval takes on its maximum value. This will be proved in Chapter 2, II.1.3.8.

As long as a function can be integrated, the metrics in (1.4.1) make sense. For example, one could allow functions that are not continuous, but only *piecewise continuous*, like the one in Figure 14. ("Piecewise continuous" means that the function is continuous except possibly at a finite number of points).

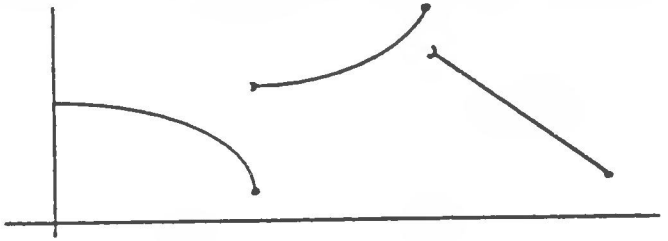


Figure 14. *A piecewise continuous function*

It is also possible to define these metrics on sets of piecewise continuous functions on intervals of the form $(-\infty, b]$, $[a, \infty)$, or $(-\infty, \infty)$. The integrals may then diverge, however, and one must restrict the definition to a suitable subclass of functions. Finally, one could also allow unbounded piecewise continuous functions on finite or infinite intervals, as in the following example.

1.4.2 Example. Let $f(t) = t^{-1/3}$ on $(0, 1]$. Or, let $g(t) = t^{-1/3}$ on $(0, 2)$ and $= t^{-2}$ on $[2, \infty)$. ∎

The advantage of Definition 1.4.1 is that the set $C([a, b])$ has a clean, unambiguous description. The disadvantage is that too many perfectly reasonable functions are excluded. On the other hand, the "Pseudo-Definition" to follow (so called to admit that it is vague even by applied mathematics standards) allows a greater variety of functions, but the description is messy. There is no point in trying for a definition without loopholes; people who want rigor should learn measure theory, which was invented precisely because of the gap between the classes of continuous and integrable functions.

1.4.3 Pseudo-Definition. *For any a, b, $(-\infty \le a < b \le +\infty)$ and $1 \le p < \infty$, define*

$$L_0^p([a, b]) = \text{ set of piecewise continuous functions for which } \int_a^b |f(t)|^p \, dt < \infty.$$

Define

$$L_0^\infty([a, b]) = \text{ set of bounded piecewise continuous functions.}$$

Convention. Whenever I write $[a, b]$, that closed interval can be replaced by a half-open interval $(a, b]$ or $[a, b)$, or by an open interval (a, b). ∎

Remarks

A. When $1 \le p < \infty$ in 1.4.3, f could be unbounded, and the integral, interpreted as an improper Riemann integral, could still be finite. To illustrate the point, I compute the integral

$$\int_a^b |g(t)|^p \, dt$$

for the function g in Example 1.4.2. It is

$$\int_0^2 (t^{-1/3})^p \, dt + \int_2^\infty (t^{-2})^p \, dt$$
$$= (1 - p/3)^{-1} t^{1-p/3} \Big|_0^2 + (1 - 2p)^{-1} t^{1-2p} \Big|_2^\infty.$$

The first term is finite at $t = 0$ as long as $1 - (p/3) > 0$ or $p < 3$, and the second term is finite at $t = \infty$ as long as $1 - 2p < 0$ or $p > 1/2$. Hence, the function g belongs to the spaces $L_0^p([0, \infty))$ for $1 \le p < 3$.

B. The subscript "0" is nonstandard notation, and is used because in all the literature, L^p without the zero refers to a more general set of functions. Those

more general spaces are defined by the use of measure theory, and are called *Lebesgue spaces* (Section II.2.6). When you see the symbol l^p in a book or paper, it always refers to sequence spaces; the symbol L^p usually refers to function spaces, but in very abstract discussions it covers sequence spaces as well. ∎

• **1.4.4 Exercise.** For each of the following functions f, find the set of $p \in [1, \infty]$ for which $f \in L_0^p$ on the given interval.

a) $f(t) = e^{-t}$ on $[0, \infty)$;
b) $f(t) = 1/t$ on $[1, \infty)$;
c) $f(t) = 1/\sqrt{t}$ on $(0, 1]$;
d) $f(t) = \ln t$ on $(0, 3]$;
e) $f(t) = 1/\sqrt{t}$ on $(0, \infty]$;
f) $f(t) = 1/\sqrt{|t|}$ on $0 < |t| \leq 1$, $= 0$ at $t = 0$, and $= 2/t^2$ on $|t| > 1$. ∎

• **1.4.5 Exercise.** For each of the following functions f, find the set of $\alpha \in \mathbf{R}$ for which $f \in L_0^p$ on the given interval, taking i) $p = 1$, ii) $p = 2$, iii) $p = \infty$.

a) $f(t) = t^\alpha$ on $(0, 1]$;
b) $f(t) = t^\alpha$ on $[1, \infty)$;
c) $f(t) = t^\alpha$ on $(0, \infty)$;
d) $f(t) = t^\alpha$ on $(0, 1]$, and $= t^{-\alpha}$ on $[1, \infty)$. ∎

• **1.4.6 Exercise.** Is there a function f on $(0, \infty)$ which belongs to L_0^1 but not to L_0^2? Is there one that belongs to L_0^2 but not to L_0^1? ∎

• **1.4.7 Exercise.** Same question, but the domain is $(0, 1)$. ∎

1.4.8 Definition. *The L^p - metric on the set $L_0^p([a, b])$ is given by*

$$\|f - g\|_p \stackrel{\text{def}}{=} \sqrt[p]{\int_a^b |f(t) - g(t)|^p \, dt} \quad , \quad \text{for } 1 \leq p < \infty \tag{1.4.9}$$

and by

$$\|f - g\|_\infty \stackrel{\text{def}}{=} \sup_{a \leq t \leq b} |f(t) - g(t)| \tag{1.4.10}$$

for $p = \infty$. ∎

Remarks

A. ***Hölder and Minkowski inequalities.*** The "metric" just defined is not a metric until the triangle inequality is established. The proof, in outline, works just as it did for the sequence spaces $l^p(\mathbf{R}, \mathbf{N})$. One proves Hölder's inequality, then Minkowski's inequality, and finally deduces the triangle inequality. The details will not be presented — there would be little point, since without the

official definition of the L^p spaces the argument would not be sufficiently general. The proof of these inequalities for L^p metrics on the space $C([a,b])$ of continuous functions on a finite interval uses only standard facts about integrals; it is left to the reader (see the exercises). In any case, I will make free use of all these inequalities when the time comes.

B. One should picture these metrics at least for $p = 1, 2, \infty$. In Figures 15 and 16, $[a,b] = [0,1]$.

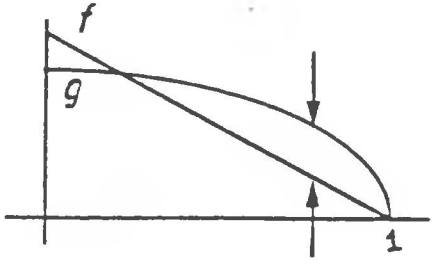


Figure 15. *Distance for* $p = \infty$

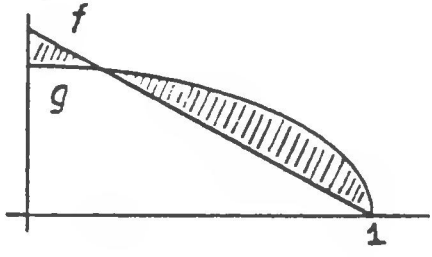


Figure 16. *Distance for* $p = 1$

$p = 2$ is probably the most useful metric in applications, but is harder to picture. You have to square the difference, $|f(t) - g(t)|$, which can expand or contract the area between the functions' graphs. Then take the total distorted area, and find its square root. I don't know how to draw this. You try it.

C. If the idea of "distance between functions" is new to you, it may surprise you to learn that the $\|\cdot - \cdot\|_2$ metric, which is not so easy to picture, is actually the most important. There are several reasons.

1) The metric is the natural generalization of Euclidean distance,

$$\sqrt{\sum_{j=1}^{n} |x_j - y_j|^2} \to \sqrt{\sum_{j=0}^{\infty} |x_j - y_j|^2} \to \sqrt{\int_a^b |f(t) - g(t)|^2 dt}$$

and it has several pleasant properties we have yet to discuss.

2) In numerical analysis, a function is represented by its values $f(t_1), \dots f(t_n)$ at a finite number of points. If the distance between functions is measured by the l^2 metric, $\| \cdot - \cdot \|_2$, then the distance between discrete approximations will naturally be measured by the usual Euclidean metric in $\mathbf{R}^n$. This is often convenient.

3) The integral

$$\int_a^b |f(t)|^2 dt \tag{1.4.11}$$

arises frequently in applications. If $f(t)$ is a "signal" or a "wave form", then (1.4.11) represents energy or power, depending on context. $\| \cdot - \cdot \|_2$ therefore compares two physical quantities in a natural way. In the engineering literature,

$$\sqrt{\int_a^b |f(t) - g(t)|^2 dt}$$

is called the "root mean square," or r.m.s., difference. There is a natural and simple relation between the r.m.s. difference of two functions and the l^2 distance between their Fourier coefficients (see (1.1.19)). ∎

The next set of exercises translates some of the key results about sequence spaces to functions spaces. Some things are unchanged, except for notation, while other results are no longer true. You should read the exercises, even if you do not work them; the ideas will come up again. Look for these two points, in particular:

Inequalities. Hölder's and Minkowski's inequalities remain true after the obvious replacement of sum by integral;

Inclusion relations. The pleasant inclusion relation $l^p(\mathbf{R}, \mathbf{N}) \subset l^q(\mathbf{R}, \mathbf{N})$ whenever $p < q$ (Proposition 1.1.31) is ***false*** for function spaces. On a finite interval, one has ***exactly the opposite inclusion,***

$$L_0^p([a,b]) \supset L_0^q([a,b]) \quad \text{for } p < q,$$

while on an infinite interval ***there are no guaranteed inclusions at all.***

1.4.12 Exercise. Hölder's inequality for $f \in L_0^p([a,b])$ and $g \in L_0^q([a,b])$ is

$$\int_a^b |f(t)g(t)|\, dt \le \sqrt[p]{\int_a^b |f(t)|^p\, dt}\ \sqrt[q]{\int_a^b |g(t)|^q\, dt}.$$

Prove it by following the argument in Subsection 1.3. Do not worry about improper integrals, discontinuities, etc. It is enough to work with continuous functions on a bounded interval. ∎

1.4.13 Exercise. Formulate and prove the integral analog of Minkowski's inequality. ∎

1.4.14 Exercise. State and prove the Cauchy - Schwarz inequality for L^2, following the outline in Exercise 1.3.36. ∎

• **1.4.15 Exercise.** Use Hölder's inequality in Exercise 1.4.12, applied to the functions f and $g \equiv 1$, to prove: if $[a, b]$ is a *finite* interval, and if $f \in L_0^p([a, b])$ for some $p \in (1, \infty)$, then $f \in L_0^1([a, b])$. (The result is not as trivial as it may seem, since the functions under consideration need not be bounded. For instance, take $f(t) = t^{-1/4}$ and $p = 2$). Show by constructing a counterexample that this may not be true if $[a, b]$ is an unbounded interval. ∎

1.4.16 Exercise*. Let M be the set of functions from the interval $[0, 1]$ to $\mathbf{R}^n$; an element of M is a *vector* of functions, $\mathbf{f} = (f_1, \ldots, f_n)$. Assume also that each component f_j is piecewise continuous. Let M^p be the subset consisting of functions f for which

$$\int_0^1 \sum_{j=1}^n |f_j(x)|^p \, dx < \infty.$$

a) Prove *Hölder's inequality*, which now takes the form

$$\int_0^1 \sum_{j=1}^n |f_j(x) g_j(x)| \, dx \leq \sqrt[p]{\int_0^1 \sum_{j=0}^n |f_j(x)|^p \, dx} \sqrt[q]{\int_0^1 \sum_{j=0}^n |g_j(x)|^q \, dx}$$

for $\mathbf{f} \in M^p$, $\mathbf{g} \in M^q$, and $1/p + 1/q = 1$. (Hint: The Hölder inequality in $\mathbf{R}^n$ gives

$$\sum_{j=1}^n |f_j(x) g_j(x)| \leq \sqrt[p]{\sum_{j=0}^n |f_j(x)|^p} \sqrt[q]{\sum_{j=0}^n |g_j(x)|^q}.$$

Abbreviate the two functions on the right side, setting $\sqrt[p]{\sum} = F$, $\sqrt[q]{\sum} = G$. We see that $F \in L^p([0, 1])$, $G \in L^q([0, 1])$. Now use the Hölder inequality for integrals to finish the proof.)

b) Formulate and prove the Minkowski inequality in the space M^p.

Incidentally, M^p is only a temporary notation. Those spaces are usually called $L^p([0, 1], \mathbf{R}^n)$, which is too messy to use in just one problem. ∎

• **1.4.17 Exercise.** Define $f_n(t) = t^n$, $t \in [0, 1]$, $n \in \mathbf{N}$. Let $g(t) \equiv 0$. Find $\|f_n - g\|_p$ for $p = 1, 2, \infty$. Find $\|f_n - f_m\|_p$ for $p = 1, 2, \infty$. Discuss. ∎

1.4.18 Exercise*. Let $1 \leq p \leq \infty$. Find an $f \in L_0^p([0, \infty))$ which belongs to no $L_0^q([0, \infty))$ for $q \neq p$. ∎

1.4.19 Exercise. Let $1 \le p < r < q \le \infty$. Show that if $f \in L_0^p([0,\infty)) \cap L_0^q([0,\infty))$, then also $f \in L_0^r([0,\infty))$. ∎

To promote familiarity with the L^p metrics, I will now compare different metrics in the spirit of the discussion of equivalence in Subsection 1.2. Recall the main points: the l^p metrics on $l^1(\mathrm{R},\mathrm{N})$ are "really different" in that convergence in l^p does not imply convergence in l^q when $p > q$, but does imply convergence when $p < q$. The situation with function spaces is worse: on unbounded intervals $[a,b]$, convergence in L^p is altogether unrelated to convergence in L^q.

1.4.20 Example. Consider the sequence of functions defined, for $n > 1$, by

$$f_n(t) = \begin{cases} 0, & \text{if } 0 \le t < 1/n, \\ 1/\sqrt{t}, & \text{if } 1/n \le t \le 1. \end{cases}$$

and let $g(t) \equiv 0$. $f_n(t)$ is drawn in Figure 17.

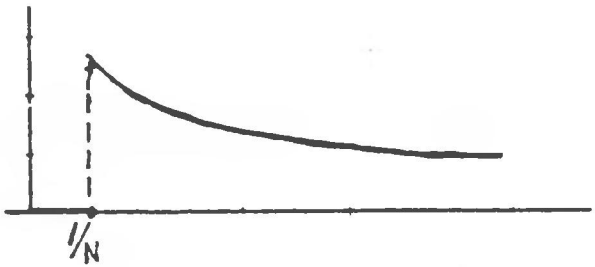


Figure 17. f_n

Now,

$$\|f_n - g\|_1 = \int_{1/n}^{1} \frac{1}{\sqrt{t}}\,dt = 2[1 - \frac{1}{\sqrt{n}}],$$

while

$$\|f_n - g\|_2 = \sqrt{\int_{1/n}^{1} (\frac{1}{\sqrt{t}})^2 dt} = \sqrt{\ln n}.$$

We can therefore have functions that stay at a finite distance from each other relative to the metric $\|\cdot - \cdot\|_1$ but go infinitely far apart in $\|\cdot - \cdot\|_2$. This should suggest that the metrics are not equivalent, and indeed they are not. You can look at this phenomenon in a slightly different way, as described in Proposition 1.2.3:

Set $h_n = f_n/\sqrt{\ln n}$ for $n > 1$, and write 0 for the zero function g. We then have

$$\|h_n - 0\|_1 \to 0 \text{ as } n \to \infty,$$

while

$$\|h_n - 0\|_2 = 1 \text{ for all } n.$$

Thus, the sequence h_n converges to the zero function in the L^1 metric, but it does not converge to the zero function in the L^2 metric. One of the two conditions for equivalence in Definition 1.2.2 is therefore violated. The situation is described by Figure 18, recalled from Subsection 1.2.

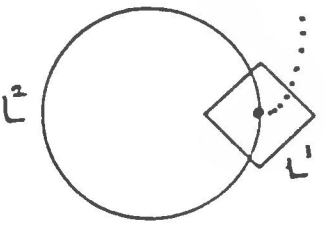


Figure 18. *L^1 convergence does not imply L^2 convergence*

On the other hand, convergence in L^2 does imply convergence in L^1. This fact depends a) on Hölder's inequality, and b) on the boundedness of the domain, $(0,1)$, of the functions in question. The argument goes as follows.

Let $f \in L_0^2((0,1))$, and set $g(t) \equiv 1$ for $t \in (0,1)$. Because $(0,1)$ is a finite interval,

$$\int_0^1 1^2\, dt < \infty,$$

and $g \in L_0^2$. Now apply Hölder's inequality with $p = q = 2$:

$$\int_0^1 |f(t)| \cdot 1\, dt \le \sqrt{\int_0^1 1^2\, dt}\sqrt{\int_0^1 |f(t)|^2\, dt} < \infty. \tag{1.4.21}$$

Thus, $f \in L_0^1$, but more is true. If you replace f by $g_n - g$, where $g_n, g \in L_0^2$, then (1.4.21) says

$$\|g_n - g\|_1 \le \|g_n - g\|_2.$$

Hence, if g_n converges to g in L_0^2 (meaning that $\|g_n - g\|_2 \to 0$ as $n \to \infty$), then also $\|g_n - g\|_1 \to 0$, and g_n converges to g in L_0^1. See Figure 19.

The argument for L^1 convergence fails if $[a,b]$ is infinite, for then

$$\int_a^b 1^2\, dt = \infty.$$

The next example takes up the case of infinite $[a,b]$. ∎

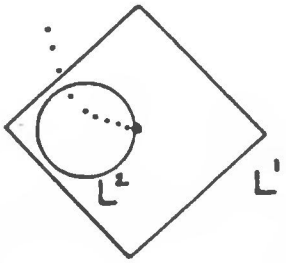


Figure 19. *L^2 convergence does imply L^1 convergence*

1.4.22 Example. Define a sequence of functions on $(0, \infty)$ as follows:

$$f_n(t) = \begin{cases} 1/n, & \text{if } n < t < 2n, \\ 0, & \text{otherwise.} \end{cases}$$

Compute:

$$\int_0^\infty |f_n(t)|\, dt = 1,$$

and

$$\int_0^\infty |f_n(t)|^2\, dt = 1/n.$$

Therefore, with 0 again denoting the zero function,

$$\|f_n - 0\|_2 \to 0 \quad \text{but} \quad \|f_n - 0\|_1 \not\to 0$$

as $n \to \infty$. Here is a sequence that converges in the L^2 metric, but not in the L^1 metric. Note again that this becomes possible when the interval of integration is infinite.

On this infinite interval, it is also easy to invent a sequence that converges in L^1 but not in L^2: just take the h_n from the preceding example, and think of them as functions on $(0, \infty)$ by declaring them to be zero for $t > 1$. ∎

1.4.23 SUPPLEMENT: Sobolev Spaces

To measure, simultaneously, the distance between two functions and their first k derivatives, one often uses the *Sobolev metric*[7]

$$\|f - g\|_{p,k} \stackrel{\text{def}}{=} \sqrt[p]{\int_a^b |f(t) - g(t)|^p + |f'(t) - g'(t)|^p + \cdots + |f^{(k)}(t) - g^{(k)}(t)|^p dt}.$$

[7] After S. L. Sobolev, Russian mathematician (1908-1989)

Of course, f and g must both have k derivatives (continuously differentiable, say), which excludes functions like $|t|$.

These metrics arise when one studies differential equations, for example

$$f''(t) = h(t) \quad \text{on } a \le t \le b, \quad f(a) = f(b) = 0.$$

(h is given, f is to be found). To compare the solutions f_1, f_2 corresponding to two different right-hand sides, h_1 and h_2, one may want to know: how close is f_1 to f_2 and f_1' to f_2' in terms of the "distance" between h_1 and h_2? For instance, the two functions in Figure 20 are probably "close" if you only compare the graphs according to $\|f_1 - f_2\|_\infty$ or according to $\|f_1 - f_2\|_2$, but the many little wiggles in f_2 might be of physical importance, and would be detected by comparing f_1' and f_2'.

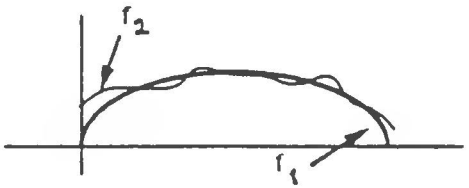


Figure 20. *What Sobolev norms measure*

1.4.24 Exercise. Prove that the Sobolev metrics satisfy the triangle inequality. (Hint: use Exercise 1.4.16.) ∎

1.4.25 SUPPLEMENT: Functions of Several Variables

So far, we have only considered functions of one variable. In higher - dimensional problems, one expects to meet functions of several variables. Typical problems include: wave propagation in space, fluid dynamics, tomography. There are analogues of the L^p spaces (and of Sobolev spaces, Supplement 1.4.23) in every (even infinite) number of dimensions. A new practical difficulty arises: it is much more difficult to find computable examples. There is also a theoretical problem: when one studies functions defined on a subset $D \subset \mathbf{R}^2$, say, the nature of the boundary of D plays a crucial role. If this boundary has corners, the technical difficulties may be more demanding than they would be for a smooth boundary. The second type of question is beyond the scope of the present discussion. The difficulties with explicit computations will become evident soon enough.

The basic formulas are easy. For functions on $D \subset \mathbf{R}^2$, for example, we define the L^p metric by

$$\sqrt[p]{\iint_D |f(x_1, x_2) - g(x_1, x_2)|^p \, dx_1 dx_2}.$$

The extension to higher - dimensional domains and to $p = \infty$ will be evident. It is more difficult than before to give even a Pseudo - Definition of the class of functions under consideration; therefore, I will not attempt to define anything.

•**1.4.26 Exercise.** Write Hölder's and Minkowski's inequalities for functions defined on a domain D in $\mathbf{R}^2$. ∎

•**1.4.27 Exercise.** Let $D = \{(x_1, x_2) \mid x_1^2 + x_2^2 \leq 1\}$. Let $f(x_1, x_2) = x_1 x_2 / (x_1^2 + x_2^2)^2$ for $(x_1, x_2) \neq (0, 0)$, and $= 0$ for $(x_1, x_2) = (0, 0)$. Is

$$\int\int_D |f(x_1, x_2)|\, dx_1 dx_2 < \infty? \text{ ∎}$$

1.4.28 Exercise*. Let $D = \{(x_1, x_2) \mid |x_1| + |x_2| > 1\}$. For what values of a and b is

$$\int\int_D \frac{1}{(|x_1|^a + |x_2|^b)^2}\, dx_1\, dx_2 < \infty? \text{ ∎}$$

• **1.4.29 Exercise.** The calculation of multiple integrals is sometimes facilitated by the use of spherical polar coordinates. The formulas for two and three dimensions are in every calculus text. In four dimensions, the transformation is described by:

$$\begin{aligned} x_1 &= r \sin\theta_1 \sin\theta_2 \sin\theta_3, \\ x_2 &= r \sin\theta_1 \sin\theta_2 \cos\theta_3, \\ x_3 &= r \sin\theta_1 \cos\theta_2, \\ x_4 &= r \cos\theta_1. \end{aligned}$$

Check that $x_1^2 + x_2^2 + x_3^2 + x_4^2 = r^2$. Find the volume element $dx_1\, dx_2\, dx_3\, dx_4 = r^3\, dr\, d\omega$, where $d\omega$ = a function of $\theta_1, \theta_2, \theta_3$ times $d\theta_1\, d\theta_2\, d\theta_3$. Generalize: in n dimensions, the volume element is $r^{n-1}\, dr\, d\omega$. ∎

• **1.4.30 Exercise.** Using polar coordinates, find those α for which

$$\int \cdots \int_D (x_1^2 + \ldots + x_n^2)^{-\alpha}\, dx_1 \ldots dx_n < \infty,$$

where $D = \{(x_1, \ldots, x_n) \mid x_1^2 + \ldots + x_n^2 \leq 1\}$. ∎

1.4.31 Exercise. Same question, but take $D = \{(x_1, \ldots, x_n) \mid x_1^2 + \ldots + x_n^2 \geq 1\}$. ∎

1.4.32 Exercise. Same D as in Exercise 1.4.30, but take the function $(1 - x_1^2 - \ldots - x_n^2)^{-\alpha}$ instead. ∎

• **1.4.33 Exercise.** For which α is

$$\int \cdots \int_{\mathbf{R}^n} (1 + x_1^2 + \ldots + x_n^2)^{\alpha}\, dx_1 \ldots dx_n < \infty? \text{ ∎}$$

1.4.34 Exercise. Let D be as in (1.4.30). To which L^p spaces does the function $(x_1^2 \ldots + x_n^2)^{-1}$ belong? ∎

1.4.35 Exercise. Let D be as in (1.4.31). To which L^p spaces does the function $(x_1^2 \ldots + x_n^2)^{-1}$ belong? ∎

1.4.36 Exercise. Let D be as in (1.4.31). To which L^p spaces does the function $(1 - x_1^2 - \ldots - x_n^2)$ belong? ∎

1.4.37 Exercise. Same as the preceding exercise, but take the function $(1 - x_1^2 - \ldots - x_n^2)^{-1}$. ∎

Miscellaneous Exercises

• **1.4.38 Exercise.** Prove that the L^∞ and L^2 metrics on $C([a, b])$ are not equivalent. ∎

1.4.39 Exercise. (A. Torchinsky) Let $f \in L_0^2([0, \pi])$. Is it possible to have simultaneously

$$\int_0^\pi (f(t) - \sin t)^2 \, dt \le 4/9 \quad \text{and} \quad \int_0^\pi (f(t) - \cos t)^2 \, dt \le 1/9? \text{ ∎}$$

• **1.4.40 Exercise.** Since $e^{-t}/t \le e^{-t}$ when $t \ge 1$, we can estimate

$$\int_1^\infty \frac{e^{-t}}{t} \, dt \le \int_1^\infty e^{-t} \, dt = e^{-1} \sim .368\ldots .$$

Show how to get the better estimate

$$\int_1^\infty \frac{e^{-t}}{t} \, dt \le \frac{e^{-1}}{\sqrt{2}} \sim .26\ldots . \text{ ∎}$$

1.4.41 Exercise*. Let $f \in L_0^p$, $g \in L_0^q$, $h \in L_0^r$, where $1/p + 1/q + 1/r = 1$. The domain of integration is some interval (a, b). Show that

$$\int_a^b |f(t)g(t)h(t)| \, dt \le \sqrt[p]{\int_a^b |f(t)|^p \, dt} \sqrt[q]{\int_a^b |g(t)|^q \, dt} \sqrt[r]{\int_a^b |h(t)|^r \, dt}.$$

How does this inequality imply the usual Hölder inequality? ∎

• **1.4.42 Exercise.** In $L_0^2([0, 1])$: Choose the parameters a, b, c to minimize the distance between the function $f(t) = \sin \pi t$ and i) the constant function $g_1(t) \equiv a$, ii) the linear function $g_2(t) = at + b$, and iii) the quadratic function $g_3(t) = at^2 + bt + c$. ∎

• **1.4.43 Exercise.** Suppose that $\int_0^\infty |f(t)| \, dt < \infty$. Prove, or give a counterexample:

$$\lim_{t \to \infty} |f(t)| = 0. \text{ ∎}$$

1.4.44 Exercise. Here is a sample of the kind of handwaving argument people like to apply to integrals.

"Let $\int_1^\infty |f(t)|^2 \, dt < \infty$. Since this integral is finite, $|f|^2$ must approach zero, as $t \to \infty$, a little faster than $1/t$. So $|f|$ goes to zero a little faster than $1/\sqrt{t}$. Therefore, $|f(t)|/\sqrt{t}$ goes to zero a little faster than $1/t$, and therefore

$$\int_1^\infty \frac{|f(t)|}{\sqrt{t}} \, dt < \infty."$$

Is the conclusion correct? If yes, give a careful proof. If no, give a counterexample. ∎

Section 2 Normed Linear Spaces

It was pointed out in Remark 1.0.8 that, while there may be a well-defined *distance* between the points of a metric space, the individual points need not have a *size*. How, in fact, could one assign *size* to the points of a general set X? Perhaps in a fairly arbitrary way: for each $x \in X$, pick a real number $s(x)$ and call it the "size of x". We probably don't like to think about negative sizes, so let us require that $s(x) \geq 0$ for all x. Should the sizes of neighboring points be close to each other? If so, we need to know when two points in X are "neighbors," and we therefore need a metric on X, or some reasonable way of measuring "closeness." But think about the real numbers, which serve as a model for many abstractions. What is the size of "2"? This is not a silly question. If you interpret 2 as the point that lies 2 units to the right of the origin on the number line, then that point has no width, and so if "size" means "length" or "extent", the size of "2" should be 0. If you are inclined to say that 2 has size 2, then you really mean that *size=distance from the origin*.

Both proposals are reasonable, but they lead to very different theories. If every point has size 0 because it has no width, there is no future in talking about relations between the sizes of points. This approach leads to *measure theory*, which is concerned with the size of whole sets of points. If "size" is to abstract the concept of absolute value of a real number, then one needs a metric space with a distinguished reference point, and the set X will be far from general. It will have to have a metric, and there must be a uniquely defined reference point.

This section deals with the second approach. "Absolute value", when generalized, becomes something called a *norm*. A norm measures the size of elements of certain vector spaces. We discuss vector spaces in Subsection 1, norms in Subsection 2. The applications in the present section are to matrix analysis in Subsection 3 and to ordinary and partial differential equations in Subsection 4.

Subsection 2.1 Linear Spaces

This subsection summarizes the properties of vector spaces. In the sophisticated circles of higher mathematics, one says *linear space* rather than *vector space*, but the words refer to the same object. It is assumed that the reader knows the elements of linear algebra, so the discussion is kept brief.

2.1.1 Definition. *A vector space, or linear space, over the field*[8] $\mathbf{F}$ *is a set* V *equipped with two operations:*

an addition defined on elements of V,

a multiplication of elements of V *by elements of* $\mathbf{F}$.

These operations satisfy the following axioms:

1) *Axioms for a "commutative group": if* $\mathbf{v}, \mathbf{w} \in V$ *then there is an element* $\mathbf{v}+\mathbf{w} \in V$; *this operation satisfies:*
   - *i)* $\mathbf{v}+\mathbf{w} = \mathbf{w}+\mathbf{v}$ *for all* $\mathbf{v}, \mathbf{w} \in V$,
   - *ii)* $\mathbf{u}+(\mathbf{v}+\mathbf{w}) = (\mathbf{u}+\mathbf{v})+\mathbf{w}$ *for all* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$;
   - *iii) there is an element* $\mathbf{0} \in V$ *such that* $\mathbf{v}+\mathbf{0} = \mathbf{v}$ *for all* $\mathbf{v} \in V$;
   - *iv) for each* $\mathbf{v} \in V$, *there is an element* $-\mathbf{v}$ *such that* $\mathbf{v}+(-\mathbf{v}) = \mathbf{0}$.
2) *Multiplication by scalars from* $\mathbf{F}$: *if* $\mathbf{v} \in V$ *and* $\alpha \in \mathbf{F}$, *there is an element* $\alpha\mathbf{v} \in V$; *this operation satisfies:*
   - *v)* *for* $\mathbf{v}, \mathbf{w} \in V$ *and* $\alpha \in \mathbf{F}$, $\alpha(\mathbf{v}+\mathbf{w}) = \alpha\mathbf{v}+\alpha\mathbf{w}$;
   - *vi)* *for* $\mathbf{v} \in V$ *and* $\alpha, \beta \in \mathbf{F}$, $(\alpha+\beta)\mathbf{v} = \alpha\mathbf{v}+\beta\mathbf{v}$;
   - *vii)* *for* $\mathbf{v} \in V$ *and* $\alpha, \beta \in \mathbf{F}$, $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$;
   - *viii)* *for* $\mathbf{v} \in V$, $1 \cdot \mathbf{v} = \mathbf{v}$, *where* 1 *is the identity in* $\mathbf{F}$. ∎

The axioms in Definition 2.1.1 are the product of years of investigation: what abstract properties of vectors in $\mathbf{R}^3$ allow one to deduce *all* relevant properties? Admirable though the axioms may be, it is not terribly important to memorize them. All rules for addition of vectors in $\mathbf{R}^3$, and for multiplication of such vectors by scalars, remain valid in general linear spaces.

Most of the familiar linear spaces carry a lot of extra structure. It is important to realize that Definition 2.1.1 implies nothing else about metrics, norms, inner products, dimension, etc. All those features are put in one at a time, and the fact that they are pretty much logically independent leads to a zoo of types of linear spaces with different structures, listed in a table in EDM II, Article 424.

2.1.2 Examples

A. If V is the set of vectors $a\vec{\mathbf{i}}+b\vec{\mathbf{j}}$ in $\mathbf{R}^2$ (with tail at the origin), all the axioms are satisfied when the scalars (α, β etc.) come from the field $\mathbf{R}$ of real numbers.

---

[8] For definition of "field", see EDM II. In analysis, generally speaking, one uses $\mathbf{F} = \mathbf{R}$ or $\mathbf{F} = \mathbf{C}$. Other fields (finite fields, quadratic number fields) play a role in coding theory, symbolic dynamics, etc.

B. $\mathbf{R}^n$ is a linear space over the field R; $\mathbf{C}^n$ is a linear space over the field C.

C. $l^p(\mathbf{R}, \mathbf{N})$ is a linear space over R; $l^p(\mathbf{C}, \mathbf{N})$ is a linear space over C.

D. Almost all the metric spaces described in Section 1 are linear spaces. The set of sequences of 0's and 1's in Example 1.1.45 is an exception.

E. The set X of polynomials $p(x) = a_n x^n + \ldots + a_1 x + a_0$ of degree $\leq n$ and having *rational* coefficients is a linear space over the field Q of rational numbers.

If the polynomials are only allowed to have integer coefficients, the set X *looks* like a linear space, but the scalars do not come from a field, they come from a "ring" (multiplicative inverses are missing from Z). (The set X is then called a *module* — an important concept, but not in this course.) ∎

Convention. From now on, we will only be concerned with linear spaces over two fields: the real numbers R or the complex numbers C. It should always be clear from the context which field is being considered. ∎

2.1.3 Exercise. Let V be a linear space. Use the axioms in (2.1.1) to prove: for every $\mathbf{v} \in V$, $0 \cdot \mathbf{v} = \mathbf{0}$. ∎

• **2.1.4 Exercise.** Let V be the set of continuous real-valued functions on $[0, 1]$ satisfying $\int_0^1 x f(x)\, dx = 0$. Is V a linear space? ∎

• **2.1.5 Exercise.** Let V be the set of real-valued, twice differentiable solutions of the differential equation $y''(x) + xy(x) = 0$ on $0 \leq x \leq 1$. Is V a linear space? Suppose the differential equation is replaced by $y''(x) + xy(x) = \sin x$ — is the set V of its solutions a linear space? ∎

The next few definitions summarize some standard ideas from linear algebra: dimension, linear dependence, and basis. The only new wrinkle may be the infinite-dimensional setting.

2.1.6 Definition. *Let V be a linear space over $\mathbf{F}$. A set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset V$ is said to be* ***linearly dependent*** *if there are scalars $a_1, \ldots, a_n \in \mathbf{F}$, not all zero, such that*

$$\sum_{j=1}^{n} a_j \mathbf{v}_j = 0.$$

If there are no such scalars, the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is said to be ***linearly independent.*** ∎

2.1.7 Definition. *Let V be a linear space over the field $\mathbf{F}$. Let N be the maximal number of linearly independent elements of V. Then N is called the* ***dimension*** *of V. (This number N could be ∞.)* ∎

The trivial linear space whose only element is 0 is assigned dimension 0. This is consistent with Definition 2.1.7: the only element of V, namely 0, satisfies $1 \cdot 0 = 0$, and since the scalar 1 is not zero, the set $\{0\}$ is not a linearly independent set.

2.1.8 Definition. *Let V be a linear space over $\mathbf{F}$ of dimension $N < \infty$. A linearly independent set $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ is called a basis of V.* ∎

Every finite-dimensional linear space has a basis. Indeed, in order to have dimension N the linear space V must have a linearly independent subset B with N elements but no linearly independent subsets containing $N + 1$ elements. The subset B, guaranteed to exist by the definition of dimension, will be a basis.

The importance of a basis lies in the following familiar property.

2.1.9 Proposition. *Let $\dim V < \infty$, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ be a basis of V. Then every $\mathbf{v} \in V$ has a unique expression in the form*

$$\mathbf{v} = \sum_{j=0}^{N} a_j \mathbf{v}_j, \quad a_1, \ldots, a_N \in \mathbf{F}. \text{ ∎}$$

When a linear space V is infinite-dimensional, it is much more difficult to find and work with bases. There are two fundamentally different possibilities. First, one can try to generalize the familiar finite-dimensional linear combinations (2.1.9) by allowing infinite series. This requires a notion of convergence in V, a metric or something similar. If there is no metric, infinite series can make no sense. One then retreats to the second alternative, the *algebraic* basis. This is a subset $B = \{\mathbf{v}_\lambda\}_{\lambda \in \Lambda}$, consisting of elements $\mathbf{v}_\lambda$ indexed by a possibly very large set Λ. For example, Λ could be $\mathbf{N}$, $\mathbf{Z}$, $\mathbf{R}$ or even something larger. The set B should have these properties:

1) Every finite subset of B is linearly independent;
2) Every $\mathbf{v} \in V$ can be written as a sum $\sum_{\lambda \in \Lambda} a_\lambda \mathbf{v}_\lambda$, in which *only finitely many of the scalars a_λ are nonzero.* Infinite linear combinations are not allowed, because we don't know how to think about the convergence of series.

Algebraic bases exist, according to some very deep axioms of set theory (axiom of choice or Zorn's lemma); see EDM II, Article 256.E. They are, however, not much use in analysis, simply because all the tools of analysis, such as convergence, continuity, and other limiting processes have been stripped away.

The upshot of this discussion is: we cannot go much further in generalizing linear algebra to infinite dimensions until many other tools have been developed. For the time being, we will generalize only the most elementary aspects of linear algebra.

• **2.1.10 Exercise.** Find three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbf{R}^2$ with this property: any two of the three vectors are linearly independent, but all three of them taken together are linearly dependent. Therefore, to determine whether a set of vectors is linearly independent, one must look at all the vectors at once. ∎

• **2.1.11 Exercise.** Are the functions $f_1(x) = 1$, $f_2(x) = x$, $f_3(x) = x^2$ independent in the linear space $L^2_0([0,1])$? ∎

2.1.12 Exercise. Show that the functions $\sin n\pi x$, $n = 1, 2, \ldots$ are linearly independent in $C([0,1])$ (see (1.4.1) for the definition of this linear space). ∎

• **2.1.13 Exercise.** Prove that $\mathbf{R}^2$ has dimension 2. (This is a dull result, but it may provide some practice in writing proofs). ∎

• **2.1.14 Exercise.** Let $\mathbf{e}_j$ be the sequence $(0, 0, \ldots, 1, 0, \ldots) \in \mathbf{R}^{\mathbf{N}}$ in which the only nonzero entry is a 1 in the j^{th} slot. Show that any finite subset of the form $\{\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_k}\}$ is independent. Conclude that the dimension of $\mathbf{R}^{\mathbf{N}}$ and of all the spaces $l^p(\mathbf{R}, \mathbf{N})$, $l^p(\mathbf{C}, \mathbf{N})$ is ∞. ∎

2.1.15 Exercise. Let $p = 1/2$. We saw in (1.1.7) that $l^p(\mathbf{R}, \mathbf{N})$ is not a metric space. Is it a linear space (over $\mathbf{R}$)? (Hint: the least obvious point is whether the sum of two elements in the set again belongs to the set). ∎

Subsection 2.2 Norms

We will now introduce an additional structure: a *norm* on a linear space, which is a gadget that tells us how to measure the size of a vector. It will generalize the notion of *absolute value* on $\mathbf{R}$ or of the *length* $\sqrt{x_1^2 + x_2^2}$ of a vector (x_1, x_2) in $\mathbf{R}^2$.

2.2.1 Definition. *A norm on a linear space V over $\mathbf{F}$ is a function $\|\cdot\|$ on V, satisfying:*

$$\|\mathbf{v}\| \geq 0, \quad \text{and } \|\mathbf{v}\| = 0 \text{ iff } \mathbf{v} = 0, \tag{2.2.2}$$

$$\|\alpha\mathbf{v}\| = |\alpha|\,\|\mathbf{v}\| \quad \text{for every } \alpha \in \mathbf{F}, \tag{2.2.3}$$

$$\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\| \quad \text{for all } \mathbf{v}, \mathbf{w} \in V. \tag{2.2.4}$$

A linear space V equipped with a norm $\|\cdot\|$ is called a ***normed linear space.*** ∎

Just as the absolute value (resp. modulus) of a real (resp. complex) number is the distance of that number from the origin, so is the norm of a vector in a normed linear space V the distance of the (tip of the) vector from the origin (the vector 0).

2.2.5 Proposition. *Let V be a normed linear space over $\mathbf{F}$ with norm $\|\cdot\|$. Then*

$$d(\mathbf{v}, \mathbf{w}) \stackrel{\text{def}}{=} \|\mathbf{v} - \mathbf{w}\|$$

is a metric on V. ∎

Proof. The axioms for a metric, (1.0.5,6,7), are easy to check:

$$d(\mathbf{v}, \mathbf{w}) \stackrel{\text{def}}{=} \|\mathbf{v} - \mathbf{w}\| \geq 0, \text{ and } \|\mathbf{v} - \mathbf{w}\| = 0 \text{ iff } \mathbf{v} - \mathbf{w} = 0 \text{ by (2.2.2).}$$

$$d(\mathbf{v},\mathbf{w}) = \|\mathbf{v}-\mathbf{w}\| = \|(-1)(\mathbf{w}-\mathbf{v})\| = |-1|\,\|\mathbf{w}-\mathbf{v}\| = d(\mathbf{w}-\mathbf{v}).$$

$$d(\mathbf{v},\mathbf{z}) = \|\mathbf{v}-\mathbf{z}\| = \|(\mathbf{v}-\mathbf{w})+(\mathbf{w}-\mathbf{z})\| \geq \|\mathbf{v}-\mathbf{w}\|+\|\mathbf{w}-\mathbf{z}\| = d(\mathbf{v},\mathbf{w})+d(\mathbf{w},\mathbf{z}). \blacksquare$$

It was noted in (2.1.2.D) that most of the examples in Section 1 are linear spaces. All the ones that are linear spaces are also normed linear spaces. *Not all linear spaces are normed*; a very useful example will appear in Section 4.

The concept of norm simplifies much of the notation in Section 1. There was no compelling reason to start the notes with metric spaces as opposed to normed linear spaces; I just thought the idea of metric to be a little more basic. Also, I have found that people get very attached to the efficient symbol $\|\mathbf{v}\|_p$, in fact so quickly that they tend not to remember what is being abbreviated. The reader who has faithfully written down the horrible sums in Section 1 should by now be unable to forget them.

The next few examples list the norms for the metric spaces we met in Section 1. The norm of an element $\mathbf{v}$ is just the distance of $\mathbf{v}$ from the null vector $\mathbf{0}$. In each case, I suggest you prove that the properties (2.2.2,3,4) of the purported norms are really satisfied. The arguments (relying on Minkowski's inequality and on simpler ideas) are quite routine, but practice will be useful if the ideas are new to you. Tricky spots will be identified.

2.2.6 Examples

A. For $V = \mathbf{R}^2$, the usual Euclidean norm, or size, of a vector $\mathbf{x} = (x_1, x_2)$,

$$\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$$

satisfies (2.2.2) and (2.2.3). The only nontrivial assertion is that the triangle inequality (2.2.4) holds. This is guaranteed by Minkowski's inequality (1.3.2) in $\mathbf{R}^2$, with $p = q = 2$,

$$\sqrt{(x_1+y_1)^2+(x_2+y_2)^2} \leq \sqrt{x_1^2+x_2^2} + \sqrt{y_1^2+y_2^2}.$$

B. The spaces $l^p(\mathbf{R}, \mathbf{N})$ carry the norms

$$\|\mathbf{x}\|_p \stackrel{\text{def}}{=} \sqrt[p]{\sum_{j=0}^{\infty} |x_j|^p} \quad , \quad 1 \leq p < \infty \tag{2.2.7}$$

$$\|\mathbf{x}\|_\infty \stackrel{\text{def}}{=} \sup_{j \geq 0} |x_j|.$$

The l^p distance $\|\mathbf{x} - \mathbf{y}\|_p$ discussed in Section 1.1 is equal to the size, or norm, of the sequence $\mathbf{x} - \mathbf{y}$ whose terms are $x_0 - y_0, x_1 - y_1, x_2 - y_2, \ldots$.

C. The sets $L_0^p([a,b])$ are normed linear spaces, with norm

$$\|f\|_p \stackrel{\text{def}}{=} \sqrt[p]{\int_a^b |f(t)|^p\,dt} \quad , \quad 1 \le p < \infty, \tag{2.2.8}$$

$$\|f\|_\infty \stackrel{\text{def}}{=} \sup_{a \le t \le b} |f(t)|.$$

The proof that $\|f\|_p = 0 \Longrightarrow f \equiv 0$ is left as exercise. (In the Lebesgue theory of L^p spaces, where the functions in L^p might be discontinuous *everywhere*, this step involves a significant change in the definition of the linear space L^p. See sections II.4 and II.6.)

D. The Sobolev norms for $1 \le p < \infty$ are

$$\|f\|_{p,k} \stackrel{\text{def}}{=} \sqrt[p]{\int_a^b |f(t)|^p + |f'(t)|^p + \ldots + |f^{(k)}(t)|^p dt}. \tag{2.2.9}$$

See Exercise 1.4.24 for the Minkowski inequality in Sobolev spaces.

E. The coin-toss space M of sequences $\underline{\epsilon}$ of 0's and 1's is *not* a normed linear space, because it is not even a linear space. Still, one might try to define the size of a sequence as

$$\|\underline{\epsilon}\| \stackrel{\text{def}}{=} \sum_{j=-\infty}^{\infty} \frac{\epsilon_j}{2^{|j|}}.$$

This is a reasonable definition, but without the additional linear space structure, it does not lead to helpful insights. ∎

• **2.2.10 Exercise.** a) Is $\int_0^1 x|f(x)|\,dx$ a norm on the linear space $C([0,1])$?
b) Is $\int_0^1 (1/2 - x)|f(x)|\,dx$ a norm on this same space? ∎

• **2.2.11 Exercise.** Let V be the set of continuously differentiable functions f on $[0,1]$ with $f'(1/2) = 0$. Show that V is a linear space. Which of the following formulas define norms on V?

a) $\|f\| = f(1/2)$,
b) $\|f\| = \max_{0 \le x \le 1} |f(x)|$,
c) $\|f\| = \max_{0 \le x \le 1} |f'(x)|$,
d) $\|f\| = \int_0^1 |f(x)|\,dx$,
e) $\|f\| = \int_0^1 |f'(x)|\,dx$. ∎

2.2.12 Exercise. Let $0 < p < 1$. Does (2.2.7) define a norm on $l^p(\mathbb{R}, \mathbb{N})$? ∎

• **2.2.13 Exercise.** Refer to Exercise 1.1.53. Use that metric to define a norm on the linear space of real 2×2 matrics. Prove that you have defined a norm. ∎

2.2.14 Exercise. Refer to Exercise 1.1.54. Is the surface M of the unit sphere a normed linear space? ∎

Subsection 2.3 Application: Norms of Matrices

Until now, all of our linear spaces have been generalizations of the classical vector space $\mathbb{R}^n$: the elements of those new spaces are sequences of numbers, indexed by integers or by a continuous variable. We now introduce a different class of normed linear spaces, which finds application in the analysis of linear equations. Applications to the estimation of errors made in the course of numerical solution will be emphasized.

To illustrate the issues involved, we look at a simple but instructive example, which I borrow from W. G. Strang, *Linear Algebra and its Applications*, Chapter 7.2. For a more thorough discussion of the issues I discuss below, you can just continue to read in Strang's text.

Let A be the matrix

$$\mathrm{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1.0001 \end{pmatrix}.$$

The solution of $\mathrm{Ax} = \mathrm{b}$ is

$$\mathrm{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \underbrace{\begin{pmatrix} 10001 & -10000 \\ -10000 & 10000 \end{pmatrix}}_{\mathrm{A}^{-1}} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 10000(b_1 - b_2) + b_1 \\ 10000(b_2 - b_1) \end{pmatrix}. \tag{2.3.1}$$

When a computer solves this equation, the right-hand side $\left(\begin{smallmatrix} b_1 \\ b_2 \end{smallmatrix}\right)$ may not be stored exactly. Rather, the true values b_1, b_2 may be truncated to a finite number of (binary) digits, resulting in a modified right-hand side $\mathrm{b}' = \left(\begin{smallmatrix} b_1' \\ b_2' \end{smallmatrix}\right)$. We then have the true solution x of $\mathrm{Ax} = \mathrm{b}$, and the computed solution x′ of $\mathrm{Ax}' = \mathrm{b}'$. This is still not the whole story. The coefficient matrix A may not be stored accurately, and further errors may be introduced by the arithmetic operations needed to solve the system with its inaccurate A and b. These latter difficulties will be ignored, and we concentrate on the simplest problem: to estimate the error $\mathrm{x} - \mathrm{x}'$ as it depends on $\mathrm{b} - \mathrm{b}'$, assuming all other operations to be carried out with infinite precision.

Subtracting $\mathrm{Ax}' = \mathrm{b}'$ from $\mathrm{Ax} = \mathrm{b}$ gives

$$\mathrm{A}(\mathrm{x} - \mathrm{x}') = \mathrm{b} - \mathrm{b}'. \tag{2.3.2}$$

Substituting (2.3.1) into (2.3.2), we get:

$$\begin{pmatrix} x_1 - x_1' \\ x_2 - x_2' \end{pmatrix} = \begin{pmatrix} 10000[(b_1 - b_1')] + (b_2 - b_2')] + (b_1 - b_1') \\ 10000[(b_2 - b_2') - (b_1 - b_1')] \end{pmatrix}.$$

It is time to use numbers. Let $b_1 = b_1'$; then the difference between b_2 and b_2' is amplified by a factor of 10000:

$$\begin{pmatrix} x_1 - x_1' \\ x_2 - x_2' \end{pmatrix} = 10000(b_2 - b_2') \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

If, say, $\mathbf{b} = \binom{1}{0}$ and $\mathbf{b}' = \binom{1}{.0001}$, then

$$\mathbf{x} - \mathbf{x}' = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Is this a significant error in the solution? Not really. We find, for the exact and perturbed solutions:

$$\mathbf{x} = \begin{pmatrix} 10001 \\ -10000 \end{pmatrix} \text{ and } \mathbf{x}' = \begin{pmatrix} 10000 \\ -9999 \end{pmatrix}.$$

The correct solution $\mathbf{x}$ *is huge, and the error* $\binom{1}{-1}$ *is small compared to* $\mathbf{x}$.

Suppose, however that $\mathbf{b} = \binom{1}{1}$ and $\mathbf{b}' = \binom{1}{1.0001}$. Then

$$\mathbf{x} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \ \mathbf{x}' = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \text{ and the error } \mathbf{x} - \mathbf{x}' \text{ is } \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

This is a significant error; its l^2 *norm,* $\sqrt{2}$*, is larger than the* l^2 *norm,* 1*, of the correct solution. A change of* .0001 *in the right-hand side has altered the solution beyond recognition.*

These calculations suggest a more precise statement of the problem:

(2.3.3) Compare the relative error in the solution, $\dfrac{\|\mathbf{x} - \mathbf{x}'\|}{\|\mathbf{x}\|}$,

to the relative error in the input. $\dfrac{\|\mathbf{b} - \mathbf{b}'\|}{\|\mathbf{b}\|}$.

Note that no particular norm $\|\cdot\|$ has been chosen yet: any of the $\|\cdot\|_p$ norms might be used, or perhaps other norms that we have not mentioned.[9] The outline of the solution of (2.3.3) (or rather, of a possible and useful solution) motivates the subsequent discussion.

Starting with an arbitrary but fixed norm $\|\cdot\|$ on vectors, we will define a related norm on matrices, so that the following property holds:

(2.3.4) whenever $\mathbf{a} = \mathbf{C}\mathbf{y}$, then $\|\mathbf{a}\| \leq \|\mathbf{C}\| \cdot \|\mathbf{y}\|$.

[9] The text *Analysis of Numerical Methods*, by E. Isaacson and H. B. Keller, has a thorough discussion of norms in connection with the numerical analysis of linear systems.

Here C is a matrix, $\|C\|$ is its norm (yet to be defined), and y and a are arbitrary vectors related only by $Cy = a$. Before we look at details about matrix norms, let us show how (2.3.4) is brought to bear on the stability problem (2.3.3). It suffices, for the time being, to think of $\|C\|$ as *some* measure of the size of the matrix C.

1) Consider the equation $A^{-1}(b - b') = (x - x')$. Apply (2.3.4), setting $C = A^{-1}, y = (b - b'), a = (x - x')$:

$$\|x - x'\| \leq \|A^{-1}\| \cdot \|b - b'\|. \tag{2.3.5}$$

2) Consider the equation $Ax = b$. Apply (2.3.4) with $C = A, y = x, b = a$:

$$\|b\| \leq \|A\| \cdot \|x\|$$

or, equivalently,

$$\frac{1}{\|x\|} \leq \|A\| \cdot \frac{1}{\|b\|}. \tag{2.3.6}$$

Multiply inequality (2.3.5) by inequality (2.3.6) to get

$$\frac{\|x - x'\|}{\|x\|} \leq \|A^{-1}\| \, \|A\| \frac{\|b - b'\|}{\|b\|}. \tag{2.3.7}$$

2.3.8 Definition. *The number*

$$\operatorname{cond}(A) \stackrel{\text{def}}{=} \|A^{-1}\| \cdot \|A\|$$

is called the ***condition number*** *of the matrix A (in the norm* $\|\cdot\|$*).* ∎

The condition number plays a fundamental role in the analysis of algorithms for the solution of linear systems. If $\|A^{-1}\| \cdot \|A\|$ is small [10] then the solution of $Ax = b$ is considered to be stable under perturbations of the right side: that is the content of (2.3.7).

If, on the other hand, $\|A^{-1}\| \cdot \|A\|$ is large (much bigger than 1), one should be prepared for the worst-case scenario. Expect that the inequality in (2.3.7) is an equality for some choices of b, b': there may be vectors b, b' with small relative difference $\|b - b'\|/\|b\|$ for which the relative solution error $\|x - x'\|/\|x\|$ will be large,

$$\frac{\|x - x'\|}{\|x\|} = \operatorname{cond}(A) \times \frac{\|b - b'\|}{\|b\|},$$

[10] We will see in (2.3.46) that $\|A^{-1}\| \cdot \|A\|$ is always ≥ 1. "Small" therefore means: the condition number is not too much bigger than 1. How big it can be, and still be "small", is a question left to the numerical analysts.

where the condition number $\mathrm{cond}(\mathbf{A})$ is large. Systems with large condition number are difficult to solve, because the "instability" is a property of the matrix $\mathbf{A}$, not of the method used to solve the equations.[11]

I will not attempt to discuss any specific numerical methods. Numerical analysis has served my purpose — the need for a proper definition of *matrix norm* has been established, "proper" referring to the crucial relation (2.3.4), from which the relative-error estimate (2.3.7) follows. I now turn to the definition of the *norm of a matrix* and to some examples which will clarify the meaning of (2.3.4) and (2.3.7).

From now on, all general matrices will be $n \times n$. Their entries may be real or complex, and the matrices may act on vectors from $\mathbf{R}^n$ or from $\mathbf{C}^n$. The formulas don't change. In examples, the setting will of course be specified more precisely.

The set of $n \times n$ matrices is a linear space of dimension n^2 — the n^2 coordinates are simply arranged in a square, rather than in the customary single column. One could therefore define a norm on matrices by taking one of the standard l^p norms on $\mathbf{R}^{n^2}$ or $\mathbf{C}^{n^2}$ — e.g.,

$$\|\mathbf{A}\| = \sqrt{\sum_{i=1}^{n}\sum_{i=1}^{n} |A_{ij}|^2}, \tag{2.3.9}$$

for example

$$\left\| \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \right\| = \sqrt{1^2 + 2^2 + 3^2 + 4^2}.$$

This definition will satisfy all axioms for norms, but it turns out to be of limited use. Matrices have two functions in life: to multiply each other, and to multiply vectors. A sensible norm should relate the size of a product $\mathbf{AB}$ to the sizes of the factors $\mathbf{A}$ and $\mathbf{B}$, and it should relate the size of $\mathbf{Ax}$ to the sizes of $\mathbf{A}$ and of $\mathbf{x}$ (the latter connection was postulated in (2.3.4) and led to the introduction of the condition number). The norm (2.3.9) does neither of these things. At best, one can deduce an estimate of the form $\|\mathbf{Ax}\|_2 \leq \gamma \|\mathbf{A}\|_2 \|\mathbf{x}\|_2$, where γ is a constant that is difficult to compute but must be included in a modified definition of condition number. One therefore confines attention to a class of matrix norms that are induced in a natural way from the norms of the vectors on which the matrices act. Those norms are called *natural* matrix norms.

2.3.10 Definition. *Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$ (or $\mathbf{C}^n$). For a matrix $\mathbf{A}$, define*

$$\|\mathbf{A}\| \stackrel{\text{def}}{=} \sup_{\|\mathbf{x}\|=1} \|\mathbf{Ax}\|.$$

---

[11] Again, see *Linear Algebra and its Applications*, by W. G. Strang, for an introduction to the condition number and related matters.

(It is standard practice to use the same symbol, $\|\cdot\|$, for the norms of A *and* x.*) The matrix norm $\|\cdot\|$ is called the* ***natural matrix norminduced by the vector norm*** *$\|\cdot\|$ on* $\mathbf{R}^n$. ∎

Interpretation: Application of A to a vector x will result in Ax, which is a new vector obtained from x by rotating x and stretching (or compressing) it. $\|\mathbf{Ax}\|$ is a measure of the stretching undergone by x. One only measures the stretch of *unit* vectors for a simple reason: since, for any number α, $\mathbf{A}(\alpha\mathbf{x}) = \alpha \cdot \mathbf{Ax}$ (A is linear), one could get arbitrarily large $\|\mathbf{Ax}\|$ simply by scaling $\|\mathbf{x}\|$. This would give no information about A. Somehow, $\|\mathbf{x}\|$ has to be kept constant so that the effect of A can be estimated correctly. The definition (2.3.10) is sometimes rephrased as

$$\|\mathbf{A}\| \stackrel{\text{def}}{=} \sup\{\frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}\}, \tag{2.3.11}$$

but formula (2.3.10) is more convenient.

Hence: $\|\mathbf{A}\|$ = the maximum stretching experienced by a unit vector x. The "stretching" could be by a factor < 1, i.e., it could be a compression. "Unit vector" is measured by the chosen norm $\|\cdot\|$, which could be different from the usual Euclidean length, although that is certainly a very popular norm.

It will be shown later in this subsection that the number $\|\mathbf{A}\|$ defined in (2.3.10) satisfies properties (2.2.2,3,4) required of a norm. First, however, it is more important to cultivate an operational understanding of the definition. We will therefore study a computable example in great detail.

Let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}.$$

I will calculate a few of the possible norms. Remember that the natural norms are derived only after a choice of norm on vectors (in $\mathbf{R}^2$, in this example) has been made. I will compute the matrix norms defined, via (2.3.10), from the l^1, l^2, and l^∞ norms on $\mathbf{R}^2$. The l^2 norm is, perhaps surprisingly, the most subtle of the three; it is done last.

The calculations are a little long, so it may be worthwhile to advertise the conclusions at the outset. The l^2 norm of A is the square root of the largest eigenvalue of $\mathbf{A}^t\mathbf{A}$ ($\mathbf{A}^t$ = transpose of A); this holds for arbitrary $n \times n$ matrices. Since eigenvalues are difficult to compute except for 2×2 matrices, the norm $\|\mathbf{A}\|_2$ can in general not be found explicitly. The norms $\|\mathbf{A}\|_1$ and $\|\mathbf{A}\|_\infty$, however, can be read off from the entries of A (even in the $n \times n$ case). Hence, these norms are more convenient as long as they contain the information one wants. If the $\|\mathbf{A}\|_2$ norm is essential, it can be estimated in terms of $\|\mathbf{A}\|_1$ and $\|\mathbf{A}\|_\infty$ — see Exercise 2.3.39.

2.3.12 Example. Start with the l^∞ norm on $\mathbf{R}^2$. By definition,

$$\|\mathbf{A}\|_\infty \stackrel{\text{def}}{=} \sup_{\|\mathbf{x}\|_\infty = 1} \|\mathbf{Ax}\|_\infty.$$

Let us untangle this definition.

$$\|\mathbf{x}\|_\infty = 1 \text{ means } \max\{|x_1|, |x_2|\} = 1$$

$$\|\mathbf{Ax}\|_\infty \stackrel{\text{def}}{=} \max\{|(\mathbf{Ax})_1|, |(\mathbf{Ax})_2|\}$$

$$= \max\{|x_1 + 2x_2|, |x_1|\}.$$

To find $\|\mathbf{A}\|_\infty$, we therefore need to find the largest possible value of $|x_1 + 2x_2|$ or $|x_1|$, whichever is larger, as $\mathbf{x}$ ranges over the perimeter of the square

$$\max\{|x_1|, |x_2|\} = 1.$$

On the right side of the square, for example, $x_1 = 1$ while $-1 \le x_2 \le 1$. So

$$|x_1 + 2x_2| = |1 + 2x_2| \text{ ranges from 1 to 3 while } |x_1| = 1.$$

The largest value of either $|x_1 + 2x_2|$ or $|x_1|$ on the right side of the square is therefore 3, achieved at

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Checking the other three sides in a similar way, one finds nothing larger than 3, and $\|\mathbf{A}\|_\infty = 3$.

The geometric interpretation must be understood. Consider the vector $\binom{1}{1}$. It has $\| \ \|_\infty$ length 1, and is rotated and stretched by A:

$$\begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

This new vector has length 3 rather than $\sqrt{10}$ because we are using $\| \ \|_\infty$, and 3 is the maximum stretch in the l^∞ norm of $\mathbf{R}^2$ of any l^∞ unit vector.

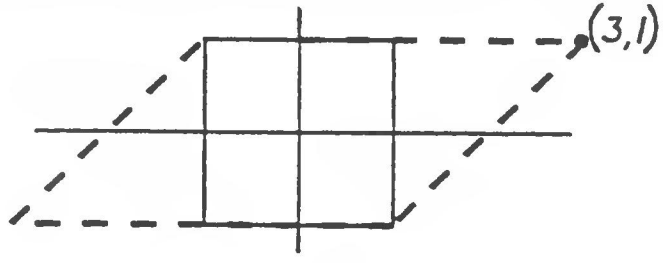


Figure 21. *The matrix* A *acting on the unit* l^∞ *sphere*

The matrix A takes the l^∞ unit "sphere", $\{\mathbf{x} \mid \|\mathbf{x}\|_\infty = 1\}$ into a parallelogram, as shown in Figure 21. The vector $\binom{1}{1}$ is moved to the vertex indicated by the fat dot, $\binom{3}{1}$. Again: the length of $\binom{3}{1}$ is *not* $\sqrt{3^2 + 1^2} = \sqrt{10}$, but, since l^∞ is used, it is $\max\{3, 1\} = 3$. ∎

2.3.13 Example. Let us now find $\|\mathbf{A}\|_1$; this means that $\mathbf{R}^2$ is equipped with the l^1 norm, $\|\mathbf{x}\|_1 = |x_1| + |x_2|$, and

$$\|\mathbf{A}\|_1 \stackrel{\text{def}}{=} \sup_{\|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1.$$

Or, more plainly:

$$\|\mathbf{A}\|_1 = \text{maximum of } |x_1 + 2x_2| + |x_1| \text{ subject to } |x_1| + |x_2| = 1.$$

This is rather painful to work out because of the absolute values. The "sphere" $|x_1| + |x_2| = 1$ is the diamond in Figure 2. On the side $x_1 + x_2 = 1$, for example, we substitute $x_2 = 1 - x_1$ into $|x_1 + 2x_2| + |x_1|$ to get $|2 - x_1| + |x_1|$; x_1 runs from 0 to 1, and the maximum of $|2 - x_1| + |x_1|$ is 2. Checking each of the other sides in turn, one finds no larger values, and therefore

$$\|\mathbf{A}\|_1 = 2.$$

Again, think about the geometric interpretation: what shape does the unit diamond get mapped to, and which vector is stretched the most in the l^1-norm of $\mathbf{R}^2$? ∎

2.3.14 Example. Finally, we compute the matrix norm $\|\mathbf{A}\|_2$ corresponding to the Euclidean vector norm $\|\cdot\|_2$ on $\mathbf{R}^2$:

$$\|\mathbf{A}\|_2 \stackrel{\text{def}}{=} \sup_{\|\mathbf{x}\|_2=1} \|\mathbf{A}\mathbf{x}\|_2.$$

Now,

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ x_1 \end{pmatrix},$$

so that

$$\|\mathbf{A}\|_2^2 = \sup_{x_1^2+x_2^2=1} (x_1 + 2x_2)^2 + (x_1)^2. \tag{2.3.15}$$

To maximize $\|\mathbf{A}\|_2$, it suffices, of course, to maximize $\|\mathbf{A}\|_2^2$, and there is a completely elementary way to find the answer: substitute $x_2 = \sqrt{1 - x_1^2}$ and maximize the resulting function of the single variable x_1. This method, however, does not work in $\mathbf{R}^n$ for $n \geq 3$. A different approach is therefore desirable.

Maximization of a function of two variables subject to constraints is a standard problem in the calculus of several variables: it is usually solved by the *method of Lagrange multipliers*. The technique is reviewed in small print; if you remember it, skip the summary.

2.3.16 Lagrange multipliers. We want to maximize a function f on $\mathbf{R}^2$ subject to a constraint $g(x_1, x_2) = 0$. In our matrix norm example,

$$f(x_1, x_2) = \|\mathbf{Ax}\|_2^2 = (x_1 + 2x_2)^2 + x_1^2 = 2x_1^2 + 4x_1x_2 + 4x_2^2,$$

and

$$g(x_1, x_2) = \|\mathbf{x}\|_2^2 - 1 = x_1^2 + x_2^2 - 1.$$

Figure 22 is drawn for this particular case, but of course the method works in general. The ellipses are the level curves of f, i.e., the curves $f(x_1, x_2) = \text{constant}$. The gradient vector field of f,

$$\vec{\nabla} f(x_1, x_2) = \frac{\partial f(x_1, x_2)}{\partial x_1}\vec{\imath} + \frac{\partial f(x_1, x_2)}{\partial x_2}\vec{\jmath}$$

indicates the direction and rate of fastest increase of f at the point (x_1, x_2). It is always perpendicular to the level curves of f. The thick circle in the figure is the curve $\Gamma : g(x_1, x_2) = 0$, the constraint. We are to find the maximum (or maxima) of f on that curve. Look at point $P \in \Gamma$, and at the gradient vector $\vec{\nabla} f$ at point P. This vector is not perpendicular to $\Gamma : g(x_1, x_2) = 0$, and therefore has a nonzero component tangent to Γ. It follows that f increases along Γ, in the direction of $\vec{\nabla} f(B)_{\text{tangent}}$, and that therefore $f(P)$ is not a maximum of f on Γ. (For the same reason, $f(P)$ is not a minimum.)

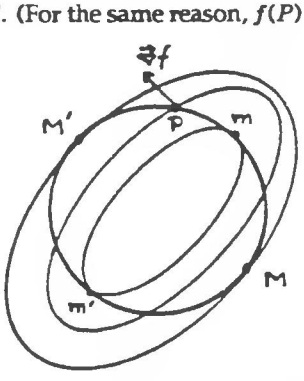


Figure 22. *Geometry of Lagrange multipliers*

A maximum (or minimum) of f on Γ can only occur at points where $\vec{\nabla} f$ is perpendicular to Γ and therefore has zero tangential component. Since $\vec{\nabla} g$ is also always perpendicular to Γ, we are looking for points (x_1, x_2) at which $\vec{\nabla} f$ and $\vec{\nabla} g$ are parallel, meaning that

$$\vec{\nabla} f(x_1, x_2) = \lambda \vec{\nabla} g(x_1, x_2). \tag{2.3.17}$$

Those points, of course, must lie on Γ,

$$g(x_1, x_2) = 0. \tag{2.3.18}$$

Equations (2.3.17) and (2.3.18) form a system of three (usually nonlinear) equations for the three unknowns x_1, x_2, λ. In the figure, the four points M, M', m, m' satisfy all three conditions. The maximum of f is assumed at M and M', and the minimum at m

and m'. When one does not have a picture available, one might evaluate f at all the solutions of (2.3.17) and (2.3.18) to determine maxima and minima. There are general prescriptions, but those will not be needed.

Functions of more variables, subject perhaps to several constraints, are treated in a similar manner. Suppose we are to find the maxima (and/or minima) of the function $f(x_1, \ldots, x_n)$, subject to the constraints $g_1(x_1, \ldots, x_n) = 0, \ldots, g_k(x_1, \ldots, x_n) = 0$. We then solve

$$\begin{aligned} \vec{\nabla} f(x_1, \ldots, x_n) &= \sum_{j=1}^{k} \lambda_j \vec{\nabla} g_j(x_1, \ldots, x_n), \\ g_1(x_1, \ldots, x_n) &= 0, \\ &\vdots \\ g_k(x_1, \ldots, x_n) &= 0, \end{aligned} \tag{2.3.19}$$

and test the points found to determine the maxima and minima. As in single-variable calculus, there are "inflection points", where $\vec{\nabla} f = 0$ and $\lambda = 0$, or (this is new) exceptional points where $\vec{\nabla} g = 0$ and λ might have to be ∞, but those will play no role in our matrix example.

The proportionality constants $\lambda_1, \ldots, \lambda_k$ in (2.3.19) are called *Lagrange multipliers.* The solutions of (2.3.19) are called *critical points* of (the constrained function) f

2.3.20 **Exercise.** Classify the critical points of $f(x_1, x_2) = x_2 - x_1^4$ on $x_1^2 + (x_2 - 1)^2 = 1$: maxima, minima, other. Figure out what "other" is. ∎

Applying the method of Lagrange multipliers to (2.3.15), we find the following system:

$$\begin{aligned} 2x_1 + 2x_2 &= \lambda x_1 \\ 2x_1 + 4x_2 &= \lambda x_2 \end{aligned} \tag{2.3.21}$$

together with the constraint

$$x_1^2 + x_2^2 = 1. \tag{2.3.22}$$

Equations (2.3.21) are – miraculously – an *eigenvalue problem*, with solutions

$$\lambda_+ = 3 + \sqrt{5}, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = c \begin{pmatrix} 1 \\ (\lambda_+ - 2)/2 \end{pmatrix}, \tag{2.3.23}$$

and

$$\lambda_- = 3 - \sqrt{5}, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = c \begin{pmatrix} 1 \\ (\lambda_- - 2)/2 \end{pmatrix}. \tag{2.3.24}$$

Here c is an arbitrary constant as far as the eigenvalue problem (2.3.21) is concerned, but the constraint (2.3.22) forces us to pick c so that the eigenvector has length 1. For each λ, there are two unit eigenvectors, which are negatives of each other. The final expression for those vectors is messy and will not be needed.

We have now found four solutions x_1, x_2, λ of the Lagrange multiplier problem (2.3.21). Two give the minimum of $2x_1^2 + 4x_1x_2 + 4x_2^2$ on $x_1^2 + x_2^2 = 1$, the other two the maximum. The desired norm of $\mathbf{A}$ will be the *square root* of that maximum, since the norm is

$$\max \sqrt{2x_1^2 + 4x_1x_2 + 4x_2^2}.$$

Mindless calculation can certainly produce the final answer, but a little thought saves work and provides additional insight.

The key question, which, I hope, has occurred to all readers, is: what is this matrix

$$\begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$$

that appeared so unexpectedly in the eigenvalue problem (2.3.21)?

To answer the question (and read off the norm of $\mathbf{A}$ without further calculation), we first note that we found the extrema not of $\|\mathbf{Ax}\|_2$, but of $\|\mathbf{Ax}\|_2^2$. Now, with $(\cdot,\cdot)$ denoting the usual dot product on $\mathbf{R}^2$, and $\mathbf{A}^t$ the transpose of $\mathbf{A}$, we have

$$\|\mathbf{Ax}\|_2^2 = (\mathbf{Ax}, \mathbf{Ax}) = (\mathbf{A}^t\mathbf{Ax}, \mathbf{x}). \tag{2.3.25}$$

Set $\mathbf{B} = \mathbf{A}^t\mathbf{A}$. This is a symmetric matrix. Quite generally, if

$$\mathbf{B} = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

then

$$(\mathbf{Bx}, \mathbf{x}) = ax_1^2 + 2bx_1x_2 + cx_2^2. \tag{2.3.26}$$

Lagrange's method, applied to (2.3.26) with constraint $x_1^2 + x_2^2 = 1$, yields the two equations

$$\begin{aligned} ax_1 + bx_2 &= \lambda x_1 \\ bx_1 + cx_2 &= \lambda x_2. \end{aligned} \tag{2.3.27}$$

This is *precisely the eigenvalue problem for the symmetric matrix* $\mathbf{B}$.

All these calculations can be summarized and generalized in a couple of lines.

2.3.28 Lemma. *Let* $\mathbf{B}$ *be a symmetric* $n \times n$ *matrix. The critical points of the function* $f(\mathbf{x}) = (\mathbf{Bx}, \mathbf{x})$, *subject to* $(\mathbf{x}, \mathbf{x}) = \|\mathbf{x}\|_2^2 = 1$, *are solutions of*

$$\mathbf{Bx} = \lambda \mathbf{x}, \quad \|\mathbf{x}\|_2 = 1,$$

i.e., normalized eigenvectors of $\mathbf{B}$. *The corresponding Lagrange multipliers are eigenvalues. The value of* f *at a critical point* $\mathbf{x}_*$ *with Lagrange multiplier* λ_* *is*

$$f(\mathbf{x}_*) = (\mathbf{Bx}_*, \mathbf{x}_*) = \lambda_*(\mathbf{x}_*, \mathbf{x}_*) = \lambda_*. \blacksquare$$

You should verify this for the matrix

$$\mathbf{B} = \mathbf{A}^t\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}$$

in our example; the general proof is left as exercise.

The computation of the l^2 norm of $\mathbf{A}$ is finally complete. By Lemma (2.3.28), the Lagrange multipliers, *alias* eigenvalues of (2.3.21), are precisely the maximum and minimum of $\|\mathbf{Ax}\|_2^2$ on $\|\mathbf{x}\|_2 = 1$; obviously, the larger one, $3 + \sqrt{5}$, is the maximum. The maximum of $\|\mathbf{Ax}\|_2$ is then its square root:

$$\|\mathbf{A}\|_2 = \sqrt{3 + \sqrt{5}}.$$

To finish this example, we review the geometric interpretation of $\|\mathbf{A}\|_2 = \sqrt{3 + \sqrt{5}}$, as in the two preceding examples. The matrix $\mathbf{A}$ takes vectors on the unit circle to vectors whose tips lie on an ellipse.

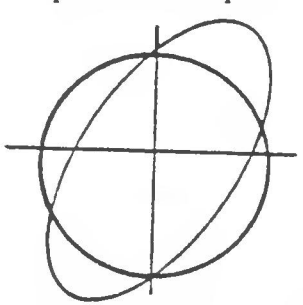


Figure 23. *The matrix* $\mathbf{A}$ *acting on the unit* l^2 *sphere*

Indeed, if

$$\begin{pmatrix} x_1' \\ x_2' \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},$$

then

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \end{pmatrix},$$

and the unit circle $x_1^2 + x_2^2 = 1$ becomes $x_1'^2 - 2x_1'x_2' + 5x_2'^2 = 4$, which is an ellipse (as sketched in Figure 23) with major semi-axis $\sqrt{3+\sqrt{5}}$. Two unit vectors (negatives of each other) are stretched to this length, and no unit vector is stretched more. Note that this ellipse is *not one of the level curves* $2x_1^2 + 4x_1x_2 + 4x_2^2 = \text{constant}$ that arise in the geometric intepretation of Lagrange's method in this example; rather, it is a level curve of the function $\|\mathbf{A}^{-1}\mathbf{x}\|_2$ (figure out why). ∎

The calculations in the last three examples generalize to arbitrary real or complex $n \times n$ matrices. The results for l^1 and l^∞ are stated as exercises. The l^2 norm deserves special emphasis because of its importance.

2.3.29 Proposition. *Let* $\mathbf{A}$ *be a real* $n \times n$ *matrix. The norm* $\|\mathbf{A}\|_2$ *equals the square root of the largest eigenvalue of* $\mathbf{A}^t\mathbf{A}$. ∎

The proof, which is immediate from (2.3.25) and Lemma 2.3.28, is left as exercise.

• **2.3.30 Exercise.** (This exercise completes the example at the beginning of this subsection). Compute the condition number, in the l^2 norm, of the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 1.0001 \end{pmatrix}.$$

Relate this condition number to the error calculations at the beginning of the subsection. ∎

• **2.3.31 Exercise.** For the matrix $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ find $\|\mathbf{A}\|_1$, $\|\mathbf{A}\|_2$, and $\|\mathbf{A}\|_\infty$. Find the eigenvalues of $\mathbf{A}$ and compare them to those matrix norms. Same question for $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. ∎

2.3.32 Exercise. Let $\mathbf{A}$ be a diagonal matrix,

$$\mathbf{A} = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

and set $M = \max\{|\lambda_1|, \ldots, |\lambda_n|\}$. Show that $\|\mathbf{A}\|_2 = M$. ∎

2.3.33 Exercise. This exercise requires some knowledge of matrix theory. Let $\mathbf{Q}$ be a real orthogonal matrix, $\mathbf{Q}^t\mathbf{Q} = \mathbf{I}$, where $\mathbf{Q}^t$ is the transpose of $\mathbf{Q}$. Recall that $\mathbf{Q}$ preserves Euclidean length. Show that $\|\mathbf{Q}\|_2 = 1$. ∎

2.3.34 Exercise. This exercise also requires knowledge of matrix theory. Let $\mathbf{A}$ be a real symmetric matrix. Remember that there exists an orthogonal matrix $\mathbf{Q}$ such that $\mathbf{Q}^t\mathbf{A}\mathbf{Q}$ is diagonal. Use the definition of $\|\mathbf{A}\|_2$ and Exercise 2.3.32 to prove that $\|\mathbf{A}\|_2$ equals the maximum of the absolute values of its eigenvalues. (The conclusion also follows easily from Proposition 2.3.29.) ∎

• **2.3.35 Exercise.** a) For a matrix $A = (a_{ij})$ on R^n, show that

$$\|A\|_\infty \leq \max_i \sum_{j=1}^{n} |a_{ij}|,$$

$$\|A\|_1 \leq \max_j \sum_{i=1}^{n} |a_{ij}|.$$

b) Show that there is always an $x \in R^n$ with $\|x\|_\infty = 1$, for which

$$\|Ax\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|.$$

Hence deduce that $\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$.
c) Prove in a similar way that $\|A\|_1 = \max_j \sum_{i=1}^{n} |a_{ij}|$.

Parts b) and c) explain the results of examples (2.3.12) and (2.3.13). ∎

• **2.3.36 Exercise.** Let $A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$. Compute $\|A\|_2$ and $\|A^{-1}\|_2$, and find the condition number. Compute the eigenvalues of A and A^{-1}, and express the condition number in terms of these eigenvalues. ∎

• **2.3.37 Exercise.** Compute the l^1 condition number of the matrix in the preceding problem. ∎

• **2.3.38 Exercise.** Let A be an invertible symmetric $n \times n$ matrix. Let $|\lambda|_{max}$ be the maximum of the absolute values of its eigenvalues, and let $|\lambda|_{min}$ be the minimum of the absolute values of its eigenvalues. Show that

$$\text{cond}(A) = |\lambda|_{max}/|\lambda|_{min}. \text{ ∎}$$

2.3.39 Exercise*. Let A be an $n \times n$ matrix. Show that

$$\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}.$$

(Hint: Exercise 1.3.33.) ∎

It is time to establish the basic properties of matrix norms. First, we must prove that Definition 2.3.10 really satisfies the axioms (2.2.2,3,4) for a norm. Then we find the relation between the norms of A, B and AB, respectively A, x and Ax. The results justify the term *natural matrix norm*. Our arguments rely on various properties of "supremum," which are summarized in the Appendix to this subsection. One other result from advanced calculus is used; if you are not familiar with it, accept it and wait for Chapter 2, where it will be proved in a general setting.

2.3.40 Proposition. *$||\mathbf{A}||$, as defined in (2.3.10), satisfies conditions (2.2.2,3,4) and is therefore really a norm.* ∎

Proof. It could be confusing if the symbol $||\cdot||$ were used to denote both a norm on vectors and a purported norm on matrices. We therefore reserve the symbol $||\cdot||$ for the chosen vector norm on $\mathbf{R}^n$ or $\mathbf{C}^n$, and write

$$N(\mathbf{A}) = \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}||$$

for what we hope will be the corresponding natural matrix norm. It is to be shown that

(2.3.41) $$N(\mathbf{A}) \geq 0, \text{ and } N(\mathbf{A}) = 0 \text{ iff } \mathbf{A} = 0;$$

(2.3.42) $$N(\alpha\mathbf{A}) = |\alpha|\, N(\mathbf{A}) \text{ for all } \alpha \in \mathbf{F};$$

(2.3.43) $$N(\mathbf{A}+\mathbf{B}) \leq N(\mathbf{A}) + N(\mathbf{B}).$$

First, however, there is an implicit claim:

$$\text{For every } \mathbf{A}, \quad \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}|| < \infty.$$

This is based on two lemmas.

Lemma 1. *Fix $\mathbf{A}$. The function $\mathbf{x} \longmapsto ||\mathbf{Ax}||$ from $\mathbf{R}^n$ to $[0,\infty)$ is continuous.* ∎

Proof. It is to be shown that $\big|\, ||\mathbf{Ax}|| - ||\mathbf{Ay}|| \,\big|$ is less than an arbitrarily prescribed $\epsilon > 0$, provided $||\mathbf{x}-\mathbf{y}|| <$ a sufficiently small $\delta > 0$.

Now, $||\mathbf{Ax}|| \leq ||\mathbf{A}(\mathbf{x}-\mathbf{y})|| + ||\mathbf{Ay}||$ by the triangle inequality, whence

i) $$||\mathbf{Ax}|| - ||\mathbf{Ay}|| \leq ||\mathbf{A}(\mathbf{x}-\mathbf{y})||.$$

Also,

$$||\mathbf{Ay}|| \leq ||\mathbf{A}(\mathbf{x}-\mathbf{y})|| + ||\mathbf{Ax}||$$

whence

ii) $$-||\mathbf{A}(\mathbf{x}-\mathbf{y})|| \leq ||\mathbf{Ax}|| - ||\mathbf{Ay}||.$$

Combining i) and ii) gives:[12]

$$\big|\,\|\mathbf{Ax}\| - \|\mathbf{Ay}\|\,\big| \le \|\mathbf{A}(\mathbf{x}-\mathbf{y})\|. \tag{2.3.44}$$

In order to make $\big|\,\|\mathbf{Ax}\| - \|\mathbf{Ay}\|\,\big| <$ a given $\epsilon > 0$, we should therefore make $\|\mathbf{A}(\mathbf{x}-\mathbf{y})\|$ small enough. We show that this can be done. The rest of the proof is perhaps more complicated than one might expect, but then we have assumed *nothing* about the vector norm $\|\cdot\|$ other than that it satisfies (2.2.2,3,4). In particular, it need not be an l^p norm. (See Exercise 2.3.52 for a different kind of norm on $\mathbf{R}^n$.)

Let $\mathbf{e}_1 \ldots, \mathbf{e}_n$ be the unit vectors in $\mathbf{R}^n$: $\mathbf{e}_i = (0, \ldots, 1, \ldots, 0)^t$, with the 1 in the i-th place. Write

$$\mathbf{A}\Big(\sum_{i=1}^{n} x_i \mathbf{e}_i\Big) = \sum_{k=1}^{n} \Big(\sum_{i=1}^{n} A_{ki} x_i\Big)\mathbf{e}_k.$$

Then

$$\mathbf{x}-\mathbf{y} = \sum_{i=1}^{n} (x_i - y_i)\mathbf{e}_i,$$

and

$$\mathbf{A}(\mathbf{x}-\mathbf{y}) = \sum_{i=1}^{n} (x_i - y_i)\mathbf{A}\mathbf{e}_i = \sum_{i=1}^{n} (x_i - y_i) \sum_{k=1}^{n} A_{ki}\mathbf{e}_k.$$

Hence

$$\|\mathbf{A}(\mathbf{x}-\mathbf{y})\| \le \sum_{i=1}^{n} |x_i - y_i| \sum_{k=1}^{n} |A_{ki}|\,\|\mathbf{e}_k\|$$

by the triangle inequality and the rule $\|\alpha \mathbf{v}\| = |\alpha| \cdot \|\mathbf{v}\|$, applied repeatedly.

Pick a number M so that the *fixed* quantity

$$\sum_{k=1}^{n} |A_{ki}|\,\|\mathbf{e}_k\| \le M, \quad \text{all } i.$$

Then $\|\mathbf{A}(\mathbf{x}-\mathbf{y})\| \le M \sum_{i=1}^{n} |x_i - y_i|$. If $\sum_{i=1}^{n} |x_i - y_i| < \epsilon/M \stackrel{\text{def}}{=} \delta$, then we have $\big|\,\|\mathbf{Ax}\| - \|\mathbf{Ay}\|\,\big| < \|\mathbf{A}(\mathbf{x}-\mathbf{y})\| < \epsilon$. ∎

---

[12] This inequality is one of the more useful consequences of the triangle inequality. One first sees it in calculus, in the form $\big|\,|r| - |s|\,\big| \le |r-s|$. See also Exercise 1.0.11.

Lemma 2. *A continuous function on the set* $\{||\mathbf{x}|| = 1\} \subset \mathbf{R}^n$ *is bounded.* ∎

This is a very special case of the following theorem: a continuous function on a compact set assumes its maximum and its minimum values. The theorem requires ideas we have not yet developed, but the result may be familiar from advanced calculus. The proof is given in II.1.3.8.

We now turn to (2.3.41, 42, 43).

Proof that $N(\mathbf{A}) \geq 0$: because $||\cdot||$ is a norm on vectors, $||\mathbf{Ax}|| \geq 0$. $N(\mathbf{A})$ is the supremum of a certain set of non-negative numbers, and is therefore also non-negative.

Proof that $N(\mathbf{A}) = 0$ iff $\mathbf{A} = 0$: i) Suppose that $N(\mathbf{A}) = 0$. Then all $||\mathbf{Ax}||$ must be zero (for if one of them, say $||\mathbf{Ax}_0||$, were $= a > 0$, then the supremum of $||\mathbf{Ax}||$ would be $\geq a$). Now, $||\mathbf{Ax}|| = 0 \Rightarrow \mathbf{Ax} = 0$: if the norm of the vector $\mathbf{Ax}$ is zero, then the vector $\mathbf{Ax}$ must be zero (remember that $||\cdot||$ is given to be a norm on vectors). So $\mathbf{Ax} = 0$ for all unit vectors ($||\mathbf{x}|| = 1$). Scale x: $\mathbf{A}(\alpha\mathbf{x}) = \alpha(\mathbf{Ax}) = 0$, and so $\mathbf{Ax} = 0$ for *all* vectors. The only matrix with that property has all entries zero, i.e., $\mathbf{A} = 0$. This proves the "only if" part.

ii) Suppose that $\mathbf{A} = 0$. Then $||\mathbf{Ax}|| = 0$ for all $\mathbf{x}$ with $||\mathbf{x}|| = 1$, and one has $\sup_{||\mathbf{x}||=1} ||\mathbf{Ax}|| = 0$. Hence $N(\mathbf{A}) = 0$ — this proves the "if" part.

Proof that $N(\alpha\mathbf{A}) = |\alpha|N(\mathbf{A})$:

$$
\begin{aligned}
N(\alpha\mathbf{A}) &= \sup_{||\mathbf{x}||=1} ||(\alpha\mathbf{A})\mathbf{x}|| = \sup_{||\mathbf{x}||=1} ||\alpha \cdot \mathbf{Ax}|| \\
&= \sup_{||\mathbf{x}||=1} |\alpha| \; ||\mathbf{Ax}|| \quad (||\;|| \text{ is a norm on vectors}) \\
&\overset{*}{=} |\alpha| \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}|| = |\alpha|N(\mathbf{A}).
\end{aligned}
$$

(For the equality marked $\overset{*}{=}$, see the appendix to this subsection).

Proof that $N(\mathbf{A}+\mathbf{B}) \leq N(\mathbf{A}) + N(\mathbf{B})$:

$$
\begin{aligned}
N(\mathbf{A}+\mathbf{B}) &= \sup_{||\mathbf{x}||=1} ||(\mathbf{A}+\mathbf{B})\mathbf{x}|| = \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}+\mathbf{Bx}|| \\
&\leq \sup_{||\mathbf{x}||=1} (||\mathbf{Ax}|| + ||\mathbf{Bx}||) \qquad (*) \\
&\leq \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}|| + \sup_{||\mathbf{x}||=1} ||\mathbf{Bx}|| \qquad (**) \\
&= N(\mathbf{A}) + N(\mathbf{B})
\end{aligned}
$$

The steps marked (*) and (**) should be thought about and understood. In (*), we start with the ordinary triangle inequality for the norm $||\cdot||$. We have

$||\mathbf{Ax} + \mathbf{Bx}|| \leq ||\mathbf{Ax}|| + ||\mathbf{Bx}||$ for every $\mathbf{x}$. Since the left side never exceeds the right side, it should be clear that the supremum of the left cannot exceed the supremum of the right. Step (**) is equally believable: the supremum of a sum is no greater than the sum of the suprema of the individual summands. Such rules are summarized in the Appendix.

This concludes the proof. ∎

Now that $N(\mathbf{A})$ is known to satisfy all the conditions on a norm, we revert to the customary notation, $||\mathbf{A}||$. It will always be clear from the context whether a symbol represents a vector or a matrix, and there should never be any doubt about whether $||\cdot||$ is a vector norm or a matrix norm.

A few of the more important properties of natural matrix norms are derived next.

2.3.45 Proposition. *Let* $\mathbf{I}$ *be the* $n \times n$ *identity matrix, let* $\mathbf{A}$ *be an* $n \times n$ *matrix, let* $||\cdot||$ *be a vector norm on* $\mathbf{R}^n$*, and denote the corresponding natural matrix norm by the same symbol. Then*

a) $||\mathbf{I}|| = 1$;

b) $||\mathbf{Ax}|| \leq ||\mathbf{A}||\,||\mathbf{x}||$ *for all* $\mathbf{A}$ *and* $\mathbf{x}$;

c) If $\mathbf{A}, \mathbf{B}$ *are two* $n \times n$ *matrices, then* $||\mathbf{AB}|| \leq ||\mathbf{A}||\,||\mathbf{B}||$. ∎

Proof. (a) $||\mathbf{I}|| = \sup_{||\mathbf{x}||=1} ||\mathbf{Ix}|| = \sup_{||\mathbf{x}||=1} ||\mathbf{x}|| = \sup_{||\mathbf{x}||=1} 1 = 1.$

(b) If $\mathbf{x} = 0$, then $||\mathbf{Ax}|| = 0$ and $||\mathbf{x}|| = 0$, so the inequality is valid. Assume $\mathbf{x}' \neq 0$. Then $\mathbf{x}'/||\mathbf{x}'||$ is a unit vector:

$$\left\| \frac{\mathbf{x}'}{||\mathbf{x}'||} \right\| = \frac{1}{||\mathbf{x}'||} ||\mathbf{x}'|| = 1$$

(because of the property of norms: $||\alpha \mathbf{x}'|| = |\alpha|\,||\mathbf{x}'||$). But then

$$\frac{||\mathbf{Ax}'||}{||\mathbf{x}'||} = ||\frac{1}{||\mathbf{x}'||}\mathbf{Ax}'|| = ||\mathbf{A}(\frac{\mathbf{x}'}{||\mathbf{x}'||})|| \leq \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}|| = ||\mathbf{A}||.$$

The $\leq$ part is true because $\mathbf{x}'/||\mathbf{x}'|| \stackrel{\text{def}}{=} \mathbf{y}$ has length one, whence

$$||\mathbf{Ay}|| \leq \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}||.$$

Assertion (b) now follows from the first and last terms in this string of equalities and inequalities.

(c) By (2.3.10), $||\mathbf{AB}|| \stackrel{\text{def}}{=} \sup_{||\mathbf{x}||=1} ||(\mathbf{AB})\mathbf{x}||$. Estimate the right-hand side:

$$\|(\mathbf{AB})\mathbf{x}\| = \|\mathbf{A}(\mathbf{Bx})\| \le \|\mathbf{A}\|\,\|\mathbf{Bx}\| \quad \text{by part (b)}$$

$$\le \|\mathbf{A}\| \cdot \|\mathbf{B}\|\,\|\mathbf{x}\| \quad \text{by (b) again}$$

$$= \|\mathbf{A}\| \cdot \|\mathbf{B}\| \quad \text{since } \|\mathbf{x}\| = 1.$$

Therefore

$$\|\mathbf{AB}\| \le \sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\|\,\|\mathbf{B}\| = \|\mathbf{A}\|\,\|\mathbf{B}\|. \blacksquare$$

2.3.46 Corollary. *The condition number,* $\operatorname{cond}(\mathbf{A}) = \|\mathbf{A}\|\,\|\mathbf{A}^{-1}\|$, *of a matrix* $\mathbf{A}$ *satisfies*

$$\|\mathbf{A}\|\,\|\mathbf{A}^{-1}\| \ge 1. \blacksquare$$

Proof. Use (2.3.45), parts a) and c): $\|\mathbf{A}\|\,\|\mathbf{A}^{-1}\| \ge \|\mathbf{A}\mathbf{A}^{-1}\| = \|\mathbf{I}\| = 1. \blacksquare$

• **2.3.47 Exercise.** For a 2×2 matrix $\mathbf{A}$, let λ_1, λ_2 be the eigenvalues (not necessarily distinct), and define $N(\mathbf{A}) = \max\{|\lambda_1|, |\lambda_2|\}$. Is it true that $N(\mathbf{A}+\mathbf{B}) \le N(\mathbf{A}) + N(\mathbf{B})$? That $N(\mathbf{AB}) \le N(\mathbf{A})N(\mathbf{B})$? ∎

• **2.3.48 Exercise.** Let $\|\cdot\|$ be a natural matrix norm on $\mathbf{R}^2$. Show that $\|\mathbf{A}^2\| \le \|\mathbf{A}\|^2$. Is it true that $\|\mathbf{A}^2\| = \|\mathbf{A}\|^2$? ∎

• **2.3.49 Exercise.** Give examples to show that if $\mathbf{A}$ and $\mathbf{B}$ are similar matrices on $\mathbf{R}^2$ (i.e., there is an $\mathbf{S}$ such that $\mathbf{B} = \mathbf{SAS}^{-1}$) it need not be true that $\|\mathbf{A}\|_1 = \|\mathbf{B}\|_1$ or $\|\mathbf{A}\|_\infty = \|\mathbf{B}\|_\infty$. Are the $\|\cdot\|_2$ norms of similar matrices necessarily the same? ∎

2.3.50 Exercise*. Prove that for every natural matrix norm (on $n \times n$ matrices) one has the inequality $\|\mathbf{A}^n\| \le \|\mathbf{A}\|^n$. Suppose that $\|\mathbf{A}\|_\infty < 1$. Show that the series $\mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \dots$ converges, in the sense that each entry of this matrix series is a convergent series of real (or complex) numbers. Show that the sum of this series is $(\mathbf{I} - \mathbf{A})^{-1}$. ∎

2.3.51 Exercise. For what matrices on $\mathbf{R}^n$ is $\|\mathbf{A}\|_2 = \|\mathbf{A}^{-1}\|_2 = 1$?. ∎

2.3.52 Exercise*. Show that

$$\|\mathbf{x}\| \stackrel{\text{def}}{=} \sqrt{\sum_{j=1}^{n} j|x_j|^2}$$

defines a norm on $\mathbf{R}^n$. Compute the natural matrix norm of

$$\begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix}$$

that arises from this vector norm. Give a method for calculating this matrix norm in $\mathbf{R}^n$. (Hint: if $(\cdot,\cdot)$ is the usual inner product on $\mathbf{R}^2$, then one can write $\|\mathbf{x}\|^2 = (\mathbf{Jx}, \mathbf{x})$, where

$$\mathbf{J} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.) \blacksquare$$

2.3.53 Exercise. Suppose **A** is a complex $n \times n$ matrix. What is the analog of Proposition 2.3.29? ∎

Appendix: sup, max, inf, min

The concepts in the title are fundamental, but a bit subtle, and a beginner could easily be persuaded to make incorrect statements. This Appendix presents a short review of these topics. The reader who remembers them from advanced calculus should skip this material.

2.3.54 Notation. *Let S be a subset of the set of real numbers. We write* $\max S$ *for the largest number in S and* $\min S$ *for the smallest number in S.* ∎

There may or may not be a largest or smallest number in S. For example, when $S = (0,1)$ neither $\max S$ nor $\min S$ exist, while if $S = [0,1]$, then $\max S = 1$ and $\min S = 0$. If S is a finite set, $\max S$ and $\min S$ will exist.

2.3.55 Definition. *If a number M satisfies $s \leq M$ for all $s \in S$, M is called an* **upper bound** *for S. If M is an upper bound for S, and if no $M' < M$ is an upper bound for S, we call M the* **least upper bound** *and write $M = \text{lub}\, S$ or, more frequently, $M = \sup S$. If there is no M satisfying $s \leq M$ for all $s \in S$, we define* $\sup S$ *to be $+\infty$. (In words: S is not bounded on the right).* ∎

2.3.56 Proposition. *Suppose that* $\sup S < \infty$. *Then*

$$M = \sup S \text{ iff: } i) \text{ for all } s \in S,\ s \leq M$$
$$ii) \text{ for any } \epsilon > 0, \text{ there is an } s \in S \text{ such that } M - \epsilon < s. \blacksquare$$

• **2.3.57 Exercise.** Prove this result. ∎

2.3.58 Example. $S = \{1-1/2, 1-1/3, \ldots, 1-1/n, \ldots\}$. It is true that $1-1/n < 2$ for any $n \geq 2$, but 2 is not the smallest number with this property. So $\sup S \neq 2$. It is also true that $1 - 1/n < 1$ for any $n \geq 2$. But now, for any $\epsilon > 0$, there will be an n so that $1 - \epsilon < 1 - 1/n < 1$. Hence no number smaller than 1 is an upper bound, and $\sup S = 1$. ∎

Now we establish some useful properties of sup.

2.3.59 Proposition. *For $S \subset \mathbf{R}$ and $\alpha > 0$, write $\alpha S = \{\alpha s \mid s \in S\}$. Then, if* $\sup S$ *is finite,* $\sup \alpha S = \alpha \sup S$. ∎

Proof. To show that two numbers are equal, one often proves that each is $\geq$ the other. We show:

$$i) \quad \sup \alpha\, S \leq \alpha \sup S$$

and

$$ii)\quad \alpha \sup S \leq \sup \alpha S.$$

Equality then follows.

i) $s \leq \sup S$ for all $s \epsilon S$, by definition of $\sup S$. Since $\alpha > 0$, this implies $\alpha s \leq \alpha \sup S$ for all $s \in S$. So every element of αS is $\leq \alpha \sup S$. Hence $\alpha \sup S$ is an upper bound for αS. The *least* upper bound for αS is $\sup \alpha S$, by definition. No upper bound for αS can be less than $\sup \alpha S$; hence,

$$\sup \alpha S \leq \alpha \sup S.$$

ii) Since $\sup \alpha S$ is an upper bound for αS, every element of αS is $\leq \sup \alpha S$, or

$$\alpha s \leq \sup \alpha S \ , \ \text{all } s \in S.$$

Equivalently,

$$s \leq \frac{1}{\alpha} \sup \alpha S \ , \ \text{all } s \in S.$$

So, $(\sup \alpha S)/\alpha$ is an upper bound for S. It is no smaller than the least upper bound for S, $\sup S$. Therefore

$$\sup S \leq \frac{1}{\alpha} \sup \alpha S$$

or

$$\alpha \sup S \leq \sup \alpha S. \blacksquare$$

The next two results are for special types of sets.

Let I be a set ("index set"), and let f be any function from I to R. For our set S, we take $S = \{f(x) | x \in I\}$, and write

$$\sup_{x \in I} f(x) \quad \text{instead of} \quad \sup S.$$

For example, I could be the set $\{\mathbf{x} \mid ||\mathbf{x}|| = 1\}$, and $f(\mathbf{x})$ could be $||\mathbf{Ax}||$. Then

$$\sup_{\mathbf{x} \in I} f(\mathbf{x}) \ \text{ is the same as } \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}||.$$

2.3.60 Proposition. *If f and g are two functions from I to $\mathbf{R}$ satisfying $f(x) \leq g(x)$ for all $x \in I$, then*

$$\sup_{x \in I} f(x) \leq \sup_{x \in I} g(x). \blacksquare$$

Proof. For all $x' \in I$, $f(x') \leq g(x') \leq \sup_{x \in I} g(x)$. So, $\sup_{x \in I} g(x)$ is an upper bound for $\{f(x') | x' \in I\}$. The least upper bound, $\sup_{x \in I} f(x)$, is $\leq$ any other upper bound. Hence,

$$\sup_{x \in I} f(x) \leq \sup_{x \in I} g(x). \blacksquare$$

Remark. With $f(\mathbf{x}) = \|\mathbf{Ax} + \mathbf{Bx}\|$ and $g(\mathbf{x}) = \|\mathbf{Ax}\| + \|\mathbf{Bx}\|$, this result gives step (*) in the proof of Proposition 2.3.40. ∎

2.3.61 Proposition. *For any $f, g : I \to \mathbf{R}$,*

$$\sup_{x \in I} f(x) + g(x) \leq \sup_{x \in I} f(x) + \sup_{x \in I} g(x). \blacksquare$$

Proof. For any x', $f(x') \leq \sup_{x \in I} f(x)$ and $g(x') \leq \sup_{x \in I} g(x)$, and so $f(x') + g(x') \leq \sup_{x \in I} f(x) + \sup_{x \in I} g(x)$. The right side in the last inequality is therefore an upper bound for the set $\{f(x') + g(x') | x' \in I\}$, and the least upper bound, $\sup_{x \in I} f(x) + g(x)$, will be $\leq$ any other upper bound. ∎

Remarks

A. Proposition 2.3.61 gives step (**) in the proof of Proposition 2.3.40.

B. Most of the time, $\sup_{x \in I} f(x) + g(x) < \sup_{x \in I} f(x) + \sup_{x \in I} g(x)$. For example, let $I = \{1, 2, 3, \ldots\}$ and write "n" instead of x. Let

$$f(n) = 1 - \frac{1}{n}, \quad g(n) = \frac{1}{n}.$$

Then

$$\sup_{n \in I}(1 - \frac{1}{n}) + (\frac{1}{n}) = \sup_{n \in I} 1 = 1,$$

while

$$\sup_{n \in I}(1 - \frac{1}{n}) = 1, \sup_{n \in I} \frac{1}{n} = 1. \blacksquare$$

The definition of "inf" (= greatest lower bound) and proofs of properties of inf are left for the exercises.

• **2.3.62 Exercise.** Let $S \subset \mathbf{R}$. Define inf S. ∎

• **2.3.63 Exercise.** Suppose that $\inf S > -\infty$. Prove that $m = \inf S$ iff i) $s \geq m$ for every $s \in S$, and ii) for every $\epsilon > 0$ there is an $s \in S$ such that $s < m + \epsilon$. ∎

• **2.3.64 Exercise.** Suppose that $\inf S > -\infty$, and let $\alpha > 0$. Show that $\inf \alpha S = \alpha \inf S$. What happens when $\alpha < 0$? ∎

• **2.3.65 Exercise.** If $f(x) \leq g(x)$ for all $x \in I$, what is the relation between

$$\inf_{x \in I} f(x) \quad \text{and} \quad \inf_{x \in I} g(x)? \ ∎$$

• **2.3.66 Exercise.** What is the relation between $\inf_{x \in I} f(x) + g(x)$ and

$$\inf_{x \in I} f(x), \quad \inf_{x \in I} g(x)? \ ∎$$

2.3.67 Exercise*. Find $\sup\{\sum_{j=1}^{n} 1/j^2 \mid n = 1, 2, 3, \ldots\}$. ∎

Subsection 2.4 Application: Differential Equations

Reference has been made several times to the importance of certain measures of distance. In this subsection, we try to back up those claims with some typical applications of metrics in the theory of differential equations.

To introduce the issues, we consider the initial value problem for a system of ordinary first-order differential equations,

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}), \quad \mathbf{x}(0) = \mathbf{x}^*. \tag{2.4.1}$$

Here $\mathbf{x}(t) = x_1(t), \ldots, x_n(t))$, $\mathbf{f}$ is a vector of functions, and $\mathbf{x}^* \in \mathbf{R}^n$ is given, so that (2.4.1) is an abbreviation for

$$\begin{aligned} \dot{x}_1(t) &= f_1(x_1, \ldots, x_n) \\ &\vdots \\ \dot{x}_n(t) &= f_n(x_1, \ldots, x_n), \end{aligned}$$

subject to

$$x_1(0) = x_1^*, \ldots, x_n(0) = x_n^*.$$

There are three basic questions in the theory of the initial value problem (2.4.1):

i) Does a solution exist?

ii) If it exists, is it unique?

iii) How does it depend on the initial value $\mathbf{x}^0$? For example, if one solves (2.4.1) for two different initial values $\mathbf{x}^*$ and $\mathbf{x}^{**}$, and if $\|\mathbf{x}^* - \mathbf{x}^{**}\|$ is small in a certain norm, will the difference between the solutions also remain small? For all time t, or if not, for how long?

Questions ii) and iii) afford the most direct illustration of concepts from the theory of metric spaces, and this subsection deals only with them (the existence question i) is discussed in Section II.3.1). It should be said at the outset that there are no general results about the *stability* problem iii); some differential equations are very stable — a small change in initial condition hardly changes the solution — while others are extremely sensitive to variations in the initial values. Whether or not the distance between solutions can be estimated in terms of the distance between initial values depends strongly on the form of the function $\mathbf{f}$, and if such an estimate is possible, the most convenient metric is also dictated by $\mathbf{f}$. Often, the correct "distance" measures a quantity of physical interest: energy difference, power, dissipation rate, ... As for the uniqueness question ii), there is a quite general uniqueness theorem guaranteeing that (for reasonable $\mathbf{f}$) the problem (2.4.1) has only one solution *as long as* $|t|$ *is small*. For special types of differential equations, one can establish uniqueness of the solution for *all* t by some simple estimates involving appropriate norms. An example is presented later.

A word to the Practical Person. Practical Persons only study differential (or other) equations that arise in a Practical Problem (from physics, engineering, chemistry, ...). Such problems, experience suggests, do have a solution, and just one solution at that. Existence and uniqueness results are therefore thought to be unnecessary luxuries. This is a reasonable point of view, to some extent, but I offer two warnings. First, a good argument for uniqueness can suggest good conclusions about stability. Stability means: if initial values are close, solutions stay close. It is usually possible to deduce that, in particular, the solutions are exactly the same when the initial values are exactly the same. Thus, the techniques used to study problems ii) and iii) are often related, and stability (of trajectories, numerical solutions, ...) is certainly a matter of concern to Practical Persons. Second, questions i), ii), iii) are equally basic in the theory of partial differential equations, and there the physical problems are often so complicated that intuition may need help from mathematics. There are examples of problems with nonunique solutions for which different numerical schemes produce (distinctly) different numerical solutions. It is not always known why certain schemes pick out certain solutions; further progress will only come from a better understanding of the uniqueness question ii).

This subsection will deal with two (related) differential equations. One is a system of ordinary differential equations, describing the small vibrations of a "beaded string", a tightly stretched string carrying a finite number of point masses ("beads") at fixed positions. The other is a partial differential equation modelling the continuum limit of the beaded string, in which the beads merge together to form a continuous mass distribution. This equation is the classical *wave equation*. The limiting process will show, once again, the passage from $\mathbb{R}^n$ to L^2. Some other equations, both ordinary and partial, are treated in supplements and exercises.

Example. *The Beaded String*

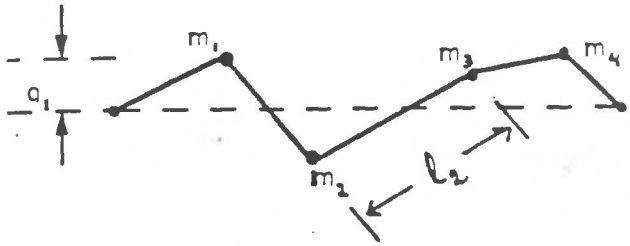


Figure 24. *The beaded string*

We have a string, stretched tight by application of a constant tension T and tied down at both ends, and beads of masses $m_1, \ldots, m_n$ fixed at certain positions on the string.[13] The $n+1$ segments between the masses, or between mass and endpoint, have lengths $\ell_0, \ldots, \ell_n$. We want to study *small* oscillations of the beads in the direction perpendicular to the rest position of the string (the dotted line in the figure). *Small oscillation* means: it is assumed that the beads have no horizontal motion, and the string cannot stretch: the ℓ_j are fixed. Let q_j be the displacement of mass j in the vertical direction. According to Newton's third law, $F = ma$, the differential equation governing the motion of the j^{th} bead has the form

$$m_j \ddot{q}_j(t) = \text{ force acting on bead } j.$$

We compute the right-hand side.

The forces acting on bead j are T^l and T^r, directed, respectively, along the left and right segments of string connecting bead j to beads $j-1$ and $j+1$. The left-pulling tension T^l has vertical and horizontal components T^l_V and T^l_H shown in the figure below; the right-pulling tension has components T^r_V and T^r_H (not shown). If the bead has no horizontal motion, as is assumed, it must be that $T^l_H = T^r_H$. Denote the common value by T_j.

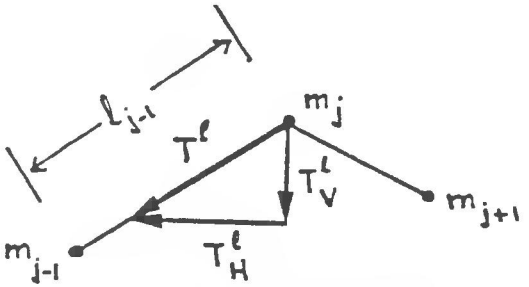


Figure 25. *Forces acting on a bead*

---

[13] This particular problem, according to the book *Oscillation Matrices, Oscillation Kernels, and Small Vibrations of Mechanical Systems*, by F. R. Gantmakher and M. G. Kreĭn, led d'Alembert to his famous solution of the wave equation and D. Bernoulli to the introduction of trigonometric (=Fourier) series into analysis.

Similar triangles in the left half of Figure 25 produce

$$\frac{T_V^l}{T_j} = -\frac{q_j - q_{j-1}}{\ell_{j-1}}; \tag{2.4.2}$$

the minus sign arises because the vertical component of the left-pulling tension points downwards if $q_{j+1} > q_j$ (and upwards if $q_{j+1} < q_j$). Likewise,

$$\frac{T_V^r}{T_j} = -\frac{q_j - q_{j+1}}{\ell_j}. \tag{2.4.3}$$

The total vertical force on bead j is

$$T_V^l + T_V^r = T_j \left(\frac{q_{j-1}}{\ell_{j-1}} - (\frac{1}{\ell_{j-1}} + \frac{1}{\ell_j}) q_j + \frac{q_{j+1}}{\ell_j} \right). \tag{2.4.4}$$

Finally, one argues that since the vertical displacement is assumed to be small, the vertical forces T_V^l, T_V^r must also be small, so that the horizontal tension T_j is practically equal to the overall constant tension T. We therefore replace T_j in (2.4.4) by T. If this sounds hokey, think about it as applied mathematicians do: Let the angle between the horizontal and the $\overline{q_{j-1}q_j}$ segment of string be ϵ, then $T_V^l = T^l \sin\epsilon \sim \epsilon T^l$, while $T_H^l = T^l \cos\epsilon \sim T^l(1 - \epsilon^2/2)$. We keep terms of order ϵ, but drop terms of order ϵ^2 or higher.

At any rate, the equations for the small-amplitude oscillations of the beaded string become:

$$m_j \ddot{q}_j = \frac{T}{\ell_{j-1}} q_{j-1} - (\frac{T}{\ell_{j-1}} + \frac{T}{\ell_j}) q_j + \frac{T}{\ell_j} q_{j+1}. \tag{2.4.5}$$

When $j = 0$ and $j = n + 1$, it is understood that $q_0 = q_{n+1} = 0$. This expresses the fact that the string is tied down at the ends; one should picture two additional beads at the two ends of the string, bead 0 and bead $n + 1$, which are prevented from moving, so that their displacements q_0, q_{n+1} are always zero.

The system (2.4.5) is of second, rather than first, order, but that is a minor point: one could (and in physics, does) introduce n more variables by defining $p_j = m_j \dot{q}_j$. Then (2.4.5) turns into a system of $2n$ first-order equations,

$$\begin{aligned} \dot{q}_j &= \frac{p_j}{m_j}, \\ \dot{p}_j &= \frac{T}{\ell_{j-1}} q_{j-1} - (\frac{T}{\ell_{j-1}} + \frac{T}{\ell_j}) q_j + \frac{T}{\ell_j} q_{j+1}, \end{aligned}$$

for $j = 1, \ldots, n$. More important, the system is *linear* in the unknown functions q_j (thanks to the assumed smallness of the oscillations), and it is what physicists call

a *conservative system*. This means that friction has been hypothesized away — once the beaded string is set into motion, it will never stop. Its energy is conserved, and in fact the energy (kinetic plus potential) will define the metric we will use to draw conclusions about the solutions of equations (2.4.5). To this end, we next find an expression for the total energy in the beaded string.

The energy of a mechanical system comes in two parts: *kinetic energy*, or energy due to motion, and *potential energy*, or energy due to position. The kinetic energy of a point mass moving at velocity v is $mv^2/2$, so the total kinetic energy of the beads is

$$\text{K.E.} = \frac{1}{2}\sum_{j=1}^{n} m_j \dot{q}_j^2. \tag{2.4.6}$$

The potential energy of a point mass acted upon by a force is, by definition, the work it takes to move the mass from an arbitrary but fixed reference position to its present position. "Work" has the technical meaning "force exerted × distance moved" if the force is constant; if the force varies with position (as it usually does), *force exerted* $= \tilde{F}(x)$, then, for motion along a straight line,

$$\text{work} = \int_{x_0}^{x} \tilde{F}(x')\,dx'. \tag{2.4.7}$$

One can run (2.4.7) backwards as well: if $V(x)$ is the work it takes to move the mass from the reference position to point x, then

$$\tilde{F}(x) = \frac{\partial V(x)}{\partial x}.$$

In mechanics problems, it is customary to work not with the force $\tilde{F}$ exerted to move a particle to some point, but instead with the force F that acts on the particle and opposes the effort to move it. If F (gravity, for example) acts on the particle, then I exert an equal but opposite force, $\tilde{F} = -F$, to overcome the resistance. Hence

$$F(x) = -\frac{\partial V(x)}{\partial x}. \tag{2.4.8}$$

This all carries over to systems consisting of several masses. If the potential energy (*alias* work) of an n-particle system is $V(q_1, \ldots, q_n)$, then the force on particle # j is

$$F_j(q_1, \ldots, q_n) = -\frac{\partial V(q_1, \ldots, q_n)}{q_j}.$$

Since we know the force acting on bead j from (2.4.5), we can compute the potential energy V:

$$\text{P.E.} = V(q_1, \ldots, q_n) = \frac{T}{2}\sum_{j=1}^{n} \frac{1}{\ell_j}(q_{j+1} - q_j)^2. \tag{2.4.9}$$

There is a constant of integration that has been set to zero; this amounts to choosing the reference position as the horizontal equilibrium position of the string.

Further manipulations are best done in matrix notation. Let

$$\mathbf{q} = (q_1, \ldots, q_n)^t, \quad \dot{\mathbf{q}} = (\dot{q}_1, \ldots, \dot{q}_n)^t$$

be the (column) position and velocity vectors of the beads. Let M be the diagonal matrix with entries $m_1, \ldots, m_n$, and let L be the negative of the symmetric matrix defined by the right-hand side of (2.4.5). Its diagonal entries are $L_{jj} = T(1/\ell_{j-1} + 1/\ell_j)$, the entries $L_{j,j+1}$ on the first superdiagonal are $-T/\ell_j$, and all other entries above the diagonal are zero:

$$\mathbf{L} = \begin{pmatrix} L_{11} & L_{12} & 0 & \cdots & 0 \\ L_{12} & L_{22} & L_{23} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \ddots & L_{n-1,n-1} & L_{n-1,n} \\ 0 & \cdots & \cdots & L_{n-1,n} & L_{nn} \end{pmatrix}. \tag{2.4.10}$$

The initial value problem for the beaded string now takes the form

$$\mathbf{M}\ddot{\mathbf{q}}(t) = -\mathbf{L}\mathbf{q}(t), \quad \mathbf{q}(0) = \mathbf{q}_\bullet, \ \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}_\bullet. \tag{2.4.11}$$

The vectors $\mathbf{q}_\bullet, \dot{\mathbf{q}}_\bullet$ list, respectively, the initial positions and velocities of the n beads.

Now write the total energy of the beaded string in matrix form. From (2.4.6), it is clear that

$$\text{K.E.} = \frac{1}{2}(\mathbf{M}\dot{\mathbf{q}}, \dot{\mathbf{q}}), \tag{2.4.12}$$

where $(\cdot, \cdot)$ is the usual inner product on $\mathbf{R}^n$. The potential energy is

$$\text{P.E.} = V(\mathbf{q}) = \frac{T}{2}(\mathbf{L}\mathbf{q}, \mathbf{q}); \tag{2.4.13}$$

the conversion of (2.4.9) into (2.4.13) is a simple calculation left to the reader. Note that $-\bar{\nabla}V(\mathbf{q})$ is precisely the (vector) right-hand side of (2.4.5); this generalizes the relation (2.4.8) between potential energy and acting force.[14]

Define the *total energy of a position-velocity state* $(\mathbf{q}, \dot{\mathbf{q}})$ to be

$$E(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}(\mathbf{M}\dot{\mathbf{q}}, \dot{\mathbf{q}}) + \frac{1}{2}(\mathbf{L}\mathbf{q}, \mathbf{q}). \tag{2.4.14}$$

---

[14] Compare also Lemma 2.3.28, which asserts that the gradient of the quadratic form $(\mathbf{B}\mathbf{x}, \mathbf{x})/2$ equals the vector $\mathbf{B}\mathbf{x}$.

2.4.15 Proposition. *The function*

$$\|(\mathbf{q}, \dot{\mathbf{q}})\|_E \stackrel{\text{def}}{=} \sqrt{E(\mathbf{q}, \dot{\mathbf{q}})}$$

is a norm on $\mathbf{R}^{2n}$. *(It is called the energy norm.)* ∎

Proof. We need to verify conditions (2.2.2,3,4). Consider (2.2.2) first. It is clear from (2.4.6) and (2.4.9) that the total energy is non-negative; it is zero precisely when $\dot{q}_j = 0$ for all j and $q_j = q_{j+1}$ for all j $(= 1, \ldots, n)$. Since q_{n+1}, the right end of the string, is zero, all the q_j must be zero.

The scaling condition (2.2.3) is obviously true. The square root in the formula for the norm makes the scaling work.

The easiest way to get at Minkowski's inequality is with a bit of linear algebra. (If the bit is not familiar, you deserve the punishment of having to find another proof.) Since M and L are both positive definite symmetric matrices, they have (positive definite) square roots. Define

$$\dot{\mathbf{Q}} = \frac{1}{\sqrt{2}}\sqrt{\mathbf{M}}\dot{\mathbf{q}}, \quad \mathbf{Q} = \frac{1}{\sqrt{2}}\sqrt{\mathbf{L}}\mathbf{q}.$$

The energy norm now reads

$$\|(\mathbf{Q}, \dot{\mathbf{Q}})\|_E = \sqrt{(\dot{\mathbf{Q}}, \dot{\mathbf{Q}}) + (\mathbf{Q}, \mathbf{Q})},$$

which is nothing but the standard Euclidean (or l^2) norm on $\mathbf{R}^{2n}$. It satisfies the Minkowski inequality, and, by changing back to $(\mathbf{q}, \dot{\mathbf{q}})$, one verifies that the inequality also holds in those variables. ∎

Now comes the point of the preceding definition.

2.4.16 Proposition. *The energy of a solution of the beaded string equation is constant in time,*

$$E(\mathbf{q}(t), \dot{\mathbf{q}}(t)) = E(\mathbf{q}(0), \dot{\mathbf{q}}(0)).$$

Put differently, the energy norm of a solution does not change with time. ∎

Proof. Differentiate (2.4.16) with respect to t:

$$\frac{d}{dt}E(\mathbf{q}(t), \dot{\mathbf{q}}(t)) = (\mathbf{M}\ddot{\mathbf{q}}(t), \dot{\mathbf{q}}(t)) + (\mathbf{L}\mathbf{q}(t), \dot{\mathbf{q}}(t)) = ([\mathbf{M}\ddot{\mathbf{q}}(t) + \mathbf{L}\mathbf{q}(t)], \dot{\mathbf{q}}(t)) = 0:$$

the matrix differential equation (2.4.11) was used in the last step. ∎

2.4.17 Corollary. *Let $q(t), q_*(t)$ be two solutions of the beaded string equation. Then*

$$E(q(t) - q_*(t), \dot{q}(t) - \dot{q}_*(t)) = E(q(0) - q_*(0), \dot{q}(0) - \dot{q}_*(0)).$$

In particular, the distance in the energy norm between two solutions is independent of time t. ∎

2.4.18 Corollary. *If the initial value problem (2.4.11) has a solution at all, that solution is unique.* ∎

Proof. Suppose there are two solutions $q(t), q_*(t)$ with initial position q^0 and initial velocity $\dot{q}^0$. Then by Corollary 2.4.17,

$$E(q(t) - q_*(t), \dot{q}(t) - \dot{q}_*(t)) = E(q^0 - q^0, \dot{q}^0 - \dot{q}^0) = 0.$$

Since $E(\cdot, \cdot)$ is the square of a norm, it can be zero only if the argument is zero: for every t,

$$q(t) = q_*(t) \quad \text{and} \quad \dot{q}(t) = \dot{q}_*(t). \ ∎$$

These two corollaries answer, respectively, question ii) (uniqueness) and question iii) (stability) from the beginning of this subsection. The argument was pretty simple from the mathematics point of view, and absolutely obvious from the physics point of view. The key was the correct choice of norm in Proposition 2.4.15.

Remarks

A. The energy norm (2.4.15) on $\mathbf{R}^{2n}$ is equivalent to all the l^p norms, and it should therefore be possible to prove a stability and uniqueness result in, say, the l^∞ norm. It would also be absolute lunacy to attempt this — I won't even put it into the exercises. The beaded string equations illustrate perfectly how the correct choice of one of many equivalent metrics can make life (and computations) pleasant.

B. Question i), existence of a solution to the initial value problem, is easily answered: one can give explicit formulas for the solution (see Exercise 2.4.37). If I can write it down, it must exist. ∎

Example. *Wave Equation*

This example treats the continuum limit of the beaded string. Organize the equations (2.4.5) (repeated here)

$$m_j \ddot{q}_j = \frac{T}{\ell_{j-1}} q_{j-1} - \left(\frac{T}{\ell_{j-1}} + \frac{T}{\ell_j}\right) q_j + \frac{T}{\ell_j} q_{j+1}$$

in the more suggestive form

$$\frac{1}{T}\ddot{q}_j = \frac{q_{j+1}-q_j}{m_j \ell_j} - \frac{q_j - q_{j-1}}{m_j \ell_{j-1}}. \tag{2.4.19}$$

Suppose that $m_j = m/n$ and $\ell_j = 1/(n+1)$: the n masses are all equal, with total mass m, and they are evenly spaced along the string. Equation (2.4.19) then becomes

$$\frac{m}{T}\ddot{q}_j = \frac{n(n+1)}{n^2}\,\frac{q_{j+1}-2q_j+q_{j-1}}{1/n^2}. \tag{2.4.20}$$

When taking the continuum limit, one thinks of q_j as the value $q(j/n)$ of a funtion $q(x)$. By Taylor's formula,

$$q_{j+1} = q(\frac{j}{n}+\frac{1}{n}) = q(\frac{j}{n}) + \frac{1}{n}\,q_x(\frac{j}{n}) + \frac{1}{2n^2}\,q_{xx}(\frac{j}{n}) + \dots. \tag{2.4.21}$$

where subscript x denotes the partial derivative with respect to x (the displacement q also depends on time t, which was omitted from this formula). There is a similar formula for q_{j-1}. Substituting both of those into (2.4.20), one finds, to leading order in $1/n$:

$$\frac{m}{T}\frac{\partial^2 q(x,t)}{\partial t^2} = \frac{\partial^2 q(x,t)}{\partial x^2}. \tag{2.4.22}$$

Equation (2.4.22) is known as the *Wave Equation*. It governs the small (transverse) oscillations of a homogeneous[15] string, of total mass m, subject to a constant tension T. The ends are tied down. One should give the string an initial displacement $q(0,x) = f(x)$ and an initial velocity $q_t(0,x) = g(x)$; subscript t denotes the partial t-derivative. The *initial/boundary value problem for the wave equation* (abbreviated IBVP) on the interval $0 \le x \le 1$ is then posed as follows:

$$\begin{aligned} &q_{tt}(t,x) = q_{xx}(t.x),\ x \in [0,1] \\ &q(0,x) = f(x),\ q_t(0,x) = g(x),\ x \in [0,1]. \\ &q(t,0) = q(t,1) = 0,\ t \ge 0 \end{aligned} \tag{2.4.23}$$

The energy norm for this IBVP is the continuum limit of the energy norm (2.4.15) for the beaded string: the kinetic energy is

$$\frac{m}{2}\sum_{j=1}^{n}\frac{1}{n}\dot{q}_j^2 \longrightarrow \frac{m}{2}\int_0^1 q_t(t,x)^2\,dx;$$

[15] If the masses m_j are allowed to vary, the continuum limit will describe the small oscillations of an inhomogeneous string. See Exercise 2.4.33.

the potential energy is

$$\frac{T}{2}\sum_{j=1}^{n}\frac{(q_{j+1}-q_j)^2}{1/(n+1)} \longrightarrow \frac{T}{2}\int_0^1 q_x(t,x)^2\,dx.$$

The total energy of the string at time t is

$$E(q(t,x),q_t(t,x)) \stackrel{\text{def}}{=} \frac{1}{2}\int_0^1 mq_t(t,x)^2 + Tq_x(t,x)^2\,dx. \tag{2.4.24}$$

The energy in the continuous string is constant, as it was for the beaded string:

2.4.25 Proposition. *The energy (2.4.24) is independent of time. Let W be the space of pairs (f,g) of continuously differentiable functions on $[0,1]$, with f satisfying $f(0)=f(1)=0$. Define the energy norm*

$$\|(f,g)\|_E = \sqrt{\frac{1}{2}\int_0^1 f'(x)^2 + g(x)^2\,dx}.$$

This is indeed a norm on W, and

$$\|(q(t,\cdot),q_t(t,\cdot))\|_E$$

is independent of time t when $q(t,x)$ solves the IBVP (2.4.23). ∎

Remark. The dot $\cdot$ in place of the variable x indicates that the latter is integrated out in the computation of the energy norm. ∎

Proof. To show that E is constant in t, one proves that $dE/dt = 0$. The technique is used all the time in partial differential equations. To simplify the notation, I will write $E(t)$ rather than $E(q_x(t,x),q_t(t,x))$. Furthermore, I take $m = T = 1$: this does not affect the arguments.

$$\frac{dE(t)}{dt} = \frac{1}{2}\int_0^1 2\cdot q_{tt}(t,x)q_t(t,x) + 2\cdot q_{xt}(t,x)q_x(t,x)\,dx.$$

In the first term of this integral, replace q_{tt} by q_{xx}, using the wave equation. The second part of the integral is transformed by integration by parts:

$$\int_0^1 q_{xt}(t,x)q_x(t,x)\,dx = q_t(t,x)\cdot q_x(t,x)\Big|_{x=0}^{x=1} - \int_0^1 q_t(t,x)\cdot q_{xx}(t,x)\,dx.$$

Look at the boundary term, $q_t q_x\big|_0^1$. Since $q(t,0) = 0$ for all t, we have $q_t(t,0) = 0$, and likewise $q_t(t,1) = 0$. The boundary term vanishes. The leftover terms give

$$\frac{dE(t)}{dt} = \int_0^1 q_{xx}(t,x)q_t(t,x) - q_t(t,x)q_{xx}(t,x)\,dx = 0,$$

as claimed.

Next, we verify the conditions (2.2.2,3,4) that must be satisfied by a norm. Non-negativity is obvious. Suppose now that $\|(f,g)\|_E = 0$. Then

$$\int_0^1 f'(x)^2 + g(x)^2\,dx = 0,$$

so that $f'^2 + g^2 \equiv 0$. Hence $f' \equiv 0$ and $g \equiv 0$. The function f is therefore constant, but since $f(0) = 0$, that constant must be zero. (Note how the boundary condition is used.)

The scaling property of norms, (2.2.3), is clearly satisfied. The Minkowski inequality (2.2.4) follows from Exercise 1.4.16. ∎

2.4.26 Corollary. *Let $q(t,x), q^*(t,x)$ be two solutions of the IBVP (2.4.23) for the wave equation. Then*

$$\|(q(t,\cdot) - q^*(t,\cdot), q_t(t,\cdot) - q_t^*(t,\cdot))\|_E = \|(q(0,\cdot) - q^*(0,\cdot), q_t(0,\cdot) - q_t^*(0,\cdot))\|_E.$$ ∎

2.4.27 Corollary. *If a solution to (2.4.23) with given initial functions $q(0,x) = f(x), q_t(0,x) = g(x)$ exists at all, then it is unique.* ∎

Proof. Since the initial values of the two solutions are equal, the energy norm of $(q - q^*, q_t - q_t^*)$ is zero. Hence $q_x - q_x^* \equiv 0$, which means that $q(t,x) = q^*(t,x)+$ a function $c(t)$ of t alone. But because $q(t,0) \equiv 0$, $c(t)$ is identically zero, and the solutions q and q^* are the same. ∎

Existence of a solution to the initial/boundary value problem (2.4.23) is shown, again, by exhibiting explicit formulas. The method of *Fourier series* gives the most convenient expression for the solution. One assumes that

$$q(t,x) = \sum_{n=1}^{\infty} a_n(t) \sin n\pi x,$$

and substitutes this into the wave equation $q_{tt} = q_{xx}$ to find

$$\ddot{a}_n(t) = -n^2\pi^2 a_n(t).$$

Hence

$$a_n(t) = \alpha_n \sin n\pi t + \beta_n \cos n\pi t.$$

The initial conditions

$$q(0,x) = \sum_{n=1}^{\infty} a_n(0) \sin n\pi x = f(x)$$

$$q_t(0,x) = \sum_{n=1}^{\infty} \dot{a}_n(0) \sin n\pi x$$

determine α_n and β_n. For details, see Chapter 3. ∎

There are problems other than the particular initial/boundary value problem (2.4.23) that can be posed for the vibrating string. The string might be infinitely long, tied down at one end ($x = 0$), or it might cover the whole x-axis, with tension being supplied by mysterious little people pulling at $x = \pm\infty$, or the ends might be free to slide up and down a pair of rods at $x = 0, 1$, and so forth. These are all (fairly simple) models for interesting physical processes. While the example has used the setting of a vibrating string, the wave equation describes many other phenomena: small amplitude water waves propagating along a channel, small amplitude sound waves propagating in one direction, and so forth. The variations are left for the exercises.

• **2.4.28 Exercise.** Prove uniqueness of the solution of the initial value problem for the wave equation on the whole line:

$$\begin{aligned} &q_{tt}(t,x) = q_{xx}(t,x), \quad t \geq 0, x \in \mathbf{R}, \\ &q(0,x) = f(x),\ q_t(0,x) = g(x), \quad x \in \mathbf{R}, \end{aligned} \tag{2.4.29}$$

assuming that all functions in question belong to $L_0^2(\mathbf{R})$ as function of x and, together with their derivatives, have limit zero as $|x| \to \infty$. ∎

• **2.4.30 Exercise.** Verify that the solution of (2.4.29) is given by *d'Alembert's formula*,

$$q(t,x) = \frac{f(x+t) + f(x-t)}{2} + \frac{1}{2} \int_{x-t}^{x+t} g(s)\, ds.$$

Note that the availability of an explicit formula does not in itself guarantee that this known solution is the *only* solution. ∎

2.4.31 Exercise. Change the boundary condition in (2.4.23) to $q_x(t,0) = q_x(1,0) = 0$, $t \geq 0$. Do uniqueness and stability still hold? (In this example, the energy "norm" is no longer a norm, but one can still draw conclusions about uniqueness and stability.) ∎

2.4.32 Exercise*. Suppose I wiggle the left end, $x = 0$, of a semi-infinite string ($0 \leq x < \infty$) periodically in time: $q(t,0) = \sin t$. Is the solution of the wave equation,

$$q_{tt}(t,x) = q_{xx}(t,x), \quad 0 \leq x < \infty, \ 0 \leq t$$

with initial conditions

$$q(0, x) = f(x), \quad q_t(0, x) = g(x)$$

unique? If the initial data are close, are the solutions close (by what measure of "closeness")? (Note that the set of solutions of the wave equation with wiggled left end is no longer a linear space. One uses ideas *suggested* by norms, without actually *having* a norm in the official sense.) ∎

2.4.33 Exercise*. Let $m(x), 0 \leq x \leq 1$, be a continuously differentiable positive function. Suppose the beaded string carries n equally spaced masses $m_j = m(j/n)$. By taking the limit $n \to \infty$, find the inhomogenous wave equation governing the small oscillations of a string carrying the mass distribution $m(x)$. Find the energy, define an energy norm, and investigate uniqueness and stability. ∎

• **2.4.34 Exercise.** The simplest harmonic oscillator is described by the ordinary differential equation $\ddot{q}(t) = -q(t)$. The function q gives the displacement from equilibrium of a unit mass attached to the end of a spring with Hooke's constant = 1 (see Figure 26). The *kinetic energy* is $mv^2/2 = \dot{q}^2/2$. The *potential energy* is $q^2/2$. The total energy is

$$E(t) \stackrel{\text{def}}{=} \frac{1}{2}\dot{q}(t)^2 + \frac{1}{2}q(t)^2.$$

Find the equation for $q(t)$. Introduce an energy norm on $\mathbf{R}^2 = \{(q, \dot{q})\}$, and show that it is constant along solutions of the spring equation. ∎

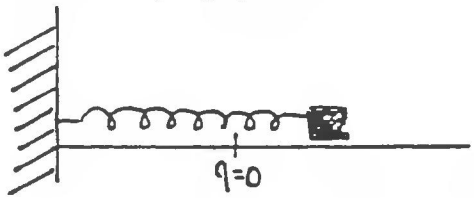


Figure 26. *Mass on a spring*

• **2.4.35 Exercise.** Suppose that the mass in the preceding problem is subject to a frictional force proportional to the velocity, $F_{\text{friction}}(q, \dot{q}) = -c\dot{q}$, with $c > 0$. Show that energy *decreases* along solutions. Use the energy norm to prove uniqueness and stability. ∎

2.4.36 Exercise*. Consider a chain of unit masses connected by springs satisfying Hooke's law with spring constant 1, "force on mass = - displacement from equilibrium". (Figure 27). Let q_j be the displacement of mass j from equilibrium. Derive the equations

$$\ddot{q}_j = q_{j-1} - 2q_j + q_{j+1}.$$

(They are the same as the beaded string equations (2.4.5), for the choice $m_j = 1, \ell_j = 1$, $T = 1$.) Find the energy, the energy norm, and prove uniqueness and stability. ∎

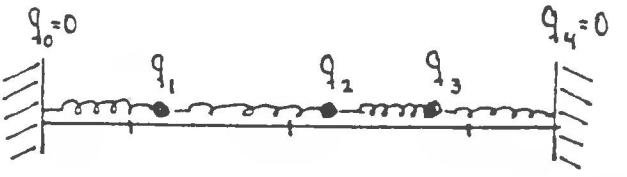


Figure 27. *The harmonic chain*

2.4.37 Exercise*. Solve the beaded string equations

$$M\ddot{q} = -Lq,$$

as follows. Set $Q = \sqrt{M}q$, and find

$$(*) \qquad \ddot{Q} = -\tilde{L}Q,$$

where $\tilde{L}$ is the symmetric matrix

$$\tilde{L} = (\sqrt{M})^{-1}L(\sqrt{M})^{-1}.$$

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\tilde{L}$, and choose eigenvectors $Q^{(1)}, \ldots, Q^{(n)}$. Express the general solution of (*) in the form

$$\sum_{j=1}^{n} a_j(t)Q^{(j)},$$

and use this to find the solution of the initial value problem for the beaded string equation. ∎

This subsection concludes with two supplements. There is a short description of the three fundamental partial differential equations of mathematical physics, and then some key formulas from vector calculus are reviewed and applied to uniqueness and stability problems for those equations.

2.4.38 SUPPLEMENT: The classical partial differential equations

NOTATION AND CONVENTIONS. The symbol x (note: no longer bold face!) will denote an element of $\mathbf{R}^n$, $x = (x_1, \ldots, x_n)$. The value of n will be clear from the context: x may lie in $\mathbf{R}$, or $\mathbf{R}^2$, *etc.* $\triangle$ is the Laplace operator

$$\triangle \stackrel{\text{def}}{=} \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_n^2}$$

that acts on functions of x. ∇ is the gradient operator applied to scalar-valued functions of n variables,

$$\nabla u(x) = (\partial u/\partial x_1, \ldots, \partial u/\partial x_n).$$

Partial derivatives are usually indicated by subscripts:

$$u_x = \frac{\partial u}{\partial x}, \quad u_{tt} = \frac{\partial^2 u}{\partial t^2},$$

and so on. Functions are assumed to be differentiable as many times as necessary, and domains $D \subset \mathbf{R}^n$ are assumed to have smooth boundary. ∎

I will describe —albeit briefly— the three classical partial differential equations of mathematical physics. The unknown functions will be called u, rather than q: I think of q as displacement of a mass, and the letter u seems a little more neutral.

2.4.39 Example. *The Wave Equation.* The wave equation in one space dimension has already been introduced; this example gives the n-dimensional generalization. The unknown function

$$u(t, x), \quad t \geq 0. x \in D \subset \mathbf{R}^n,$$

is to be determined from the equation

$$u_{tt}(t, x) = \triangle u(t, x), \quad t \geq 0, x \in D, \tag{2.4.40}$$

supplemented by the *initial conditions*

$$u(0, x) = f(x), \quad u_t(0, x) = g(x), \quad x \in D, \tag{2.4.41}$$

where f and g are given functions, and the *boundary conditions*

$$u(t, x) = 0 \quad \text{for all } x \in \partial D, \tag{2.4.42}$$

where ∂D denotes the *boundary of* D. Condition (2.4.42) is only one of many possible ones.

The solution $u(t, x)$ is supposed to describe the displacement of a vibrating "membrane" over the n-dimensional domain D.

i) If $n = 1$ and $D = [0, 1]$, then the "membrane" is a string which, at rest, covers the unit interval. The value of the solution u at a point (t_0, x_0) represents the displacement of the string from its equilibrium, at time t_0 and position x_0. The boundary condition (2.4.42) signifies that the ends of the string are tied down at the points $x = 0, 1$. The initial function f describes the displacement of the string at the initial time $t = 0$, and the function g describes the initial velocity imparted to the string. This case was already discussed above.

ii) If $n = 2$, the "membrane" should be thought of as the membrane of a drum of shape D, tied down at the rim ∂D of the drum. Again, f and g describe the initial displacements and velocities of the points of the drum membrane, and $u(t, x)$ is the subsequent displacement of point x as function of time t.

iii) If $n = 3$, all of space "vibrates", so to speak. The value $u(t, x)$ of a solution may be thought of as the strength of a sound wave, or of an electromagnetic wave, at point x and time t. ∎

2.4.43 Example. *The Heat Equation.* The unknown function $u(t, x)$, for $x \in D \subset \mathbf{R}^n$ and $t \geq 0$, is to be determined from the equation

$$u_t(t, x) = \Delta u(t, x). \quad t \geq 0, x \in D. \tag{2.4.44}$$

augmented by the *initial condition*

$$u(0, x) = f(x), \quad x \in D, \tag{2.4.45}$$

where f is a given function, and by the *boundary condition*

$$u(t, x) = 0, \quad \text{for all } x \in \partial D. \tag{2.4.46}$$

The interpretation is as follows: the n-dimensional region D is heated, and the temperature at point x and time $t = 0$ is $f(x)$. The boundary ∂D is kept at temperature 0 (the boundary condition). Then, $u(t, x)$ represents the temperature at point x at time $t > 0$.

Boundary conditions other than (2.4.46) are possible. For example, requiring that the *normal derivative* of u vanish along the boundary amounts to asking that the boundary be insulating, i.e. it allows no heat to leave the domain.

The heat equation is derived from a physical law called *Fourier's law of heat conduction*. Its applicability to real life is described in various physics texts. ∎

2.4.47 Example. *The Laplace Equation.* The function $u(x), x \in D \subset \mathbf{R}^n$, is to be determined from the equation

$$\Delta u(x) = 0, \quad \text{for } x \in D, \tag{2.4.48}$$

augmented by the *boundary condition*

$$u(x) = \varphi(x), \quad \text{for } x \in D, \tag{2.4.49}$$

where φ is a given function. This problem (equation plus boundary condition) has many, many applications. We mention only one.

Change the boundary condition in the heat equation, (2.4.46), to $u(t, x) = \varphi(x)$. Assume, for sake of physical realism, that $\varphi \geq 0$. This means: some experimental setup keeps the temperature at point $x \in \partial D$ at the non-negative value $\varphi(x)$. A solution of Laplace's Equation (2.4.48) with its boundary condition (2.4.49) may be interpreted as a *time-independent* solution of this heat equation. In other words, the initial temperature distribution f from (2.4.45) has had a chance to settle down and reach equilibrium, and the resulting *steady - state* temperature distribution u is the one permitted by the externally imposed boundary temperature φ. ∎

The three basic questions raised at the beginning of this subsection are again the starting point for the theory of these partial differential equations:

i) Is there a solution u?
ii) If there is, is the solution unique?

iii) Does it change by just a little bit if the external conditions, initial and/or boundary, are changed by a small amount?

The many physical applications of the equations make a fourth question equally important:

iv) Can one find explicit formulas for the solutions? Can one find approximate values for the solutions?

It is fair to say that much of mathematical analysis during the past two centuries was developed in order to understand those three partial differential equations arising in physics.

2.4.50 SUPPLEMENT: Green's formulas and energy estimates

The crucial step in the proof of conservation of energy in the one-dimensional wave equation was integration by parts (see (2.4.25)):

$$\int_0^1 a_x b_x \, dx = a_x b \Big|_0^1 - \int_0^1 a_{xx} b \, dx;$$

in that example, we had $a = q$ and $b = q_t$. In higher dimensions, one also uses integration by parts, which there goes under the name *Green's formula*, see EDM II, Appendix A, Table 3:

$$\int_D \nabla a \cdot \nabla b \, dV = \int_{\partial D} b \frac{\partial a}{\partial n} \, dS - \int_D b \triangle a \, dV. \tag{2.4.51}$$

Here dV is the volume element in D, dS is the element of surface area in ∂D, and $\mathbf{n}(x)$ is the unit outward normal at $x \in \partial D$. The directional derivative in the normal direction is written

$$\frac{\partial a}{\partial n} = \nabla a \cdot \mathbf{n}.$$

Multiple integrals are, for ease of notation, represented by a single $\int$ sign, but the dimension is arbitrary. If $D \subset \mathbf{R}^2$, then dV is the element of area $dx\,dy$, if $D \subset \mathbf{R}^3$, then $dV = dx\,dy\,dz$, and so forth.

If one writes (2.4.51) a second time with a and b interchanged and subtracts one of the two formulas from the other, one finds another useful identity, also called *Green's formula*:

$$\int_D a \triangle b - b \triangle a \, dV = \int_{\partial D} a \frac{\partial b}{\partial n} - b \frac{\partial a}{\partial n} \, dS. \tag{2.4.52}$$ ∎

• **2.4.53 Exercise.** Derive (2.4.52) from (2.4.51). ∎

Green's formula (2.4.51) is now used to deduce a stability/uniqueness result for the multi-dimensional heat equation.

2.4.54 Example. Consider the (2.4.43), taking D to be a *bounded* domain. The energy in a solution u is defined to be

$$E(u)(t) \stackrel{\text{def}}{=} \frac{1}{2}\int_D u(t,x)^2\,dV. \tag{2.4.55}$$

Now compute the time-derivative of E:

$$\begin{aligned}\frac{dE(u)(t)}{dt} &= \int_D u(t,x)u_t(t,x)\,dV \\ &= \int_D u(t,x)\triangle u(t,x)\,dV \quad \text{(use the equation)} \\ &= \int_{\partial D} u(t,x)\frac{\partial u(t,x)}{\partial n}\,dS - \int_D \|\nabla u(t,x)\|_2^2\,dV \quad \text{(Green's formula (2.4.51))} \\ &= -\int_D \|\nabla u(t,x)\|_2^2\,dV \quad \text{(use boundary condition).}\end{aligned}$$

The last expression is ≤ 0 because the integrand is non-negative, and we conclude:

$$\frac{dE(u)(t)}{dt} \leq 0. \tag{2.4.56}$$

The energy decreases or, in more physical terms, it dissipates into the outside world which is not described by the heat equation.

Inequality (2.4.56) leads to a stability result about the distance between two solutions. Let u and v solve the Heat Equation with initial conditions f and $\tilde{f}$, and with zero boundary conditions. Then the difference $w = u - v$ solves the heat equation with initial condition $f - \tilde{f}$. Since energy is non-increasing, the energy in w at time $t > 0$ is $\leq$ the energy at time $t = 0$:

$$\int_D (u(t,x) - v(t,x))^2\,dV \leq \int_D (f(x) - \tilde{f}(x))^2\,dV.$$

In particular, if $f \equiv \tilde{f}$, then $u \equiv v$: the solution to the initial/boundary value problem is unique. ∎

2.4.57 Example. Next, consider the Laplace equation (2.4.47). The arguments are closely related to those just presented for the heat equation. This is to be expected, since Laplace's equation is the steady-state version of the heat equation, see (2.4.47).

Let $D \subset \mathbf{R}^n$ be a bounded region. Let M be the set of functions on D for which the second partial derivatives are continuous and which are zero on the boundary ∂D of D. This set becomes a normed linear space if we define

$$\|u\| \stackrel{\text{def}}{=} \sqrt{\int_D \|\nabla u(x)\|_2^2\,dV}. \tag{2.4.58}$$

Let us outline the proof of this assertion.

It is clear that $\|u\| = 0$ if $u = 0$. Conversely, suppose that $\|u\| = 0$. Since the integrand in (2.4.58) is non-negative and the integral is zero, the integrand must in fact be zero. It follows that $\|\nabla u(x)\|_2^2 \equiv 0$, which in turn implies that $\nabla u(x) = 0$. If all partial derivatives of u are zero, then u is constant. But on ∂D, $u = 0$, so the constant must be zero, i.e., $u \equiv 0$. This proves (2.2.2).

The scaling condition (2.2.3) is trivially satisfied.

Minkowski's inequality (2.2.4) is a consequence of Exercise 1.4.16.

Now suppose there are two solutions u, v of the boundary value problem (2.4.48,49), both assuming the boundary values φ on ∂D. The difference $w = u - v$ again solves $\triangle w = 0$ in D, and $w = 0$ on ∂D. Using Green's formula (2.4.51), compute:

$$\int_D \|\nabla w(x)\|_2^2 \, dV = \int_{\partial D} w \cdot \frac{\partial w}{\partial n} \, dS - \int_D w(x) \triangle w(x) \, dV. \tag{2.4.59}$$

The first term on the right is zero because w vanishes on the boundary. The second term is zero because $\triangle w$ is zero. Hence, the integral on the left is zero, and as we just saw, this implies that $w \equiv 0$. The solution of the boundary value problem (2.4.48, 49) is therefore unique, at least in the class of functions with continuous second partials. ∎

Green's formulas (2.4.51) and (2.4.52) are ideal for *isotropic* problems. "Isotropic" means that all directions in space are equivalent. In the three classical equations, this feature is encoded in the Laplace operator $\triangle$: it is invariant under rotations and translations. Consider the wave equation in all of $\mathbf{R}^3$, for example:

$$u_{tt} = \triangle u.$$

If you find a solution $u(t, x_1, x_2, x_3)$, then I can build another solution. Define $\mathbf{y} = \mathbf{a} + \mathbf{Q}\mathbf{x}$, with an orthogonal matrix $\mathbf{Q}$; then $u(t, y_1, y_2, y_3)$ will again solve the wave equation. I have translated and rotated your solution. I could not do this in the presence of *anisotropy*. "Anisotropic" means that spatial direction influences the physical process under investigation. Think about a circular drum (the head of a banjo). If you stretch the membrane unevenly, the waves will propagate unevenly, even though the drum is perfectly symmetric. The sound will not be quite right. You could also produce anisotropy by messing with boundary conditions: let heat flow in a uniformly conducting circular plate, but keep part of the boundary at zero temperature while heating the rest to a fixed nonzero value.

There are, evidently, at least two ways to destroy isotropy: change the medium (tighten the drum unevenly) or change the boundary conditions (heat the boundary unevenly). Boundary-induced anisotropy is built into all the IBVP's of Supplement 2.4.38. I want to consider the wave-, heat, and Laplace-equation in an *anisotropic medium*. This requires a change in the $\triangle$ operator.

If the anisotropy is constant, depending on direction but not on position, then $\triangle u$ is replaced by

$$\sum_{i,j=1}^{n} a_{ij} u_{x_i x_j}. \tag{2.4.60}$$

Here $\mathbf{A} = (a_{ij})$ is a constant matrix; physical laws usually require that $\mathbf{A}$ be symmetric and positive definite. If the properties of the medium depend on position as well as on direction, then Δu is replaced by

$$\sum_{i,j=1}^{n} \left(a_{ij}(x)u_{x_i}\right)_{x_j}. \tag{2.4.61}$$

Again, $(a_{ij}(x))$ is, for each x, a positive definite symmetric matrix; assume also that the entries a_{ij} are continuously differentiable in x.

Energy estimates for the anisotropic wave, heat, or Laplace equation require new rules from multivariable calculus. The basic one, from which all others follow, is one of *Gauss's formulas*; it is a direct generalization of the Fundamental Theorem of Calculus,

$$\int_a^b f'(x)\,dx = f(b) - f(a).$$

2.4.62 Proposition. *Let $D \subset \mathbf{R}^n$ be a bounded open set with smooth boundary, and let u be continuously differentiable on $\overline{D}$. For $x \in \partial D$, let $\mathbf{n}(x) = (n_1(x), \ldots)$ be the unit outward normal. Then*

$$\int_D \frac{\partial u(x)}{\partial x_j}\,dV = \int_{\partial D} u(x)n_j(x)\,dS. \tag{2.4.63}$$ ∎

One can write (2.4.63) in a perhaps more familiar vector form:

$$\int_D \nabla u(x)\,dV = \int_{\partial D} u(x)\mathbf{n}(x)\,dS. \tag{2.4.64}$$

Just as in one-variable calculus, if you replace the function u in the "Fundamental Theorem" by a product vu, you get the rule for integration by parts.

2.4.65 Integration by parts. *Let n_i be the i^{th} component of the unit outward normal to ∂D. Then*

$$\int_D \frac{\partial u}{\partial x_i}\,v\,dV = \int_{\partial D} uv\,n_i\,dS - \int_D \frac{\partial v}{\partial x_i}u\,dV. \tag{2.4.66}$$ ∎

In vector notation:

$$\int_D v\nabla u\,dV = \int_{\partial D} uv\,\mathbf{n}\,dS - \int_D u\nabla v\,dV. \tag{2.4.67}$$

Example. Integration by parts is now applied to establish conservation of energy for the *anisotropic wave equation*,

$$u_{tt}(x,t) = \sum_{i,j=1}^{n} a_{ij}u_{x_ix_j}. \tag{2.4.68}$$

Again, equation (2.4.68) is to be solved in a bounded domain $D \subset \mathbf{R}^n$, subject to the boundary condition $u(x,t) = 0$ when $x \in \partial D$. The initial conditions $u(x,0) = f(x)$, $u_t(x,0) = g(x)$ are also prescribed. Once it is proved that energy is conserved, stability and uniqueness statements follow as in the other examples.

The first task is, obviously, to get a formula for the energy. This can be done in two ways. 1) Know some physics, think of a problem described by (2.4.68), and write the energy for that problem. 2) Know some mathematics, diddle around with analogs of (2.4.24), and stop when you find a norm-like expression whose t-derivative is zero by virtue of equation (2.4.68). Call that expression "energy", and sneer condescendingly at anyone who asks for a physical interpretation. I follow the second path. (But see Exercise 2.4.78 for the professional secret.)

2.4.69 Proposition. *The "energy"*

$$E(t) = \frac{1}{2}\int_D u_t(t,x)^2 + (\mathbf{A}\nabla u(t,x)) \cdot \nabla u(t,x)\, dV$$

is constant along solutions of (2.4.68). The formula

$$\|(f,g)\|_E \stackrel{\text{def}}{=} \sqrt{\frac{1}{2}\int_D g(x)^2 + (\mathbf{A}\nabla f(x)) \cdot \nabla f(x)\, dV}$$

defines a norm on the space of pairs (f,g) of continuously differentiable functions on the domain D, where f is required to be zero on the boundary ∂D. ∎

Proof. Let $E(t)$ be the energy at time t. Then

$$\frac{dE(t)}{dt} = \int_D u_t u_{tt} + \frac{1}{2}(\mathbf{A}\nabla u_t) \cdot \nabla u + \frac{1}{2}(\mathbf{A}\nabla u) \cdot \nabla u_t\, dV.$$

Because $\mathbf{A}$ is symmetric,

$$(\mathbf{A}\nabla u_t) \cdot \nabla u = \nabla u_t \cdot (\mathbf{A}\nabla u),$$

and so

$$\frac{dE(t)}{dt} = \int u_t u_{tt} + (\mathbf{A}\nabla u) \cdot \nabla u_t\, dV. \tag{2.4.70}$$

Write out the integral of the second term:

$$\int_D (\mathbf{A}\nabla u) \cdot \nabla u_t\, dV = \sum_{i,j=1}^n a_{ij} \int_D \frac{\partial u}{\partial x_j}\frac{\partial u_t}{\partial x_i}\, dV.$$

Integrating by parts according to (2.4.65), we convert this into

$$\sum_{i,j=1}^n a_{ij}\left(\int_{\partial D} \frac{\partial u}{\partial x_j} u_t n_i\, dS - \int_D \frac{\partial^2 u}{\partial x_i \partial x_j} u_t\, dV\right). \tag{2.4.71}$$

The boundary integral in (2.4.71) vanishes since $u(x,t) = 0$ when $x \in \partial D$, and (2.4.70) becomes

$$\int_D u_t \left(u_{tt} - \sum_{i,j=1}^{n} a_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} \, dV \right),$$

which vanishes thanks to the wave equation (2.4.68).

The verification of the norm axioms is left to the reader. ∎

• **2.4.72 Exercise.** Consider the wave equation (2.4.39). Define the energy of a solution at time t, and show that $dE/dt = 0$. Use this result to prove a uniqueness theorem. (This is a special case of Proposition 2.4.69; you should do it without looking at the text.) ∎

2.4.73 Exercise. The so-called *Neumann boundary value problem* for $\Delta u = 0$ prescribes the value of the normal derivative at the boundary, say $\partial u/\partial n = \psi$ on ∂D. The solution u may be thought of as the steady-state solution of the heat equation, when the heat flux out of the region D is prescribed. Use the argument in Example 2.4.57 to try to prove uniqueness of the solution. Something goes just a little wrong — explain. ∎

2.4.74 Exercise. Prove that the "energy norm" for the anisotropic wave equation (2.4.68) is really a norm. ∎

• **2.4.75 Exercise.** Derive Green's formula (2.4.51) from the integration by parts formula (2.4.65). ∎

2.4.76 Exercise. Prove uniqueness and stability for an anisotropic heat equation (with constant (a_{ij})). ∎

2.4.77 Exercise. Find the "energy", and prove uniqueness and stability, for the inhomogeneous anisotropic wave equation in which the Laplacean is replaced by (2.4.61). ∎

2.4.78 Exercise. The standard (mathematician's) trick to find the energy that goes with a system of differential equations is as follows.

For $\ddot{q} = -Lq$, with symmetric positive definite L, take the inner product with $2\dot{q}$,

$$2(\ddot{q}, \dot{q}) = -2(Lq, \dot{q}),$$

and recognize this as saying that

$$\frac{d}{dt}(\dot{q}, \dot{q}) = -\frac{d}{dt}(Lq, q).$$

Hence

$$\frac{d}{dt}\Big((\dot{q}, \dot{q}) + (Lq, q)\Big) = 0,$$

and the time-independent, non-negative, quantity in parentheses is then called "energy".

The same method works for the wave equation. Multiply $u_{tt} = \Delta u$ by $2u_t$ and integrate $\int_D$: that is the analog of the inner product.

$$\int_D 2u_t u_{tt} \, dV = \int_D u_t \Delta u \, dV.$$

By Green's formula and the boundary condition, one converts the right side to

$$-2 \int_D \nabla u \cdot \nabla u_t \, dV,$$

which one recognizes as

$$-\frac{d}{dt}\int_D \|\nabla u\|_2^2\, dV.$$

The expression for "energy" is then guessed as for the ordinary differential equation.

Carry out this procedure to derive the energy (2.4.69) for the anisotropic wave equation. ∎

2.4.79 Exercise*. Modify the procedure in the preceding exercise to find a norm-like quantity that decreases along the solutions of the equation

$$\ddot{q} = -Lq - C\dot{q}.$$

The matrix C is positive definite symmetric. This equation might describe a beaded string with friction. ∎

2.4.80 Exercise*. Adapt the procedure in the preceding exercise to guess the "energy" for the anisotropic heat equation. ∎

Section 3 Topology

Metrics and norms allow one to talk about "closeness" of elements of a set. One can then formulate statements involving limits, which permit the definition of "continuous function," which in turn lead to derivatives, and pretty soon one is slinging around Taylor series and multiple integrals .... But hold the excitement: we are not yet finished with "closeness." There is one more abstraction that must be discussed. It is the theory called *Topology*, and it reduces the concept of "closeness" to its most primitive form. You might wonder why the Ultimate Abstraction should be studied at the beginning of an introduction to analysis, rather than at the end. The reason is simple. Even though the major *theorems* of General Topology tend to play a role only in advanced analysis, the basic *language* and *ideas* are pervasive. It is difficult to read papers and books on analysis at any level without encountering terms like *open set, closed set, Hausdorff space, weak topology, etc.*. This section will attempt to explain what the basic concepts are, how one should think about them, and why they are useful.

To illustrate the guiding idea, we begin with a concept that is no doubt familiar (and will, in any case, be dealt with at length later on): continuity of a function. Here is the intuitive definition:

> A function f from a metric space M to a metric space N, is continuous at $x \in M$, if $f(y)$ can be made arbitrarily close to $f(x)$ by taking y sufficiently close to x.

The ϵ's and δ's will be filled in later. Notice, however, that even without them the definition is really quite meaningful. In fact, students who are struggling with ϵ's and δ's in Calculus usually get something like it as motivation, and understand it — becoming confused only when "close" is quantified.

Figure 28 illustrates this definition:

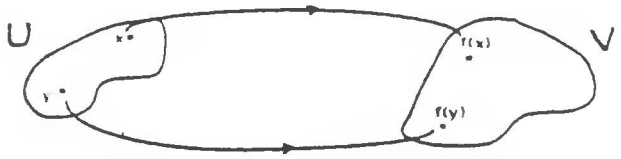
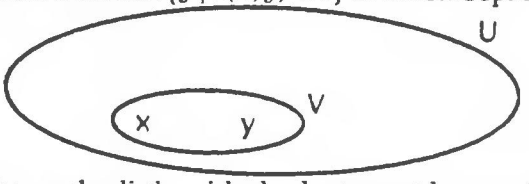


Figure 28. *Continuity*

The blob V surrounding $f(x)$ indicates how close we want $f(y)$ to be to $f(x)$. The blob U surrounding x suggests how close y must be to x if we are to achieve our goal. Now, instead of using a metric to make sure that the distance between $f(x)$ and $f(y)$ is less than ϵ, topology *uses only the shapeless blobs*. Of course, there are *some* axioms, but the important point is that distance is not mentioned. There is only a primitive concept of "closeness".

The definition to be given in a moment does the following. A collection, $\mathcal{T}$, of subsets of a set X is singled out: if x, y belong to one of those subsets, say U, then x and y are "close". If there is another one of those subsets, say V, contained in U and containing x, y ($x, y \in V \subset U$), then x and y are closer than you first thought, so to say. These distinguished subsets are called *open sets*. They are to be thought of as "completely surrounding" every point they contain. If U is an open set, and $x \in U$, then you cannot reach the point x without passing through U. In this sense, open sets generalize the sets $\{y \mid d(x,y) < \epsilon\}$ in a metric space.



For a decent theory, the distinguished subsets must have certain properties:

3.0.1 Definition. *A topological space is a set X together with a collection $\mathcal{T}$ of subsets of X satisfying*

(3.0.2) $$\emptyset \in \mathcal{T},$$

(3.0.3) $$X \in \mathcal{T}$$

(3.0.4) *If $U, V \in \mathcal{T}$, then $U \cap V \in \mathcal{T}$.*

(3.0.5) *If $\mathcal{U} \subset \mathcal{T}$ is any collection of sets in $\mathcal{T}$, then the union $\bigcup_{U \in \mathcal{U}} U$ is in $\mathcal{T}$.*

A collection $\mathcal{T}$ of sets obeying (3.02)-(3.05) is called a **topology** *on X. The sets belonging to $\mathcal{T}$ are called* **open sets**. ∎

The difference between (3.0.4) and (3.0.5) is crucial. By induction[16], one can show from (3.0.4) that the intersection of a finite number of sets from $\mathcal{T}$ again belongs to $\mathcal{T}$. Examples show that infinite intersections may not belong to $\mathcal{T}$. On the other hand, (3.0.5) says that arbitrary unions – countable, or even uncountable – of sets in $\mathcal{T}$ again belong to $\mathcal{T}$. Examples will be given soon.

Terminology. The collection $\mathcal{T}$ in the definition is called *a topology on X*. The pair $(X, \mathcal{T})$ is a *topological space*. If $\mathcal{T}$ is given or understood to be given, one often calls X a topological space. The study of topological spaces is, confusingly, also called *topology*. Abstract topological spaces (as in the definition above) were introduced as generalization of metric spaces, around 1915-25. There are several equivalent definitions, of which (3.0.1) is the most common. ∎

Remark. The four axioms in Definition 3.0.1 are quite sparse — they don't impose many conditions, and it is therefore possible to have many topologies, some really bizarre. It is natural to supplement the definition with further axioms, thereby creating a topological space with special properties. The table at the end of Article 425 in EDM II gives an idea of the number of supplementary properties that have been studied during the last 70 or so years. ∎

This section explores the implications of Definition (3.0.1). In Subsection 1, we study the now familiar metric spaces from the point of view of topology. Subsection 2 re-examines the concept of continuity, in the setting of general topological spaces. Subsection 3 introduces a particular type of topological space, the *Hausdorff space*. (There are also some concrete and important examples of non - Hausdorff spaces.) A topology is hardly ever defined by listing every set in the collection $\mathcal{T}$, and Subsection 4 describes useful methods for constructing topologies $\mathcal{T}$ in an implicit manner. Subsection 5 presents a miscellany of standard notions, many of which may be familiar in more concrete settings: *closed sets*, *accumulation points*, and so on. Some nontrivial applications then follow. In Subsection 6, the *space of sample paths of Brownian motion* is constructed, and in Subsection 7, *weak topologies* on normed linear spaces are introduced.

For a change, I should also mention topics that are *not* covered in this section. Certain topological questions can be stated in everyday languange and illustrated by pretty pictures; therefore, Topology tends to appear in non-technical books and journals, such as the *Time-Life* series or *Scientific American*. You may have heard that a topologist can't tell a doughnut from a coffee cup; that topology is "rubber-sheet geometry;" that topologists study knots and apply their insights to

---

[16] The general scheme of mathematical induction is the following. One wants to prove a statement P_n for all integers n. 1) Prove P_0; 2) assume that P_k is true; iii) prove that P_{k+1} follows from P_k. Equivalently, one may assume 2'): the statements P_n are true for $1 \leq n \leq k$, and deduce P_{k+1} from that assumption.

DNA research; that Euler discovered $V-E+F=2$ while walking over the bridges of Königsberg, and the like. These things are true to varying degrees. There are aspects of topology that have a geometric flavor, aspects with a combinatorial orientation, and areas with close ties to abstract algebra. The topology needed in this course is either very basic, too rudimentary to solve (or even allow one to state) the glamor problems of the subject, but primarily it has a different orientation: for the foundations of analysis, the unglamorous branch called *point set topology* is the most immediately applicable, and it will be our main concern.

- **3.0.6 Exercise.** Let $(X,\mathcal{T})$ be a topological space. If $U_1,U_2,U_3\in\mathcal{T}$, then $U_1\cap U_2\in\mathcal{T}$ by (3.0.4), and hence $(U_1\cap U_2)\cap U_3\in\mathcal{T}$ by another application of (3.0.4). Prove by induction that if $U_1,\ldots,U_n\in\mathcal{T}$, then $U_1\cap\ldots\cap U_n\in\mathcal{T}$. ∎

- **3.0.7 Exercise.** Let $X=\{-1,0,1\}$, and let $\mathcal{T}=\big\{X,\emptyset,\{1\},\{-1\},\{-1,1\}\big\}$. Show that the axioms for a topological space are satisfied. This may seem like a silly example, but it actually comes from the differential equation $y'=y$; see (3.3.4). ∎

- **3.0.8 Exercise.** On every set X, there are two trivial topologies: $\mathcal{T}_1=\{\emptyset,X\}$ and $\mathcal{T}_2=$ the collection of *all* subsets of X. For each of these two topologies, find the smallest open set containing a given point $x\in X$. ∎

- **3.0.9 Exercise.** Let $(X,\mathcal{T})$ be a topological space, and let $A\subset X$. Suppose that for each $x\in A$ there is a $U_x\in\mathcal{T}$ such that $x\in U_x\subset A$. Show that then $A\in\mathcal{T}$. (Hint: consider the union of all the U_x.) ∎

Subsection 3.1 Metric Space Topology

A metric space is, in a natural way, a topological space. The *metric topology* will be defined in this subsection. We will, in effect, throw away the numerical measure of distance provided by the metric, and keep only information about the relative "closeness" of pairs of points. As is explained in the next subsection, that remnant of the metric still allows one to talk about convergence of sequences and continuity of functions, just as in Calculus.

Notation. *In a metric space (M,d), the* ϵ*-ball about* x *is the set*

$$B(x,\epsilon)=\{y\,|\,d(x,y)<\epsilon\}.\ ∎$$

A point y will be "near" x if it belongs to some ball $B(x,\epsilon)$. A point z is, morally speaking, closer to x than y is if it belongs to a ball $B(x,\eta)$ contained in $B(x,\epsilon)$. A sequence $\{x_n\}$ will "approach" the point x if the x_n belong to nested balls $B(x,\epsilon_n)$ that shrink down to x. The diameters of the balls will be irrelevant; one need only know that one ball is contained in another, and that any given ball always contains a smaller one. At the end of this subsection, the essence of the relationships amongst balls $B(x,\epsilon)$ is distilled into the concept of *local basis*. The notion of "metric" is by then discarded for good.

3.1.1 Proposition – Definition. *Let (M, d) be a metric space. Let $\mathcal{T}$ be the collection of subsets of M consisting of the empty set $\emptyset$ and all the subsets of M that satisfy the following condition:*

for every $x \in U$ there is an ϵ (depending on x) such that $B(x, \epsilon) \subset U$.

Then $\mathcal{T}$ is a topology on M, called the ***metric topology.*** ∎

Figure 29 illustrates the property to be satisfied by a subset $U \subset M$ belonging to $\mathcal{T}$. Every point (x and y, in the figure) is surrounded by a ball, possibly very small, that lies entirely in U.

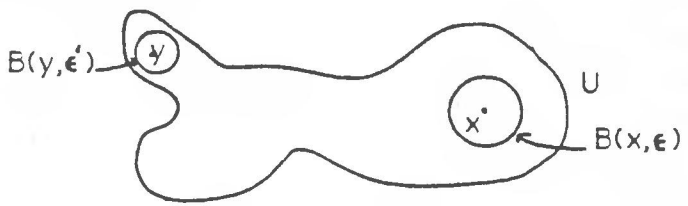


Figure 29. *How to be open*

Such sets are called *open* (Definition 3.0.1). Their characteristic property is this: *if an open set contains a point x, it also contains all points sufficiently close to x.*

Proof. (3.0.2): By definition, $\emptyset$ belongs to $\mathcal{T}$.

(3.0.3): It is clear that $M \in \mathcal{T}$: if $x \in M$, then $x \in B(x, 1) \subset M$.

(3.0.4): Suppose next that $U \in \mathcal{T}$ and $V \in \mathcal{T}$. We need to show that $U \cap V \in \mathcal{T}$. Let $x \in U \cap V$. We are to find a $B(x, \epsilon) \subset U \cap V$. By assumption, since $x \in U$ and $x \in V$ and both U, V are in $\mathcal{T}$, we have $B(x, \eta) \subset U$ and $B(x, \zeta) \subset V$ for some $\eta, \zeta > 0$. Suppose $\eta \leq \zeta$. Then $B(x, \eta) \subset B(x, \zeta)$, and $B(x, \eta) \subset U \cap V$. If $\eta \geq \zeta$, the argument is similar. (Draw a picture!)

(3.0.5): Finally, let $\mathcal{U}$ be a collection of sets in $\mathcal{T}$. If

$$x \in \bigcup_{U \in \mathcal{U}} U \stackrel{\text{def}}{=} V,$$

then x belongs to one of the sets U. Because U is open, $B(x, \epsilon) \subset U$ for some $\epsilon > 0$. But since $U \subset V$, we have $B(x, \epsilon) \subset U \subset V$, and so the union V belongs to $\mathcal{T}$. ∎

3.1.2 Corollary. *Every ϵ-ball is an open set (i.e., belongs to the metric topology $\mathcal{T}$) as defined above.* ∎

Proof. Let $y \in B(x, \epsilon)$. If $y = x$, then $B(y, \epsilon) = B(x, \epsilon)$, and the requirement of the definition is satisfied. If $y \neq x$, then $d(x, y) \stackrel{\text{def}}{=} a > 0$, and $a = d(x, y) < \epsilon$. Pick

an $\eta > 0$ such that $a + \eta < \epsilon$. As Figure 30 suggests, $B(y, \eta) \subset B(x, \epsilon)$; we now prove this. Pick $z \in B(y, \eta)$. By the triangle inequality,

$$d(x; z) \leq d(x, y) + d(y, z) < a + \eta < \epsilon.$$

Hence $z \in B(x, \epsilon)$, which shows that the picture is correct: $B(x, \epsilon)$ is open according to (3.1.1). ∎

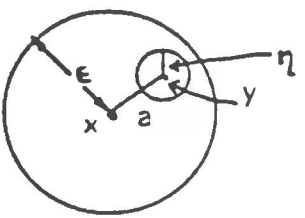


Figure 30. *Balls are open*

Next, we illustrate Proposition 3.1.1 with several simple examples.

3.1.3 Example. $M = \mathbf{R}$. An interval (a, b) is usually called *open*. It is also "open" in the sense of Proposition 3.1.1, because (a, b) is actually a ball $B(x, \epsilon)$: take $x = (b + a)/2$ and $\epsilon = (b - a)/2$. The intervals $[a, b)$, $(a, b]$, $[a, b]$ are not open, because at least one of the endpoints is not contained in any $B(a, \epsilon)$ or $B(b, \epsilon)$ belonging entirely to the interval.

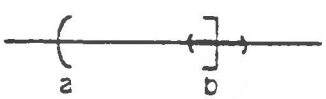


Figure 31. $(a, b]$ *is not open*

We check the axioms for a topology on open intervals.

i) the empty set $\emptyset$ is open — this is merely a convention;
ii) the whole space, R, is open — indeed, for every x there is an ϵ (e.g., $\epsilon = 1$) so that $x \in B(x, \epsilon) \subset \mathbf{R}$;
iii) if $U = (a, b)$ and $V = (c, d)$ are open, so is $U \cap V$. This is evident if $U = (a, b)$ and $V = (c, d)$. $U \cap V$ may be $\emptyset$, which is open by definition, or $U \cap V = U$ if $U \subset V$ or $U \cap V = V$ if $V \subset U$. Or, $U \cap V$ is another open interval:

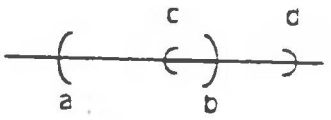


Figure 32. ***An intersection of open intervals***

iv) *every* union of open sets is open. For example, let $Y = \bigcup I$ = a union of open intervals: if $x \in Y$, then x belongs to one of the intervals whose union forms Y. Call that interval J. Because J is open, there is a $B(x, \epsilon)$ such that $x \in B(x, \epsilon) \subset J \subset Y$; so Y also satisfies the criterion for "open set."

3.1.4 Proposition. *Every (open) set, other than $\emptyset$, in the metric topology of $\mathbf{R}$ is a countable union of disjoint intervals.* ∎

Proof. Let U be an open set (in the metric topology). When $x \in U$, there is an interval (α, β) such that $x \in (\alpha, \beta) \subset U$ (since U is open). It will be important that there is in fact a *largest* open interval containing x and contained in U. This probably seems obvious, but is in fact a consequence of fundamental properties of the real numbers incorporated in their axiomatic construction. There is no time to go into those details, and we must simply accept the assertion [17].

The proof of the proposition now goes as follows. The rational numbers are countable; list them as $r_1, r_2, r_3, \ldots$. Go down the list: if $r_j \notin U$, discard it; if $r_j \in U$, let (a_j, b_j) be the largest interval containing r_j and contained in U. There will be many duplications in the list of the (a_j, b_j), but at most countably many distinct ones. Any two of those must be disjoint—if they were not, their union would be an interval longer than either of them, contrary to construction.

Now let V denote the union of these disjoint longest intervals. I claim that $U = V$. Indeed, every interval used to form V is contained in U by design, so $V \subset U$. Conversely, if $x \in U$, then $x \in (\alpha, \beta) \subset U$ for some (α, β); there is a rational number r_j in (α, β), and then the longest interval (a_j, b_j) containing r_j must be longer than (α, β), i.e. $x \in (\alpha, \beta) \subset (a_j, b_j) \subset V$. Hence $U \subset V$ and therefore $U = V$. ∎

---

[17] Here is the point. There is at least the one interval (α, β) just described. Consider the set $\{\gamma \mid x \in (\alpha, \gamma) \subset U\}$. If it is unbounded, then $x \in (\alpha, \infty) \subset U$. If it is bounded, then one of the axioms governing the real numbers says that there is a *least upper bound*, $\bar{\gamma}$. One deduces that $x \in (\alpha, \bar{\gamma}) \subset U$, but that $(\alpha, \gamma) \not\subset U$ for any $\gamma > \bar{\gamma}$. Similarly, by considering all δ such that $x \in (\delta, \bar{\gamma}) \subset U$, one arrives at the largest interval $(\bar{\delta}, \bar{\gamma})$.

The fact that every open $U \subset \mathbf{R}$ has the form

$$U = \bigcup_{n=1}^{\infty} (a_n, b_n)$$

may make matters seem simpler than they really are. The resulting sets can actually be very complicated, see Example 3.1.6. ∎

3.1.5 Remark. A *countably infinite intersection* of open intervals in R need not be open. For instance,

$$\bigcap_{n=1}^{\infty} (-\frac{1}{n}, \frac{1}{n}) = \{0\},$$

and $\{0\}$ is not an open set. No $B(0, \epsilon)$, with $\epsilon > 0$, will fit into that one-point set. On the other hand, a countably infinite intersection of intervals *may* be open. Let $I_n = (n, \infty)$. Then $\bigcap_{n=1}^{\infty} I_n = \emptyset$, which is declared to be an open set. ∎

3.1.6 Example. Here is a really complicated open subset of R. Start with the interval $[0, 1]$, and let I_1 be the open middle third, $(1/3, 2/3)$. From each of the two leftover pieces, take the open middle thirds; they are $I_2 = (1/9, 2/9)$ and $I_3 = (7/9, 8/9)$.

Figure 33. *The Cantor set*

Define $I_4, I_5, \ldots$ to be the open middle thirds of the leftover intervals at the preceding steps, see Figure 33. The union $U = \bigcup I_n$ is an open set; it is a fairly complicated sort of set. One can appreciate this best by looking at the set $C = [0, 1] - U$, the complement of U. This is the *Cantor set*, which has many interesting properties.

Notice that whenever an open middle third is removed and put into the union U, its two endpoints are left behind. Altogether, then, C consists of two endpoint for each of the countably many intervals I_n, and is therefore countable. Right? Wrong! C is an *uncountable* set. The intervals I_n get smaller and smaller, and the leftover endpoints can form sequences that tend to other points in $[0, 1] - U$, points that are not endpoints of any interval I_n. In that way, uncountably many points

make their way into C. It may seem plausible, then, that the arrangements of open intervals forming U is complex and subtle. The next two exercises describe some details, and again illustrate the many connections between different mathematical ideas. ∎

3.1.7 Exercise. This exercise gives an explicit description of the Cantor set. Every $x \in [0,1]$ has an expansion in base 3:

$$x = .\alpha_1\alpha_2\alpha_3\ldots = \sum_{n=1}^{\infty} \frac{\alpha_n}{3^n},$$

where each $\alpha_n = 0, 1$, or 2. Show that x has a unique expansion unless it is of the form $x = k/3^n$ for some integers k, n. In that case, there are two expansions. (This is analogous to .0999... = .1 in base 10.) If you cannot prove these statements, accept them and go on.

Show that the Cantor set consists of those $x \in [0,1]$ that have no 1 in their base 3 expansion, together with the x that have two base 3 expansions, one of them without 1's. (Hint: first understand the base 3 expansions of the points that lie in the middle thirds I_1, I_2, I_3.)

The Cantor set is therefore in one-to-one correspondence with the set of all sequences of 0's and 2's. An argument called the *Cantor diagonal process* shows that it is uncountable. See the next exercise. The Cantor set is also related to the metric space in (1.1.45). ∎

3.1.8 Exercise. For sake of completeness, we outline the Cantor diagonal process. Let X be the set of all sequences $(\epsilon_1, \epsilon_2, \ldots)$ of 0's and 1's. The set X is uncountable. To prove this, assume that it is countable, and arrive at a contradiction.

If it were countable, the sequences could be indexed by the positive integers (that is what "countable" means): there is a first sequence, a second sequence, and so on, and the list contains all members of X. Imagine the complete list:

$$\begin{array}{l} \alpha_1, \alpha_2, \alpha_3, \ldots \\ \beta_1, \beta_2, \beta_3, \ldots \\ \gamma_1, \gamma_2, \gamma_3, \ldots \\ \vdots \quad \vdots \quad \vdots \ldots . \end{array}$$

I now exhibit a sequence that is not contained in the list. It is constructed as follows. The first entry is different from α_1, the second entry is different from β_2, the third entry is different from γ_3, and so on. The resulting sequence differs in at least one place from every sequence in the purportedly complete list, and so the list could not have been complete. Contradiction: there cannot be a countable list.

Use the Cantor diagonal process to prove that i) the Cantor set is uncountable; ii) that the real numbers are uncountable. ∎

3.1.9 Example. $M = \mathrm{R}^2$ with the Euclidean metric. A set $U \subset \mathrm{R}^2$ is open (in the metric topology) if every $x \in U$ is contained in a ball, $B(x, \epsilon) = \{y \mid ||y - x||_2 < \epsilon\}$, which in turn is contained in U:

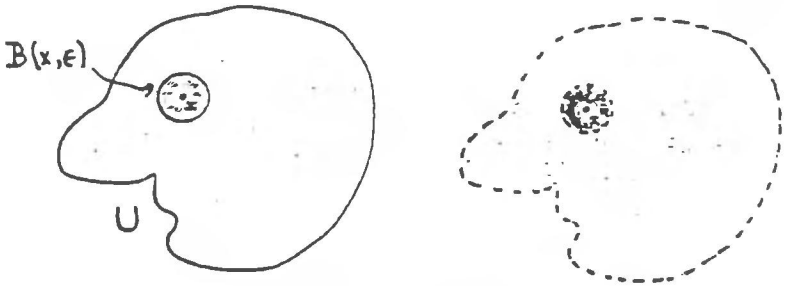


Figure 34. *How to be open in* $\mathbb{R}^2$

The "open" unit ball in $\mathbb{R}^2$, $\{y \mid \|y\|_2 < 1\}$, is indeed open in the sense of 3.1.1: Corollary 3.1.2 says just that. The "closed" unit ball (the official definition of *closed* comes later) $\{y \mid \|y\|_2 \leq 1\}$ is not open. To check this, take $y = (1, 0)$ and consider an arbitrary $B(y, \epsilon)$; it contains the point $z = (1 + \epsilon/2, 0)$ which does not satisfy $\|z\|_2 \leq 1$.

Intuitively: an open set should not contain any "boundary points," so perhaps one ought to draw its boundary as a dotted curve (Figure 34, right). This becomes tiresome after about the third sketch. ∎

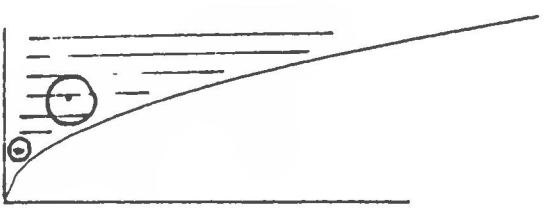


Figure 35. *An open set with very small* $B(x, \epsilon)$

3.1.10 Examples

A. Let M be a metric space with the discrete metric

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

In this case,

$$B(x, \frac{1}{2}) = \{y \mid d(x, y) < \frac{1}{2}\} = \{x\}.$$

If U is *an arbitrary* subset of M, then for any $x \in U$,

$$x \in \{x\} = B(x, \frac{1}{2}) \subset U,$$

which means that U is open. Hence, all subsets of M are open. (Such a metric was defined, and then discarded, in Example 1.1.45).

B. Let A be the following subset of $l^\infty(\mathbf{R}, \mathbf{N})$:

$$A = \{\mathbf{y} \mid |y_j| < 1/(j+1), j \geq 0\}.$$

This set is not open. For a proof, we need only find *one* $\mathbf{x} \in A$ for which no $B(\mathbf{x}, \epsilon)$ (as measured by the l^∞ metric) is contained in A. Define $\mathbf{x}$ by $x_j = 1/(j+2)$; since $|x_j| < 1/(j+1)$, $\mathbf{x}$ does belong to A. Now pick your favorite ϵ, and then let

$$y_j = \frac{1}{j+2} + \frac{\epsilon}{2}.$$

Since $|y_j - x_j| = \epsilon/2$ for all $j \geq 0$, we have $\|\mathbf{y} - \mathbf{x}\|_\infty < \epsilon$, so that $\mathbf{y} \in B(\mathbf{x}, \epsilon)$. It is not true, however, that $1/(j+2) + \epsilon/2 < 1/(j+1)$ for all $j \geq 0$, and therefore $\mathbf{y} \notin A$.

C. In $\mathbf{R}^2$, let

$$U = \{(x_1, x_2) \mid x_1 > 0, x_2 > \sqrt{x_1}\}.$$

The set is sketched in Figure 35. Evidently, the radii of the $B(\mathbf{x}, \epsilon)$ balls will become very small as $\mathbf{x}$ moves down towards the cusp. ∎

Subsection 1.2 introduced the notion of *equivalence of metrics*. It was claimed that that the numerical distance is often irrelevant, and that only "closeness" matters. Topology can now make those claims more precise.

3.1.11 Proposition. *Let M be a set equipped with two metrics, d_1 and d_2; let $\mathcal{T}_1$ and $\mathcal{T}_2$ be the corresponding metric topologies. Then d_1 and d_2 are equivalent if, and only if, $\mathcal{T}_1 = \mathcal{T}_2$.* ∎

Proof. Suppose first that d_1 and d_2 are equivalent. We need to show that the sets (of subsets of M) $\mathcal{T}_1$ and $\mathcal{T}_2$ coincide. This is done by proving that $\mathcal{T}_1 \subset \mathcal{T}_2$ *and* that $\mathcal{T}_2 \subset \mathcal{T}_1$. So, let $U \in \mathcal{T}_1$; we need to show that $U \in \mathcal{T}_2$.

To get U into $\mathcal{T}_2$, we must verify the condition from Proposition 3.1.1 for the metric space (M, d_2). Let $x \in U$. Is there a set $\{y \mid d_2(x, y) < \delta\} \stackrel{\text{def}}{=} B_2(x, \delta)$ containing x and contained in U?

Well, since $U \in \mathcal{T}_1$, there is a set $\{y \mid d_1(x, y) < \epsilon\} \stackrel{\text{def}}{=} B_1(x, \epsilon)$ with those properties. Then, because d_1 and d_2 are equivalent (see (1.1.2)) there is a $\delta > 0$ such that $x \in B_2(x, \delta) \subset B_1(x, \epsilon) \subset U$. This $B_2(x, \delta)$ is the d_2-ball required by Proposition 3.1.1, and indeed U is open in $\mathcal{T}_2$.

Next, let $V \in \mathcal{T}_2$. It must be shown to be in $\mathcal{T}_1$: that proof is exactly the same (just interchange 1 and 2.)

The converse is left as an exercise. ∎

3.1.12 Remark. We know that all the l^p metrics on $\mathbf{R}^n$ are equivalent. The more popular way to say this (which should now make sense) is: "all l^p metrics define the same topology on $\mathbf{R}^n$". Thus, any concept that can be formulated within the axiom system (3.0.1) for a topology on $\mathbf{R}^n$ will be independent of the choice of metric. ∎

• **3.1.13 Exercise.** Finish the proof of Proposition 3.1.11. Suppose that $\mathcal{T}_1 = \mathcal{T}_2$, and deduce that the condition in (1.2.1) is satisfied. You should make use of the fact (3.1.2) that balls of the type $B_i(x, \epsilon)$ are themselves open sets in the topology $\mathcal{T}_i$. ∎

I now want to extract the topological information contained in the family $\{B(x, \epsilon) \mid x \in M, \epsilon > 0\}$ of open balls in a metric space (M, d). In Proposition 3.1.1, the balls were used to define the open sets, and (this is a kind of consistency condition), the balls themselves are open sets according to Corollary 3.1.2. If you look at the proofs, you will see that the full machinery of "metric space" is not really needed. We used, of course, the trivial fact

i') $x \in B(x, \epsilon)$, for every x and every $\epsilon > 0$.

In the proof of (3.0.4), we needed to know that

ii') $B(x, \epsilon) \cap B(x, \eta)$ contains either $B(x, \epsilon)$ or $B(x, \eta)$.

The proof of Corollary 3.1.2 appealed to the triangle inequality in order to argue that

iii') If $y \in B(x, \epsilon)$, then some $B(y, \eta)$ is contained in $B(x, \epsilon)$.

In topology, one starts with a set X, and postulates the existence, for every $x \in X$, of a family $\mathcal{N}(x)$ of subsets of satisfying analogs of properties i'), ii'), iii'). The sets in $\mathcal{N}(x)$ are substitutes for the balls $B(x, \epsilon)$. The conditions i), ii), iii) below on $\mathcal{N}(x)$ are more general; they allow for "balls" that are less "symmetric" than the $B(x, \epsilon)$. Furthermore, the sets in $\mathcal{N}(x)$ need not themselves be open, but they do describe the open sets.

3.1.14 Example. For $x \in \mathbf{R}^2$ and $a, b > 0$, let $V(x, a, b)$ be the set of points inside *and on* the ellipse with center x and vertical and horizontal axes with diameters $2a$ and $2b$ respectively. Property ii') fails, because the intersection of concentric ellipses need not be an ellipse, and property iii') fails for points y on the ellipse itself. It may be intuitively clear, however, that the families $\mathcal{N}(x) = \{V(x, a, b) \mid a, b > 0\}$ describe the same notion of "closeness to x" as do the round balls $B(x, \epsilon)$. ∎

3.1.15 Definition. *Let X be a set, and for every $x \in X$, let there be given a collection $\mathcal{N}(x)$ of subsets of X satisfying the following conditions.*

i) If $V \in \mathcal{N}(x)$, then $x \in V$.

ii) If $V_1, V_2 \in \mathcal{N}(x)$, then there is a $V_3 \in \mathcal{N}(x)$ such that $V_3 \subset V_1 \cap V_2$.

iii) If $V \in \mathcal{N}(x)$, there exists a $W \in \mathcal{N}(x)$ such that a) $W \subset V$, and b) when $y \in W$, there is a $U \in \mathcal{N}(y)$ with $U \subset V$.

The collection $\{\mathcal{N}(x) \mid x \in X\}$ is called a local base. ∎

• **3.1.16 Exercise.** Verify that the elliptical neighborhoods in Example 3.1.14 satisfy conditions i), ii), and iii) for a local base. ∎

The counterpart of Proposition 3.1.1 and its Corollary is immediate (indeed, local bases are introduced precisely to generalize those results).

3.1.17 Proposition. *Let X be a set and $\{\mathcal{N}(x) \mid x \in X\}$ a local base. Define the collection $\mathcal{T}$ of subsets of X as follows: $U \in \mathcal{T}$ if, and only if, whenever $x \in U$ then there is a $V \in \mathcal{N}(x)$ such that $x \in V \subset U$. The collection $\mathcal{T}$ is a topology on X. If $V \in \mathcal{N}(x)$, then there exists an open set W such that $x \in W \subset V$. One says that $\{\mathcal{N}(x) \mid x \in X\}$ is a* ***local base for the topology*** *$\mathcal{T}$.*

Conversely, if $(X, \mathcal{T})$ is a topological space, then the system $\{\mathcal{N}(x) \mid x \in X\}$ with $\mathcal{N}(x) = \{U \mid U$ open and $x \in U\}$ is a local base for the topology $\mathcal{T}$. ∎

The last part of the proposition only says that every topology can be defined by at least one local base. Taking all the open sets containing a given x is hardly ever a convenient choice.

• **3.1.18 Exercise.** Prove Proposition 3.1.17. ∎

3.1.19 Examples

A. Let $X = [0, \infty)$. For each $x > 0$, let $\mathcal{N}(x)$ be the collection of all intervals $(x - c, x + c)$ with $x - c > 0$. Let $\mathcal{N}(0)$ be the collection of intervals of the form $[0, c)$, with $c > 0$. The system $\{\mathcal{N}(x) \mid x \in [0, \infty)\}$ is easily seen to be a local base. Furthermore, one checks (do so!) that the intervals $(x - c, x + c)$, resp. $[0, c)$, are themselves open. It may be surprising that $[0, c)$, which we normally think of as "half-open interval", is declared to be open. This happens, of course, because the set X stops at the point zero. The natural answer to "what points of X are close to 0" is exactly: "the *positive* numbers that are close to 0", i.e., the numbers in an interval $[0, c)$.

The topology on $[0, \infty)$ defined by our local base is, in fact, the metric topology arising from the usual absolute-value distance. The underlying metric is not needed to describe the local base and the open sets.

B. This example is studied in greater detail in Section 4. Let X be the set of all infinitely differentiable functions on R. It is a linear space. Let $\varphi \in X$. To describe when a function ψ is "close" to φ, we should–since the existence of all derivatives is characteristic of the space X— introduce a measure of the difference between the derivatives of ψ and those of φ. It is indeed possible to compare, simultaneously, all derivatives at all values of $x \in \mathrm{R}$ (Proposition 4.2.25), but that route is fairly inconvenient in practice. One usually proceeds as follows.

The analogs of the balls $B(x, \epsilon)$ will be the sets

$$U(\varphi, m, R, \epsilon) = \{\psi \mid \max_{0 \le k \le m} \sup_{-R \le x \le R} |\varphi^{(k)}(x) - \psi^{(k)}(x)| < \epsilon\}. \tag{3.1.20}$$

(Superscript k denotes the k^{th} derivative, with the zeroth derivative being the function itself.) In words: ψ belongs to the set (3.1.20) iff its first m derivatives

differ from those of φ by less than ϵ on the interval $[-R, R]$. A local base is defined by

$$\mathcal{N}(\varphi) = \{U(\varphi, m, R, \epsilon) \mid m = 0, 1, 2, \ldots,\ R > 0, \epsilon > 0\}.$$

The verification of the conditions in Definition 3.1.15 is left as exercise.

Note that $U(\varphi, m', R', \epsilon') \subset U(\varphi, m, R, \epsilon)$ when $m' \geq m$, $R' \geq R$, and $\epsilon' \leq \epsilon$. A function χ is "nearer" to φ than is ψ, if more derivatives of χ are closer to those of φ over a larger interval. We have not provided a numerical measure of the distance between ψ and φ. ∎

• **3.1.21 Exercise.** Verify conditions i), ii), and iii) in the definition of local base. (For iii), proceed as follows. Let $\chi \in U(\varphi, m, R, \epsilon)$; set

$$\eta = \max_{0 \leq k \leq m} \sup_{-R \leq x \leq R} |\varphi^{(k)}(x) - \chi^{(k)}(x)|,$$

and show that $U(\chi, m, R, (\epsilon - \eta)/2) \subset U(\varphi, m, R, \epsilon)$.) ∎

• **3.1.22 Exercise.** Show that every set (3.1.20) is itself open. ∎

• **3.1.23 Exercise.** One's knee-jerk reaction is probably to introduce the obvious metric on X:

$$d(\varphi, \psi) = \sup_{k \geq 0} \sup_{-\infty < x < \infty} |\varphi^{(k)}(x) - \psi^{(k)}(x)|.$$

Why doesn't this work? ∎

Subsection 3.2 Continuity

It is not yet supposed to be clear how a topology on a set X encodes a useful concept of "closeness". This subsection provides a bit of clarification by developing the definition of *continuous function* on a topological space $(X, \mathcal{T})$.

3.2.1 Example. Consider first a special case: a function f from $\mathbf{R}^2$ to $\mathbf{R}$. (Think of $f(x_1, x_2) = x_1^2 + x_2^2$, for instance.) Denote distance in $\mathbf{R}$ by $|\cdot - \cdot|$, the usual absolute value, and distance in $\mathbf{R}^2$ by $||\cdot - \cdot||_2$, the Euclidean distance.

f is continuous at (x_1^*, x_2^*) if : for every $\epsilon > 0$ there exists a $\delta > 0$ such that

$$|f(x_1, x_2) - f(x_1^*, x_2^*)| < \epsilon \quad (\text{in } \mathbf{R})$$

whenever

$$||(x_1, x_2) - (x_1^*, x_2^*)||_2 < \delta \quad (\text{in } \mathbf{R}^2.)$$

Draw a picture! ∎

Next, we define continuity for functions from one metric space to another.

3.2.2 Definition. *Let (M, d) and (N, d') be two metric spaces, and let $f : M \to N$. One says that f is* **continuous at** *$x_0 \in M$ if for every $\epsilon > 0$ there exists a $\delta > 0$, such that*

$$d'(f(x), f(x_0)) < \epsilon \quad (\text{in } N)$$

whenever

$$d(x, x_0) < \delta \quad (\text{in } M).$$

One says that f is **continuous** *if it is continuous at each $x_0 \in M$.* ∎

Finally, here is the definition for general topological spaces.

3.2.3 Definition. *Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces, and let $f : X \to Y$ be a function. One says that f is* **continuous at** *$x_0 \in X$ if:*

i) for every open set $V \subset Y$ containing $f(x_0)$,
ii) there is an open set $U \subset X$ containing x_0,
iii) so that $x \in U$
iv) guarantees that $f(x) \in V$.

One says that f is **continuous** *if it is continuous at all $x_0 \in X$.* ∎

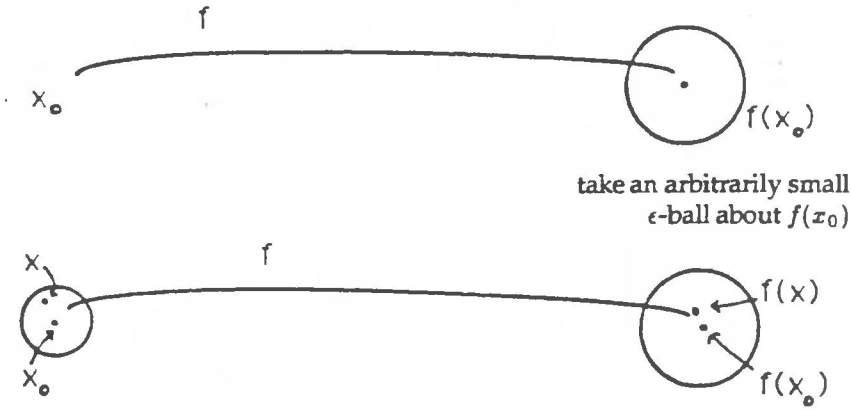
Items i) through iv) have been isolated to facilitate a point-by-point comparison with the metric space definition. Let X and Y be metric spaces with metrics d, d'. They also have topologies $\mathcal{T}, \mathcal{T}'$ as defined in (3.1.1). Assume that f is continuous at x_0 in the sense of the topology definition (3.2.3). We show how one deduces that it is also continuous in the sense of the metric definition (3.2.2).

i) V is an arbitrary open set containing $f(x_0)$. In particular, we are allowed to make it "small": choose it so that it is contained in $\{y \mid d'(f(x_0), y) < \epsilon\}$.
ii) There is a U containing x_0, with the property described in the continuity definition (3.2.3). Since U is open, there is a $\delta > 0$ such that $x_0 \in B(x_0, \delta) \subset U$.
iii) $d(x_0, x) < \delta$ guarantees that $x \in U$.
iv) Finally, $x \in U$ guarantees that $f(x) \in V$, which (by i) above) ensures that $d'(f(x_0), f(x)) < \epsilon$.

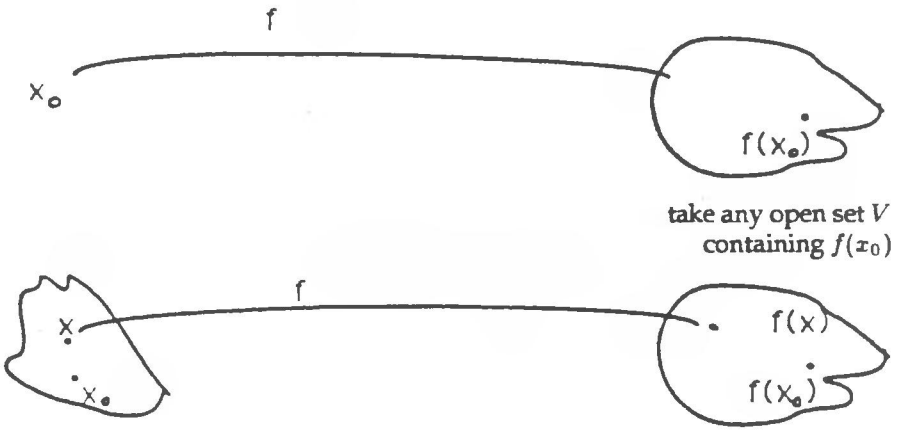
The converse, that Definition 3.2.2 implies Definition 3.2.3, is left for Exercise 3.2.39.

3.2.4 Example. At the risk of overdoing[B things, I want to draw a few more pictures. In the $\epsilon - \delta$ definition for metric spaces, we think of the regions around x_0 and $f(x_0)$ as little spheres (see the figures below). For topological spaces, we should at first think of the open sets U and V as shapeless blobs rather than spheres. This can get tedious, and need only be done until the idea that *there is no distance in general topological spaces* is firmly ingrained.

Remark. There are topological spaces that do not arise from any metric space (Proposition 3.3.8), so the correspondence between "closeness" and "open sets" as



Continuity in metric spaces



Continuity in topological spaces

just illustrated is intended mostly as a reference point. In general, if you want to think about "closeness", you should substitute "x and y are in an open set $U \in \mathcal{T}$" for "x and y are close." And in general, there is no answer to "how close are they?" When topological spaces are used, people simply stop thinking in terms of distance, and avoid loaded words like "close". ∎

Observe that, for metric spaces, Definition 3.2.3 appears to be more restrictive than Definition 3.2.2: it requires that something be true for *all* open sets containing the point $f(x_0)$, while (3.2.2) talks only about special open sets of the form $B(f(x_0), \epsilon)$. The difference is illusory, however, since the general open sets are characterized by containing some ball about each of their points. It should not come as surprise to learn that there is an analog of Definition 3.2.2 that makes use of local bases.

3.2.5 Definition. *Let $\{\mathcal{N}(X) \mid x \in X\}$ and $\{\mathcal{M}(y) \mid y \in Y\}$ be local bases on the sets X and Y (definition 3.1.15). Say that a function $f : X \to Y$ is* ***continuous at*** *x_0 if for every $V \in \mathcal{M}(f(x_0))$ there is a $U \in \mathcal{N}(x_0)$, such that $x \in U$ implies $f(x) \in V$. Again, f is called* ***continuous*** *if it is continuous at each $x_0 \in X$.* ∎

3.2.6 Proposition. *Let X and Y be equipped with local bases as in the preceding definition, and let $\mathcal{T}$, $\mathcal{S}$ denote the topologies determined on X and Y per Proposition 3.1.17. Then $f : X \to Y$ is continuous according to Definition 3.2.3 if and only if it is continuous according to Definition 3.2.5.* ∎

• **3.2.7 Exercise.** Prove this proposition. ∎

3.2.8 Example. Let X be the linear space of infinitely differentiable functions on $\mathbf{R}$, as in Example 3.1.19B. Define the function $F : X \to \mathbf{R}$ by

$$F(\varphi) = \int_{-1}^{1} (\varphi'(x))^2 \, dx.$$

Let us check that F is continuous. Suppose $F(\varphi) = c$. Given $\epsilon > 0$, we should find a set $U(\varphi, m, R, \delta)$ (cf. (3.1.20)) such that $\psi \in U(\varphi, m, R, \delta)$ implies $|F(\psi) - c| < \epsilon$. Note that this is Definition 3.2.5 of continuity.

Fix a function $\varphi \in X$. For every ψ,

$$(3.2.9)\quad |F(\psi) - c| \leq \int_{-1}^{1} |\psi'(x)^2 - \varphi'(x)^2| \, dx = \int_{-1}^{1} |\psi'(x) + \varphi'(x)| \cdot |\psi'(x) - \varphi'(x)| \, dx.$$

We need to make the expression on the right $< \epsilon$. Suppose ψ has the property that $\max_{-1 \leq x \leq 1} |\psi'(x) - \varphi'(x)| <$ some number $\eta < 1$. We want to figure out how small η must be to get (3.2.9) less than ϵ. Now,

$$|\psi'(x) + \varphi'(x)| = |(\psi'(x) - \varphi'(x)) + 2\varphi'(x)| \leq |\psi'(x) - \varphi'(x)| + 2|\varphi'(x)|.$$

Hence, for $x \in [-1, 1]$, we have $|\psi'(x) + \varphi'(x)| < 1 + 2M$, where we have defined $M = \max_{-1 \leq x \leq 1} |\varphi'(x)|$. We therefore find in (3.2.9) that

$$|F(\psi) - c| < \int_{-1}^{1} (1 + 2M) \cdot \eta \, dx = 2(1 + 2M)\eta.$$

Choose η to be less than $\epsilon/(2+4M) \stackrel{\text{def}}{=} \delta$ to get the desired conclusion: $\psi \in U(\varphi, 1, 1, \delta)$ implies that $|F(\psi) - c| < \epsilon$. ∎

Perhaps the most familiar definition of continuity is the one encountered at some point in a calculus course: $\lim_{x \to x_0} f(x) = f(x_0)$. Of course, this is just an abbreviation of 3.2.2[18]. Often, a variant involving sequences is convenient: if $x_n \to x$ then $f(x_n) \to f(x)$. This makes sense in general topological spaces, but is not necessarily equivalent to Definition 3.2.2. A brief discussion of the main points is worthwhile.

3.2.10 Definition. *Let (X, T) be a topological space, and let $\{x_n \mid n \geq 1\}$ be a subset of X. We call such a set a* **sequence in** *X. The sequence is said to converge to $x \in X$ (written $x_n \to x$) if for every neighborhood U of x there is an integer N_U (usually depending on U), such that $x_n \in U$ for every index $n \geq N_U$.* ∎

3.2.11 Proposition. *Let (M, d) and (M', d') be metric spaces. A function $f : M \to M'$ is continuous at $x \in M$ in the sense of Definition 3.2.2 if, and only if, the following condition holds:*

(3.2.12) whenever $x_n \to x$, then $f(x_n) \to f(x)$. ∎

Proof. Suppose f is continuous at x, and let $x_n \to x$. Pick an $\epsilon > 0$. By Definition 3.2.2 of continuity, there is a $\delta > 0$ such that $d(z, x) < \delta$ implies $d'(f(z), f(x)) < \epsilon$. Since $x_n \to x$, we will have $d(x_n, x) < \delta$ when $n >$ some N_δ, and then $d'(f(x_n), f(x)) < \epsilon$ for $n > N_\delta$. That means $f(x_n) \to f(x)$, and condition (3.2.12) holds.

Suppose now that (3.2.12) is true. We argue by contradiction: suppose that f is *not* continuous at x in the sense of Definition 3.2.2. That is, there exists an $\epsilon > 0$, and for each $\delta > 0$ a $z \in B(x, \delta)$, such that $d'(f(z), f(x)) \geq \epsilon$[19]. Let $\delta = 1, 1/2, 1/3, \ldots, 1/n, \ldots$ in succession, and let $x_1, x_2, \ldots$ be points (assumed to exist) which satisfy $d(x_n, x) < 1/n$ but $d'(f(x_n), f(x)) \geq \epsilon$. Then clearly $x_n \to x$, but $f(x_n) \not\to f(x)$, which contradicts (3.2.12). Hence, f must be continuous at x in the sense of 3.2.2. ∎

Half of the argument applies equally well in general topological spaces.

---

[18] Beginners usually read it "as x gets close to x_0, $f(x)$ gets close to $f(x_0)$". This is dangerously backwards. One should say "$f(x)$ will be arbitrarily close to $f(x_0)$ if only x is sufficiently close to x_0."

[19] The negation of "for every $\epsilon > 0$, blah" is "for at least one $\epsilon > 0$, *not* blah". In the definition of continuity, "blah" is "there exists a $\delta > 0$ such that bleah"; the negation is "for every $\delta > 0$, *not* bleah". Finally, "bleah" is "for every z with $d(z, x) < \delta$, $d'(f(z), f(x)) < \epsilon$", and the negation is "there is a z with $d(z, x) < \delta$ for which $d'(f(z), f(x))$ is *not* $< \epsilon$".

3.2.13 Proposition. *Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces, and suppose that $f : X \to Y$ is continuous at x in the sense of Definition 3.2.3. Then property (3.2.12) holds.* ∎

• **3.2.14 Exercise.** Prove this proposition. ∎

3.2.15 Example. In general, condition (3.2.12) does not entail continuity in the sense of 3.2.3. An example must involve a topology that does not arise from a metric (since for metric topologies, Proposition 3.2.11 does assert equivalence of the two definitions).

Introduce a topology $\mathcal{T}$ on R as follows. The open sets are $\emptyset$ and R, of course, and all sets whose complement is countable, i.e., sets of the form $\mathrm{R} - \{a_1, a_2, \ldots\}$. Verification of the topology axioms (3.0.2-5) is a useful exercise. This topology is sometimes called the *cocountable topology* on R.

Very few sequences converge in this strange topology. Indeed, suppose that $x_n \to x$. Take the open set $U = \{x\} \cup (\mathrm{R} - \{x_1, x_2, x_3, \ldots\})$. By definition of convergence, $x_n \in U$ for $n >$some N. But to form U we removed all points in the sequence from R, and added back the point x. If all $x_n, n > N$ still belong to U, it must be that they are all equal to x itself. Hence:

Only sequences which are eventually constant ($x_n = x$ for $n >$some N) can converge in the topology $\mathcal{T}$.

The condition "$x_n \to x$ implies $f(x_n) \to f(x)$" is therefore no restriction at all on the function f, since

$$x_n \to x \Rightarrow x_n = x \text{ for } n > \text{some } N \Rightarrow f(x_n) = f(x) \text{ for } n > N \Rightarrow f(x_n) \to f(x).$$

Note that the range of the function f is immaterial: if we take (3.2.12) as criterion for continuity, *every* function from $(\mathrm{R}, \mathcal{T})$ to *every* topological space is continuous!

Here is an example of an f from $(\mathrm{R}, \mathcal{T})$ to R with its usual topology which is not continuous in the sense of Definition 3.2.2. Let $f(x) = x^2$. Pick a point a. Take the interval $(a^2 - \epsilon/2, a^2 + \epsilon/2)$ as open set V containing $f(a)$. Every open set $U = \mathrm{R} - \{a_1, a_2, \ldots\}$ containing a contains arbitrarily large numbers x, and the corresponding function values $f(x) = x^2$ will also be arbitrarily large and therefore not contained in V. ∎

It is instructive to ask why the second half of the proof of Proposition 3.2.11 does not work in this example.Let us try to adapt it to the present setting. Assuming that (3.2.12) holds, let us suppose that the function f from $(\mathrm{R}, \mathcal{T})$ to a topological space $(Y, \mathcal{S})$ is not continuous at $x \in X$. There is then an open set V containing $f(x)$, such that for every open U containing x there is a $z_U \in U$ for which $f(z_U) \notin V$. To reach a contradiction, we would have to use some of the points z_U to build a sequence that converges to x; then $f(z_U) \notin U$ would be incompatible with (3.2.12). The problem is how to pick this convergent sequence. What we did in the Proposition was to choose a sequence of open sets U_n—namely, balls $B(x, 1/n)$—that "shrank down" to x. It is this "shrinking down" property of

the U_n that would force the corresponding points z_{U_n} to converge to x. In a very precise sense, that property fails for the cocountable topology. The details are left for the exercises. ∎

3.2.16 Exercise. A topological space $(X, \mathcal{T})$ is said to be *first countable* if the topology can be defined by a local base $\{\mathcal{N}(x) \mid x \in X\}$ for which each $\mathcal{N}(x)$ is a countable collection of sets. Prove that this definition can be rephrased as follows: for each $x \in X$, there exists a countable collection of open sets $\{U_1(x), U_2(x), \ldots\}$ containing x, such that whenever U is open and contains x, then $U_i(x) \subset U$ for some $U_i(x)$. ∎

3.2.17 Exercise. Suppose $(X, \mathcal{T})$ is first countable. A function f from X to a topological space $(Y, \mathcal{S})$ is continuous at x in the sense of Definition 3.2.2 if, and only if, condition (3.2.12) holds. ∎

3.2.18 Exercise. Prove that $\mathbf{R}$ with the cocountable topology in Example 3.2.15 is not first countable. (Hint: there are uncountably many countable sets of real numbers.) ∎

3.2.19 Exercise*. Can a nonconstant function be continuous from $\mathbf{R}$ with the cocountable topology in Example 3.2.15 to $\mathbf{R}$ with the usual topology? ∎

In the setting of topological spaces, there is another version of the definition of continuity. It dispenses with the ϵ and δ neighborhoods, or rather with the open sets U and V, and is as streamlined as a definition can ever be. It is also (next to the definition of $(X, \mathcal{T})$) *the fundamental definition in topology*. It is particularly convenient when one investigates general properties of continuous functions. The $\epsilon - \delta$–type Definition 3.2.2 is mostly used in the analysis of specific functions (e.g., Example 3.2.8).

To prepare the ground for this definition, we must first think about the interaction between "continuity" and "open set". These are our basic concepts so far, and we ask: where does a continuous function take an open set in the domain? where does an open set in the range come from?

3.2.20 Examples

A. The function $f(x) = \tan x$ maps the open interval $(-\pi/2, \pi/2)$ one-to-one onto the whole real axis (which is also open). So, a continuous function can *stretch* sets in its domain. (The inverse function arctan shows that a continuous function may also *shrink* sets in its domain.)

B. The function $f(x) = x^3 - 3x$ maps the open interval $(-\sqrt{3}, \sqrt{3})$ onto the closed interval $[-2, 2]$. So, a continuous function may map an open interval to a closed interval. Note that this f is not one-to-one. This same function maps the open interval $(-\sqrt{3}, 0)$ onto the half-open interval $(0, 2]$.

C. Define a function on $[0, \infty)$ by

$$f(x) = \begin{cases} 1 + x \sin x, & \text{whenever } \sin x \geq 0, \\ 1 + \frac{x}{1+x} \sin x, & \text{whenever } \sin x \leq 0. \end{cases}$$

This function is continuous, and maps $[0, \infty)$ onto $(0, \infty)$. (Its graph oscillates, getting ever closer to the x-axis at its minima, while its maxima get ever higher.) Only the point 0 is "lost" between range and domain. This function is not one-to-one.

D. Let $f(x, y) = (x^2, y)$. This is continuous from $\mathbf{R}^2$ into $\mathbf{R}^2$. It *folds* the plane along the y-axis (and stretches it in the horizontal direction).

E. There is no continuous function that maps the open interval $(-1, 1)$ onto the disjoint union $(-1, 0) \cup (0, 1)$. In words: a continuous function cannot *tear* an open set. To prove this, let $A = \{x \mid f(x) \in (-1, 0)\}$ and $B = \{x \mid f(x) \in (0, 1)\}$. Clearly, $A \cap B = \emptyset$. The set A is open. To see this, let $x \in A$, so that $f(x) \in (-1, 0)$. Pick ϵ so small that $V = (f(x) - \epsilon, f(x) + \epsilon) \subset (-1, 0)$. By continuity, there is an interval $(x - \delta, x + \delta)$ all of whose points map to V. This interval, by definition, is contained in A, so A is open. By Proposition 3.1.4, A is a countable union of disjoint intervals. Let (a, b) be one of those intervals. If $a = -1$ and $b = 1$, then f maps $(-1, 1)$ into $(-1, 0)$, contrary to assumption. So suppose, without loss of generality, that $b < 1$. The point b must belong to B, because the open intervals whose union is A are disjoint. But now B is open, by the same proof that showed A to be open, and so an interval about b belongs to B. This is impossible, since A and B are disjoint. ∎

These simple examples show that continuous functions can stretch, shrink, and fold, but they cannot tear. There are other things they cannot do or must do, and a portion of the subject of topology is concerned with analyzing such matters. It should be fairly clear, at least, that one cannot hope to characterize continuous functions by what they do to open sets. They can do too many different things.

Let us now ask the opposite question: where do open sets in the range of a continuous function come from? To answer this, we require the notion of *inverse of a function*. The main properties are stated as exercises; be sure to read them, even if you don't work them.

3.2.21 Notation. *Let X, Y be sets and let f be a function with domain contained in X and range contained in Y. For $A \subset Y$, set*

$$f^{-1}(A) = \{x \in X \mid f(x) \in A\}.$$

$f^{-1}(A)$ is called the inverse image of A; one says "f inverse of A". ∎

• **3.2.22 Exercise.** Define $f : \mathbf{R} \to \mathbf{R}$ by $f(x) = 1 + x^2$. Find $f^{-1}((2, 3))$, $f^{-1}(\{1\})$, $f^{-1}((-1, 0))$ and $f^{-1}(\mathbf{R})$. Note that $f^{-1}(A)$ may be the empty set. ∎

• **3.2.23 Exercise.** Let X, Y be sets, and let f be a function with domain $\subset X$ and range $\subset Y$. Let $\mathcal{A}$ be a collection of subsets of Y. Show that

a) $f^{-1}\left(\bigcup_{A \in \mathcal{A}} A\right) = \bigcup_{A \in \mathcal{A}} f^{-1}(A)$;

b) $f^{-1}\left(\bigcap_{A\in\mathcal{A}} A\right) = \bigcap_{A\in\mathcal{A}} f^{-1}(A)$;

c) $f(f^{-1}(A)) \subset A$. $A \subset Y$;

d) $f^{-1}(f(B)) \supseteq B$. $B \subset X$.

(Hint: remember the technique for proving equality of sets: each set is shown to be a subset of the other.) ∎

• **3.2.24 Exercise.** Find examples to show that the subset symbol in c), d) in the last exercise cannot be replaced by an = sign. ∎

In contrast to Examples 3.2.20, f^{-1}(an open set) is completely predictable.

3.2.25 Proposition. *Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces, and suppose that $f : X \to Y$ is continuous (Definition 3.2.2). If V is open in Y, then $f^{-1}(V)$ is open in X.* ∎

Proof. Let $x \in f^{-1}(V)$, so that $f(x) \in V$. By continuity, there exists an open U_x containing x such that $y \in U_x$ implies $f(y) \in V$. This says precisely that $U_x \subset f^{-1}(V)$. I now claim that

$$f^{-1}(V) = \bigcup_{x\in f^{-1}(V)} U_x.$$

Because the union of open sets on the right is again open, by (3.0.5), this would prove the assertion. Now, the left side is contained in the right side, because if $x \in f^{-1}(V)$, then $x \in U_x \subset \bigcup U_x$. The right side is contained in the left side, because $z \in \bigcup U_x$ implies $z \in$ some U_x, and by definition $U_x \subset f^{-1}(V)$. Equality follows, and the proposition is proved. ∎

In the course of the proof, we have established the following basic and useful fact.

3.2.26 Corollary (of the proof). *Let $(X, \mathcal{T})$ be a topological space. Suppose $V \subset X$ has this property: for every $x \in V$, there exists an open set U_x such that $x \in U_x \subset V$. Then V is open.* ∎

It is a surprising and fundamental (and by now quite straightforward) fact that "f^{-1}(open) =open" *characterizes* continuous functions.

3.2.27 Definition. *Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces. A function $f : X \to Y$ is said to be continuous if, for every open subset V of Y, the inverse image $f^{-1}(V)$ is an open subset of X.* ∎

3.2.28 Proposition. *Definitions 3.2.3 and 3.2.27 are equivalent.* ∎

Proof. Proposition 3.2.25 shows that Definition 3.2.3 implies Definition 3.2.27. To get the other direction, suppose that f is continuous according to 3.2.27, so that $f^{-1}(\text{open set})$ is open. If now V is any open neighborhood of $f(x_0)$, then $U \stackrel{\text{def}}{=} f^{-1}(V)$ is open. When $x \in U$, then $f(x) \in V$ by definition of f^{-1}. Hence, f is also continuous in the sense of the first definition. ∎

Here is one example of a situation in which the continuity criterion (3.2.27) is the most convenient.

3.2.29 Definition. *A topological space $(X, \mathcal{T})$ is said to be connected if there do not exist nonempty open sets $A, B \subset X$ such that $A \cap B = \emptyset$ and $A \cup B = X$. A subset $C \subset X$ is said to be connected if there do not exist nonempty open subsets A, B of X such that $(A \cap C) \cap (B \cap C) = \emptyset$ and $(A \cap C) \cup (B \cap C) = C$.* ∎

Example. The real line $\mathbf{R}$ with its usual topology is connected. The proof goes as consequence of an axiom about the real numbers, such as the "least upper bound axiom". The subset $[-2, -1] \cup [1, 2]$ of $\mathbf{R}$ is not connected. ∎

The next proposition explains Example E in (3.2.20).

3.2.30 Proposition. *Let $(X, \mathcal{T})$ be connected, and let f be continuous from $(X, \mathcal{T})$ onto the topological space $(Y, \mathcal{S})$. Then Y must be connected.* ∎

Proof. By contradiction. If Y is not connected, then $Y = A \cup B$ for nonempty, disjoint open sets A, B. Because f is continuous, $f^{-1}(A)$ and $f^{-1}(B)$ are open in X. They are nonempty and disjoint, and their union is X. This contradicts the connectedness of X. ∎
More information about connected sets may be found in Exercises II.1.1.49-54.

3.2.31 SUPPLEMENT: Homeomorphisms

We have, so far, talked about *topologies* defined on sets, introducing the concept as a generalization of metric space. There is also the subject called *Topology*, whose ultimate goal is to provide a list of all possible distinct topological spaces. That aim is too ambitious; it will never be achieved. A little discussion is still worthwhile, because it introduces some important ideas.

3.2.32 Definition. *Two topological spaces $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ are said to be homeomorphic if there exists a one-to-one function f with domain X and range Y, such that $f : X \longrightarrow Y$ and $f^{-1} : Y \longrightarrow X$ are both continuous. f is called a homeomorphism.* ∎

In other areas of mathematics, "isomorphic" is often used to say that two objects are to be considered identical. "Homeomorphic" is the standard term in topology. Homeomorphic topological spaces cannot be distinguished by any concept that makes sense in a general topological space.

3.2.33 Example. Let $X = (-1/2, 1/2)$ and $Y = \mathbf{R}$, both with their usual topologies. Define $f : X \longrightarrow Y$ by $f(x) = \tan \pi x$. f is one-to-one, and it and its inverse are both continuous. Hence $(-1/2, 1/2)$ and $\mathbf{R}$ are homeomorphic. As topological spaces, they are indistinguishable. True, $(-1/2, 1/2)$ has finite length and $\mathbf{R}$ has infinite length, but the concept of "length" is an addition to the structure defining a topology. One must go beyond the axioms for a "topology" to define "length". ∎

Roughly speaking, two topological spaces are homeomorphic if you can stretch one of them (without tearing it, Proposition 3.2.30) so that it fits on top of the other one. In the popular press, topology is therefore referred to as "rubber-sheet geometry". Just as Euclidean geometry considers two figures to be the same if they are congruent, so topology considers two figures to be the same if one can be continuously deformed into the other.

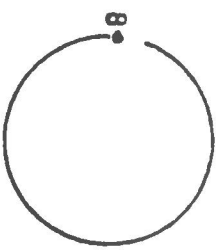
3.2.34 Example. Is a circle homeomorphic to a line? Intuitively, the answer is "no": to turn a circle into a line, you would have to tear it. It is not so easy, however, to *prove* that there is no one-to-one continuous function from circle to line, with continuous inverse from line to circle. The circle and the line are both connected; what distinguishes between them is the fact that removal of one point from a line splits it into two disjoint pieces, while removal of one point from the circle leaves one connected piece. One says that the line is "simply connected", while the circle is connected but not "simply connected". ∎

To prove that two topological spaces are homeomorphic, one generally aims to construct a homeomorphism (the function f). To prove that two spaces are *not* homeomorphic, one often proceeds indirectly. One tries to find properties that are preserved by homeomorphisms (connectedness, for example); i.e., homeomorphic spaces either both have the property or neither has it. Then, if space X has such a property while space Y doesn't, one knows that X and Y cannot be homeomorphic.

• 3.2.35 Exercise. Show that the interior of the unit circle in the plane is homeomorphic to the plane (both with their usual metric topologies). ∎

• 3.2.36 Exercise. Is a circle homeomorphic to an ellipse? If yes, construct a homeomorphism. (First, do it by picture.) ∎

3.2.37 Exercise. There is a topological way of "adding a point at infinity to $\mathbf{R}$". Let $\overline{\mathbf{R}} = \mathbf{R} \cup \{\infty\}$; ∞ is just a symbol for a point to be added to $\mathbf{R}$. We define a topology on $\overline{\mathbf{R}}$. It consists of all the usual open subsets of $\mathbf{R}$, plus unions of an open subset of $\mathbf{R}$ and a set of the form $\{t \mid |t| > \text{some } N\} \cup \{\infty\}$ (draw a picture of the new open sets). The set $\overline{\mathbf{R}}$ with the topology just described is called the *one-point compactification of* $\mathbf{R}$. Show that the one-point compactification of $\mathbf{R}$ is homeomorphic to the circle. ∎



3.2.38 Exercise*. Prove or disprove: $[0, 1)$ and $(0, \infty)$ are homeomorphic. ∎

Miscellaneous Exercises

•**3.2.39 Exercise.** Let $(M, d), (N, d')$ be metric spaces with their metric topologies, $\mathcal{T}, \mathcal{T}'$. Suppose that $f : M \to N$ is continuous at $x_0 \in M$ according to the metric definition, (3.2.2). Prove that it is also continuous according to the topological definition, (3.2.3). The converse was proved after that definition. ∎

• **3.2.40 Exercise.** Define the real-valued function f on $\mathbf{R}$ by

$$f(x) = \begin{cases} x, & \text{if } x < 0, \\ 1 + x, & \text{if } x \geq 0. \end{cases}$$

Obviously, f is not continuous. Find an open set $U \subset \mathbf{R}$ for which $f^{-1}(U)$ is not open. ∎

• **3.2.41 Exercise.** Define $f(x) = x^2$ on $\mathbf{R}$. Using nothing but the definition of f and elementary estimates, show that $f^{-1}\big((0,1)\big)$ is open. ∎

•**3.2.42 Exercise.** Let $f : \mathbf{R} \to \mathbf{R}$, and suppose that $f^{-1}\big((a,b)\big)$ is open for every interval (a,b). Let $U \subset \mathbf{R}$ be an open set. Show, using Definition 3.1.1 of "metric topology", that $f^{-1}(U)$ is open.

Generalize this result to a function $f : M \to M'$ from one metric space to another — assume that $f^{-1}\big(B(x,\epsilon)\big)$ is open for every ball $B(x,\epsilon)$ in (M', d'), and prove that $f^{-1}(U)$ is open in (M, d) for every open set $U \subset M'$. ∎

• **3.2.43 Exercise.** Let M be the set of $n \times n$ real matrices, equipped with the metric

$$d(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (A_{ij} - B_{ij})^2}.$$

Taking for granted that the function f defined by $f(\mathbf{A}) =$ determinant of $\mathbf{A}$ is continuous, show that the set of invertible matrices is open in M. ∎

•**3.2.44 Exercise.** Show that the set of invertible $n \times n$ real matrices is not connected. ∎

• **3.2.45 Exercise.** Let $(X, \mathcal{T})$, $(Y, \mathcal{S})$, and $(Z, \mathcal{R})$ be topological spaces. Suppose that $f : X \to Y$ and $g : Y \to Z$ are continuous. Prove that the composite function $g \circ f : X \to Z$, defined by $g \circ f(x) = g(f(x))$, is continuous. ∎

3.2.46 Exercise. Let f and g be continuous functions from $\mathbf{R}$ to $\mathbf{R}$ (in the metric topology), and suppose, for simplicity, that the range of both functions is all of $\mathbf{R}$. Use Definition 3.2.27 to prove that the sum $f + g$ and the product fg are also continuous functions.

Generalize to complex-valued functions on a metric space (M, d). ∎

3.2.47 Exercise. Use the preceding exercise to prove that the determinant (in Exercise 3.2.23) is continuous. ∎

3.2.48 Exercise. Let $(X, \mathcal{T})$ be the topological space in Exercise 3.0.7. Make a list of all continuous functions from X to X. ∎

3.2.49 Exercise*. Let $f : \mathbf{R} \to \mathbf{R}$ be continuous (in the metric topology of $\mathbf{R}$), and suppose that the inverse f^{-1} is also a function (i.e., for each y in the range of f, $f^{-1}(y)$ is a single point). Is f^{-1} a continuous function from the range of f to $\mathbf{R}$? Prove, or give a counterexample. ∎

Subsection 3.3 Hausdorff Spaces

Topological spaces are constrained by very few axioms, and while a surprising number of results can be deduced from those axioms, there are many properties encountered in concrete situations that do not hold in general.

3.3.1 Example. This is a very popular picture:

Here x and y are points in a topological space X, and U and V are open sets surrounding them. Obviously, if we just take U and V "small enough," in some sense, then $U \cap V = \emptyset$. Unfortunately, this is just not true in arbitrary topological spaces.

Consider $X = \{-1, 0, 1\}$ and $\mathcal{T} = \{X, \emptyset, \{-1\}, \{1\}, \{-1, 1\}\}$ as in Exercise 3.0.7. Let $x = 0$ and $y = 1$. One cannot find open sets $U, V \in \mathcal{T}$ such that $x \in U$, $y \in V$, and $U \cap V = \emptyset$. ∎

Since many arguments in analysis rely on the existence of such separating open sets, general topology appears to be too general to be applicable. The solution is clear: add more axioms to the four in Definition 3.0.1.

3.3.2 Definition. *A topological space $(X, \mathcal{T})$ is said to be a* ***Hausdorff space***[20] *if it satisfies the following condition:*

> *for every pair of points $x, y \in X, x \neq y$*
> *there are open sets U, V, such that*
> $x \in U$, $y \in V$. and $U \cap V = \emptyset$. ∎

Remark. This new axiom is called "T_2." The letter "T" stands for the German word *Trennung*, which means *separation*. There are various degrees of separation, described by different axioms called T_1, $T_{\frac{1}{2}}$, and so on. See EDM II for more information. ∎

3.3.3 Example. Most of the familiar topological spaces are Hausdorff spaces. Consider, for example, a metric space (M, d) with its metric topology. Let $x, y \in M, x \neq y$, and let $a = d(x, y) > 0$. Then $B(x, a/3)$ and $B(y, a/3)$ are open and disjoint (prove!). ∎

It is perhaps more interesting that non-Hausdorff topological spaces also have some uses.

---

[20] After Felix Hausdorff, 1868-1942, one of the creators of general topology.

3.3.4 Example. Consider the differential equation $y'(t) = y(t)$, and let X be the set of *orbits*, i.e., the *set* of *subsets* of $\mathbf{R}$ of the form

$$\{y(t) \mid y'(t) = y(t) \text{ and } -\infty < t < \infty\}.$$

Since the solutions of $y'(t) = y(t)$ look like

$$y(t) = y_0 e^t,$$

the orbits, which can be written more explicitly as $\{y_0 e^t \mid -\infty < t < \infty\}$, come in three different types:

$$\begin{aligned} &i)\ y_0 > 0 \ \Rightarrow\ \{y_0 e^t \mid t \in \mathbf{R}\} = (0, \infty) \\ &ii)\ y_0 < 0 \ \Rightarrow\ \{y_0 e^t \mid t \in \mathbf{R}\} = (-\infty, 0) \\ &iii)\ y_0 = 0 \ \Rightarrow\ \{y_0 e^t \mid t \in \mathbf{R}\} = \{0\}. \end{aligned}$$

X, the set of orbits, can be thought of as $\{-1, 0, 1\}$:

$$\begin{aligned} -1 \text{ represents } &\{-e^t \mid -\infty < t < \infty\} = (-\infty, 0) \\ 1 \text{ represents } &\{e^t \mid -\infty < t < \infty\} = (0, \infty) \\ 0 \text{ represents } &\{0 \cdot e^t \mid -\infty < t < \infty\} = \{0\}. \end{aligned}$$

It turns out to be natural (see Example 3.4.17) to declare a subset Y of the orbit space $X = \{-1, 0, 1\}$ to be *open* if the set of all initial conditions giving rise to the orbits in Y is open in $\mathbf{R}$. All the possibilities can be listed.

1) The open set $0 < y_0 < \infty$ of initial conditions has orbit $(0, \infty)$, represented by $1 \in X$. So, $\{1\}$ is open.
2) The set $-\infty < y_0 < 0$ has orbit $(-\infty, 0)$, or $-1 \in X$. So, $\{-1\}$ is open.
3) The set $\{-\infty < y_0 < 0\} \cup \{0 < y_0 < \infty\}$ has orbits $(-\infty, 0) \cup (0, \infty)$, or $\{-1, 1\} \subset X$. So, $\{-1, 1\}$ is open.
4) The set $-\infty < y_0 < \infty$ gives rise to all three orbits, $(-\infty, 0) \cup \{0\} \cup (0, \infty) = \mathbf{R}$. So, X is open.
5) There is *no* open set of initial conditions with only $\{0\}$ as orbit. So, $\{0\}$ is *not* open.

The topology defined thereby on the orbit space X is:

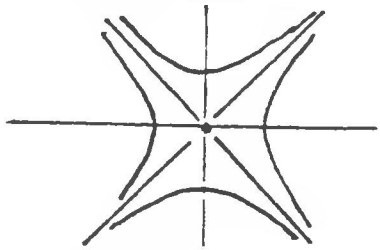
$$\mathcal{T} = \{\emptyset, X, \{1\}, \{-1\}, \{-1, 1\}\}.$$

This collection is a topology on X (Exercise 3.0.7), but it is not Hausdorff (Example 3.3.1). The construction of X may seem strange to the uninitiated, who do no more with differential equations than solve them, but in fact it is standard and useful. What is not so appealing (to the initiated) is the frequent appearance of non-Hausdorff topologies. ∎

3.3.5 Exercise. (This requires some knowledge of differential equations.) The system

$$\dot{x}(t) = y(t)$$
$$\dot{y}(t) = x(t)$$

has general solution $x(t) = a\cosh t + b\sinh t$, $y(t) = a\sinh t + b\cosh t$. The orbits are the curves $\{(x(t), y(t)) \mid t \in \mathbf{R}\}$ in the (x, y)-plane. They are the various branches and limiting cases of the hyperbolas $x^2 - y^2 = a^2 - b^2$:



Let X be the set of orbits. One point of X is one of the curves in the figure. The topology is defined as in the example above; a set U of orbits is declared to be open if the set of all initial conditions giving rise to the orbits in U is open in $\mathbf{R}^2$.

Show that there is no open subset $U \subset X$ containing the orbit $\{(0,0)\}$ that does not also contain the four straight-line orbits. Explain why the orbit space X is not Hausdorff.

Are there disjoint open sets $U, V \subset X$ each of which contains a single one of the straight-line orbits? ∎

3.3.6 Exercise*. Let M be the set of real, invertible 2×2 matrices. M can be made into a metric space with a metric derived from any one of the matrix norms from Subsection 2.3. It will not be a normed *linear* space, because the sum of two invertible matrices need not be invertible. From each possible matrix norm, M receives a metric topology. Prove that these topologies are all the same, and that a set $U \subset M$ is open if, and only if, the following condition holds:

$$\text{for each } m = (m_{i,j}) \in U, \text{ there is an } \epsilon > 0 \text{ such that}$$
$$\{p = (p_{i,j}) \in M \mid |m_{i,j} - p_{i,j}| < \epsilon \text{ for } i, j = 1,2\} \subset U. \blacksquare$$

3.3.7 Exercise*. This exercise continues the preceding one. Let $m \in M$. The *orbit through* m is the set of all matrices similar to m, i.e. the set $\{p \in M \mid p = gmg^{-1}$ for some invertible matrix $g\}$. Write $p \sim m$ if p is similar to m. Let X be the set of orbits; X lists all the invertible 2×2 matrices up to similarity. For example, the collection of all matrices

with eigenvalues 1, 2 may be thought of as a single point, represented by $\left(\begin{smallmatrix}1&0\\0&2\end{smallmatrix}\right)$. A subset $U \subset X$ is defined to be open if $\{m \in M \mid m \text{ lies on one of the orbits in } U\}$ is open in M.

Is X a Hausdorff space? (Hint: Observe that for fixed $\lambda \in \mathbf{R}$ and arbitrary $a \neq 0$,

$$\begin{pmatrix}\lambda & a\\0 & \lambda\end{pmatrix} \sim \begin{pmatrix}\lambda & 1\\0 & \lambda\end{pmatrix},$$

and let $|a| \to 0$.) ∎

One further consequence of the existence of non-Hausdorff topologies deserves mention.

3.3.8 Proposition. *Let $(X, \mathcal{T})$ be a topological space. If it does not have the Hausdorff property, then the topology cannot come from a metric on X.* ∎

Proof. We saw in Example 3.3.3 that metric topologies are Hausdorff. ∎

So, without realizing it, you have seen several examples of topological spaces in which the concept of "closeness" could not have been derived from any numerical measure of distance.

- **3.3.9 Exercise.** Prove: if $(X, \mathcal{T})$ is Hausdorff, then for every $x \in X$, the set $\{x\}$ is closed. Is the converse true? ("closed" has not yet been defined) ∎

- **3.3.10 Exercise.** Prove: in a Hausdorff space, every convergent sequence has a unique limit (compare Example 3.2.15). ∎

3.3.11 Exercise. Let $(X, \mathcal{T})$ be first countable (Exercise 3.2.16), and suppose that every convergent sequence has a unique limit. Show that $(X, \mathcal{T})$ is Hausdorff. ∎

3.3.12 Exercise. Suppose the topological spaces $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ are homeomorphic. Suppose that $(X, \mathcal{T})$ is Hausdorff. Show that then $(Y, \mathcal{S})$ is also Hausdorff. ∎

3.3.13 SUPPLEMENT: The Zariski topology

It is quite possible for one set X to be equipped with several distinct topologies. The usual topology on $\mathbf{R}$ is a Hausdorff topology, but there is another well-known topology on $\mathbf{R}$ which, it turns out, is not Hausdorff. The Zariski[21] topology is of overwhelming importance in algebraic geometry. It plays a minor role in analysis, but is still an instructive example.

The Zariski topology on R. A set $U \subset \mathbf{R}$ is declared to be open, i.e. to belong to the Zariski topology $\mathcal{T}$, if its complement $U^c = \mathbf{R} - U$ is a *finite* set, or if $U = \mathbf{R}$ or $\emptyset$, as required by Definition 3.0.1.

This may seem rather peculiar, but the definition has its uses. Before explaining where these open sets arise, I will verify the axioms:

[21] Oscar Zariski, 1899-1986, algebraist and geometer.

(3.0.2,3) $\emptyset \in \mathcal{T}$ and $\mathbf{R} \in \mathcal{T}$ by definition.

(3.0.4) If $U \in \mathcal{T}$ and $V \in \mathcal{T}$, then U^c and V^c are finite sets, and hence

$$(U \cap V)^c = U^c \cup V^c$$

is also a finite set. So $U \cap V \in \mathcal{T}$. Note, incidentally, that a countable intersection need not be in $\mathcal{T}$:

$$\left(\bigcap_{j=1}^{\infty} U_j\right)^c = \bigcup_{j=1}^{\infty} U_j^c,$$

and a countable union of finite sets is usually not finite (unless almost all the sets are empty.)

(3.0.5) Let $\mathcal{U}$ be any collection of open sets. Then

$$\left(\bigcup_{U \in \mathcal{U}} U\right)^c = \bigcap_{U \in \mathcal{U}} U^c = \text{ an intersection of finite sets.}$$

This intersection is obviously finite, and so $\bigcup_{U \in \mathcal{U}} U$ is open.

The Zariski topology generalizes to $\mathbf{R}^n$ and $\mathbf{C}^n$, but not in the way you might expect. The definition given for $\mathbf{R}$ is completely misleading. We must rephrase it to reveal the motivation behind this topology.

3.3.14 Definition. *The Zariski topology on* $\mathbf{R}$. *A subset* $U \subset \mathbf{R}$ *is open in the Zariski topology if* $U = \emptyset$ *or* $\mathbf{R}$ *or if there is a polynomial* $p_U(x)$ *such that* $U = \{x \mid p_U(x) \neq 0\}$. ∎

In words: the Zariski open subsets of $\mathbf{R}$ are complements of the set of zeros of polynomials. Since every nonzero polynomial of one variable has at most finitely many zeros, the open sets in the Zariski topology on $\mathbf{R}$ are complements of finite sets, as stated in the first definition. For example, take $U = \mathbf{R} - \{x_1, \ldots, x_n\}$ (so that $U^c = \{x_1, \ldots, x_n\}$); then $p_U(x) = \prod_{j=1}^{n}(x - x_j)$, and

$$\mathbf{R} - \{x_1, \ldots, x_n\} = \{x \mid \prod_{j=1}^{n}(x - x_j) \neq 0\}.$$

3.3.15 Definition. *The Zariski topology on* $\mathbf{R}^2$. *Let* $\{p_1, \ldots, p_n\}$ *be a finite set of polynomials in* (x, y). *Then the set*

$$U = \{(x, y) \mid p_j(x, y) \neq 0 \text{ for } j = 1, \ldots, n\}$$

is declared to be open. (The number n *and the polynomials* $p_1, \ldots, p_n$ *are arbitrary). Furthermore, those are all the open sets.* ∎

Note that $\emptyset$ and $\mathbf{R}^2$ are covered by this definition. The empty set is the complement of the set of zeros of the polynomial $p(x, y) \equiv 0$, and $\mathbf{R}^2$ is the complement of the set

of zeros of the polynomial $p(x, y) \equiv 1$. The set of solutions of a polynomial equation $p(x, y) = 0$ is called an *algebraic curve*. (E.g.: $x^2+y^2-1=0$, $y-x^2=0$, $x^3-y^2=0, \ldots$). The Zariski-open sets are, therefore, the complements of one or more (but always finitely many) algebraic curves.

The Zariski topology is important in algebraic geometry because that subject is concerned with the solutions of polynomial equations. The objects of study of algebraic geometry, therefore, are complements of Zariski-open sets, and the connection between polynomial equations and topology turns out to be useful. The verification of the axioms (3.0.1) is elementary only for R; in general, it relies on some fundamental theorems from algebra about the structure of polynomial rings.

• **3.3.16 Exercise.** Prove that the Zariski topology on R is not Hausdorff. ∎

• **3.3.17 Exercise.** Let $\mathcal{T}_1$ be the Zariski topology and $\mathcal{T}_2$ the usual (metric) topology on R. Show that $\mathcal{T}_1 \subset \mathcal{T}_2$, but that $\mathcal{T}_2$ is not a subset of $\mathcal{T}_1$. ∎

• **3.3.18 Exercise.** Continue the notation of the preceding example. Suppose that $f : (\mathrm{R}, \mathcal{T}_1) \to (\mathrm{R}, \mathcal{T}_1)$ is continuous. Is the same f also continuous from $(\mathrm{R}, \mathcal{T}_1)$ to $(\mathrm{R}, \mathcal{T}_2)$? From $(\mathrm{R}, \mathcal{T}_2)$ to $(\mathrm{R}, \mathcal{T}_1)$? From $(\mathrm{R}, \mathcal{T}_2)$ to $(\mathrm{R}, \mathcal{T}_2)$? Answer the analogous questions, starting from the assumption that f is continuous from $(\mathrm{R}, \mathcal{T}_i)$ to $(\mathrm{R}, \mathcal{T}_j)$ for some other pair i, j. ∎

3.3.19 Exercise. Equip R with the Zariski topology (3.3.13). Let $x_1, x_2, x_3, \ldots$ be a sequence consisting od distinct numbers, and let $x \in \mathrm{R}$ be arbitrary. Show that x_n converges to x. ∎

3.3.20 Exercise. Let M be the set of all quadratic equations $x^2 + bx + c = 0$ with real coefficients. One may identify M with $\mathrm{R}^2 = \{(b, c)\}$ in an obvious way; a topology on R^2 then induces a topology on M. Show that the set of all quadratic equations with *distinct* roots is Zariski open in M. (Later in this chapter we will interpret this property as giving a precise meaning to the statement: most quadratic equations have distinct roots.) ∎

3.3.21 Exercise. Let X be the set of all cubic equations $x^3 + px + q = 0$; identify it with R^2 and give it the Zariski topology. Show that the set of equations with distinct roots is a Zariski open set. (Hint: a cubic $f(x)$ of the given form has a double or triple root at x_0 if $f(x_0) = f'(x_0) = 0$.)

The absence of the x^2 term in the cubic makes no difference. One can transform x^3+ax^2+bx+c into $x^3 + px + q$ by a substitution $x \mapsto x - a/3$. This keeps distinct roots distinct. ∎

3.3.22 Exercise. Show that R with the usual topology is not homeomorphic to R with the Zariski topology. ∎

Subsection 3.4 How to Define Topologies

Let X be a set — for example, $X = \mathrm{C}^n$ or $X = l^2(\mathrm{R}, \mathrm{N})$. We want to define a topology $\mathcal{T}$ on X. How can we do this? Certainly not by listing all the sets in $\mathcal{T}$. In most examples, there are far too many open sets, and they are much too complicated, for such a direct procedure to succeed. If you look back at Proposition 3.1.1, you will see that we defined the metric topology only implicitly; the open sets U are characterized by a certain property:

if $x \in U$, then there is an ϵ-ball $B(x, \epsilon)$ such that $x \in B(x, \epsilon) \subset U$.

The open balls $B(x, \epsilon)$ are the simplest open sets (Corollary 3.1.2). They are used to build the general open set in Proposition 3.1.1. One can make their role as building blocks even more explicit (compare the proof of Proposition 3.2.25):

3.4.1 Proposition. *Let (M, d) be a metric space, and let $\mathcal{T}$ be the corresponding metric topology. Then $U \in \mathcal{T}$ if, and only if, U is a (generally uncountable) union of balls $B(x, \epsilon)$ (x will, and ϵ may, change from ball to ball.)* ∎

Note, that this is a special case of Corollary 3.2.26! It should not have been proved again.

Proof. Since a ball $B(x, \epsilon)$ is open (see (3.1.2)), a union of any number of such balls is open, by axiom (3.0.5). Conversely, suppose that $U \in \mathcal{T}$. For each $x \in U$, there is a ball $B(x, \epsilon_x) \stackrel{\text{def}}{=} B_x$ satisfying $x \in B_x \subset U$; this follows from the definition of "U is open" in Proposition 3.1.1. I claim that

$$\bigcup_{y \in U} B_y = U.$$

To prove that two sets are equal, one usually shows that each is a subset of the other. Let $x \in U$. Then $x \in B_x$ by construction, and therefore x belongs to the union. Thus,

$$U \subset \bigcup_{y \in U} B_y.$$

Suppose, conversely, that x belongs to the union. Then by definition of "union", $x \in$ some B_y; but since $B_y \subset U$, we have $x \in U$. Hence

$$\bigcup_{y \in U} B_y \subset U.$$

Equality follows. ∎

There are as many balls B_y in this union as there are points in U (usually, uncountably many). Furthermore, the B_y may overlap — this is not a disjoint union. Still, the proposition describes one method for building complicated open sets from simple open sets, and one gets *all* the open sets that way.

In this subsection, we will describe a generalization of this method, which lets one define other topologies of interest in terms of simpler building blocks called *bases* and *subbases*. Several typical applications are also presented.

3.4.2 Proposition. *Let X be a set, and let $\mathcal{B}$ be a collection of subsets of X satisfying the following conditions:*

$$\bigcup_{B \in \mathcal{B}} B = X; \tag{3.4.3}$$

$$\begin{aligned} &\text{if } B_1, B_2 \in \mathcal{B}, \text{ and } x \in B_1 \cap B_2, \text{ there is} \\ &\text{a } B_3 \in \mathcal{B} \text{ such that } x \in B_3 \subset B_1 \cap B_2. \end{aligned} \tag{3.4.4}$$

Then there is a unique topology $\mathcal{T}$ characterized by:

$$U \in \mathcal{T} \text{ iff for each } x \in U \text{ there is a } B \in \mathcal{B} \text{ such that } x \in B \subset U. \tag{3.4.5}$$ ∎

The proof is left for the exercises. The important property (3.4.4) is illustrated in Figure 36.

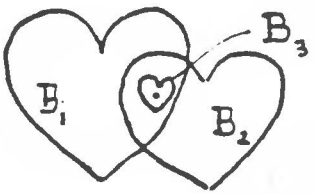


Figure 36. *A base of hearts*

Terminology. A collection of sets $\mathcal{B}$ satisfying (3.4.3), (3.4.4) is called a ***base*** for the topology $\mathcal{T}$ defined by (3.4.5). $\mathcal{T}$ is said to be ***generated by*** $\mathcal{B}$. ∎

Condition (3.4.3) should be compared with property ii) in the definition of *local base* in (3.1.15), and the connection (3.4.5) between base and topology with the connection, in Proposition 3.1.17, between local base and topology. The concepts are related, as one might guess when seeing the word "base" in both cases. The sets in a local base need not be themselves open, whereas the sets in a base must be open (as follows from (3.4.5) when $U \in \mathcal{B}$). The sets in a local base try to describe the open sets surrounding each given point of the space; the sets in a base are used to construct the open sets without reference to individual points. These slight differences notwithstanding, "base" and "local base" accomplish the same thing by essentially similar methods.

- **3.4.6 Exercise.** Prove that every $B \in \mathcal{B}$ also belongs to the collection $\mathcal{T}$ defined by (3.4.5). This is a generalization of (3.1.2). ∎
- **3.4.7 Exercise.** Prove Proposition 3.4.2, following the argument in (3.1.1). The collection $\mathcal{B}$ takes the place of the collection of all possible balls $B(x, \epsilon)$, as x ranges over X and ϵ ranges over $(0, \infty)$. ∎
- **3.4.8 Exercise.** Let $\mathcal{B}$ be a base for the topology $\mathcal{T}$. Prove that every $U \in \mathcal{T}$ can be written as a union of sets in $\mathcal{B}$. Follow the argument in Proposition 3.4.1. This result justifies the term *base*. ∎

Even bases are often too unwieldy, and so one introduces yet simpler building blocks.

3.4.9 Definition. *Let $\mathcal{B}^0$ be a collection of subsets of X with the following property: the collection $\mathcal{B}$ of all finite intersections of members of $\mathcal{B}^0$ is a base (i.e., satisfies (3.4.3) and (3.4.4)). Then $\mathcal{B}^0$ is called a* ***subbase***. ∎

3.4.10 Proposition. *Let $\mathcal{B}^0$ be any collection of subsets of X such that*

$$\bigcup_{B \in \mathcal{B}^0} B = X.$$

Then $\mathcal{B}^0$ is a subbase, i.e., the family $\mathcal{B}$ of all finite intersection of members of $\mathcal{B}^0$ is a base for a topology. ∎

Proof. We must check that $\mathcal{B}$ satisfies (3.4.3) and (3.4.4). The first condition holds by hypothesis. The second condition is satisfied in a trivial way. If $B_1, B_2 \in \mathcal{B}$, then B_1 and B_2 are finite intersections of sets in $\mathcal{B}^0$. Hence $B_1 \cap B_2$ is also such a finite intersection, which belongs to $\mathcal{B}$; it can serve as the set B_3 in (3.4.4). ∎

These concepts are now illustrated by a series of examples, many of which are stated as exercises. If you cannot (or do not want to) work all these exercises, at least read them to get a feeling for the application of the definitions.

• **3.4.11 Exercise.** Let $\mathcal{B}$ be the collection of all squares in $\mathbf{R}^2$ of the form $S(x, \epsilon) \stackrel{\text{def}}{=} \{y \mid |x_1 - y_1| < \epsilon, |x_2 - y_2| < \epsilon\}$. (These are the "balls" in the l^∞ metric.) Show that $\mathcal{B}$ is a base for a topology, i.e., verify (3.4.3) and (3.4.4). ∎

• **3.4.12 Exercise.** The topology defined in the preceding exercise is just the usual metric topology on $\mathbf{R}^2$. This follows from the equivalence of the l^2 and l^∞ metrics via (3.1.11). There is also a concept of *equivalence of bases* which does not depend on the availability of a metric.

Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be collections of subsets of X, both satisfying (3.4.3) and (3.4.4). We say that $\mathcal{B}_1$ and $\mathcal{B}_2$ are *equivalent*, if:

whenever $x \in B_1 \in \mathcal{B}_1$, there is a $B_2 \in \mathcal{B}_2$ such that $x \in B_2 \subset B_1$, and whenever $x \in B_2 \in \mathcal{B}_2$, there is a $B_1 \in \mathcal{B}_1$ such that $x \in B_1 \subset B_2$.

Compare this with the definition of equivalence of metrics, (1.2.2). Draw a picture! Now prove:

3.4.13 Proposition. *If two bases $\mathcal{B}_1$ and $\mathcal{B}_2$ are equivalent, then they generate the same topology.*

Then use the concept of equivalence of bases to prove that the squares and the circles generate the same topology on $\mathbf{R}^2$. ∎

• **3.4.14 Exercise.** An *open half-plane* in $\mathbf{R}^2$ is a set of the form

$$P(\mathbf{a}, r) \stackrel{\text{def}}{=} \{(x_1, x_2) \mid a_1 x_1 + a_2 x_2 > r\},$$

where $\mathbf{a} = (a_1, a_2) \in \mathbf{R}^2$ and $r \in \mathbf{R}$. Let $\mathcal{B}^0$ be the collection of all open half-planes. The collection $\mathcal{B}$ of all finite intersections of members of $\mathcal{B}^0$ is a base for a topology by Proposition 3.4.10. Show that this topology is just the usual metric topology on $\mathbf{R}^2$. (Hint: begin by drawing intersections of several sets from $\mathcal{B}^0$, say of the half-planes $x_1 < 1, x_1 > -1, x_2 < 1, x_2 > -2$.)

Remark. This subbase $\mathcal{B}^0$ can be generalized to normed linear spaces of infinite dimension. The topology it generates is then called the *weak topology*, which plays an important role in (for example) the analysis of partial differential equations. We will say a little more about it in Subsection 7. ∎

• **3.4.15 Exercise.** Explain why the collection of half-planes in the preceding exercise is not a base. ∎

3.4.16 Exercise. A coin is tossed every second. This goes on till the end of time, $n = \infty$. The outcome of this experiment is encoded in an infinite sequence$(\alpha_n)_{n=0}^{n=\infty}$, where each α_n is either an "H" (heads) or a "T" (tails). Let X be the set of all such sequences.

The following collection $\mathcal{B}^0$ of subsets of X is a subbase for a topology on X:

$$A(n,H) = \{\alpha \mid \alpha_n = H\}, \quad A(n,T) = \{\alpha \mid \alpha_n = T\},$$

with $n \in \mathbf{N}$ arbitrary.

i) Which, if any, of the following sets are open:
   a) the set of sequences in which the first 100 tosses are heads;
   b) the set of sequences in which there are infinitely many heads;
   c) the set of sequences $(H, H, \ldots, H, T, T, \ldots, T, H, \ldots)$, in which the initial run of heads is longer than the subsequent run of tails.

ii) Is the function $f : X \to \mathbf{R}$, defined by $f(\alpha)$ = the number of heads in the first 1000 tosses in the sequence α, continuous from X with the topology given by $\mathcal{B}^0$, to $\mathbf{R}$ with the usual topology? ∎

A particularly useful variant of the method of bases and subbases is applied in the following setting. One has a topological space $(Y, \mathcal{S})$, a set X that is to be equipped with a topology, and a function or a set of functions ***from Y to X or from X to Y***. These functions are to be used to induce a topology on X from the given topology on Y.

3.4.17 Example. Let $(Y, \mathcal{S})$ be $\mathbf{R}$ with its usual topology. Let $X = \{-1, 0, 1\}$ as in Example 3.3.4. Let p (for "projection") be the function that assigns to $y_0 \in \mathbf{R}$ the orbit representative, -1, 0 or 1. In (3.3.4), we declared a subset $U \subset X$ to be open if, and only if,

$$\{y \in \mathbf{R} \mid p(y) \in U\}$$

is open in $\mathbf{R}$. Figure 37 tries to convey the nature of this construction.

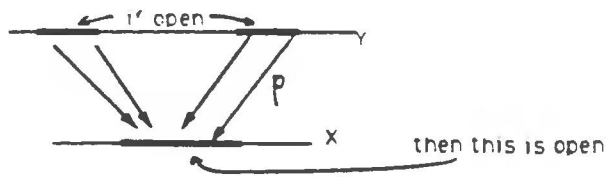


Figure 37. ***Defining a topology on X from one on Y — method #1***

We look at a subset of X and see "where it comes from." If it comes from an open set, it is itself declared to be open in X.

In fewer words: we declare $U \subset X$ to be open if, and only if, $p^{-1}(U) \subset Y$ is open. This means: we were given a topology on Y, and we define a topology on X so that the projection function p is continuous (p^{-1}(open set in X) = open set in Y). ∎

3.4.18 Example. Let $X = \mathbf{R}^2$, and let $(Y, \mathcal{S})$ be $\mathbf{R}$ with its usual topology. For each $\mathbf{a} \in \mathbf{R}^2$, let $f_{\mathbf{a}} : X \to Y$ be the function defined by $f_{\mathbf{a}}(x) = a_1x_1 + a_2x_2$. A subset of X of the form

$$U(\mathbf{a}, B) \stackrel{\text{def}}{=} \{x \mid f_{\mathbf{a}}(x) \in B \in \mathcal{S}\}$$

is declared to be open in X; here $\mathbf{a}, B$ vary over $\mathbf{R}^2$ and $\mathcal{S}$, respectively. This collection of sets is only a *subbase* for a topology on X. It contains all the open half-planes of Exercise 3.4.14, and generates the usual metric topology on $X = \mathbf{R}^2$ as was asserted in that Exercise. Figure 38 shows how this setup differs from the one in the preceding example.

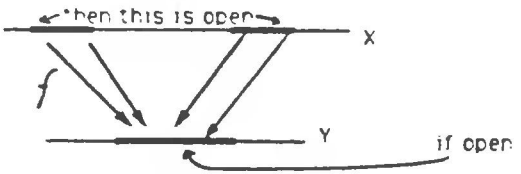


Figure 38. *Defining a topology on X from one on Y — method #2*

Take any open set in the range, and ask "where does it come from." The *preimage* is declared to be open in X.

The functions $f_{\mathbf{a}}$, by this definition, are continuous: $f_{\mathbf{a}}^{-1}(\text{open set in } Y) =$ an open set in X. ∎

The constructions of Examples 3.4.17 and 3.4.18 are summarized in a proposition.

3.4.19 Proposition. *a) Let $(Y, \mathcal{S})$ be a topological space, and let p be a function from Y to a set X. The set $\mathcal{T}$ of subsets of X defined by*

$$U \in \mathcal{T} \iff p^{-1}(U) \in \mathcal{S}$$

is a topology on X, and p is a continuous function from $(Y, \mathcal{S})$ to $(X, \mathcal{T})$.

b) Let $(Y, \mathcal{S})$ be a topological space, and let $\mathcal{F}$ be a set of functions from a set X to Y. The set $\mathcal{B}^0$ of subsets of X defined by

$$U \in \mathcal{B}^0 \iff U = f^{-1}(V) \text{ for some } V \in \mathcal{S} \text{ and some } f \in \mathcal{F}$$

is a subbase for a topology $\mathcal{T}$ on X, and all functions $f \in \mathcal{F}$ are continuous from $(X, \mathcal{T})$ to $(Y, \mathcal{S})$. ∎

• **3.4.20 Exercise.** Prove this proposition. Mimic Examples (3.4.17,18). ∎

3.4.21 Example. Let X be the set of all lines through the origin in $\mathbf{R}^n$. This is called the $n-1$-dimensional *real projective space*. We can use Proposition 3.4.19 to produce a "natural" topology on X. Namely, let $p : \mathbf{R}^n \to X$ be the projection, $p(x) =$ the unique line through the origin that passes through x. The topology $\mathcal{T}$ on X is then defined by: $U \subset X$ is open if, and only if, $p^{-1}(U)$ is an open subset of $\mathbf{R}^n$.

This definition should make perfect sense once you look at a picture for the case $n = 2$. The set of lines indicated on the left of Figure 39 is open, because the set of *points* on all those lines is open in $\mathbf{R}^2$.

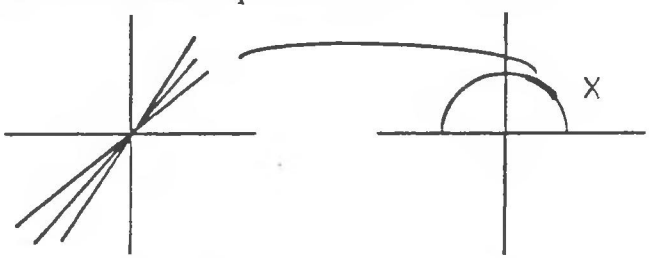


Figure 39. *The 1-dimensional real projective space*

Even more explicitly (still for $n = 2$), we can associate to the line $l_t \stackrel{\text{def}}{=}: x_1 = r\cos t, x_2 = r\sin t, r \in \mathbf{R}$, the point $(\cos 2t, \sin 2t)$ on the unit circle. We multiply t by 2 because the lines l_t and $l_{t+\pi}$ are in fact the same, and should be represented by a single point. Thus, the 1-dimensional projective space X can be identified with the unit circle, and the topology on X is just the usual metric topology on that circle. ∎

• **3.4.22 Exercise.** Let X be the set of two-dimensional planes through the origin in $\mathbf{R}^3$. Use Proposition 3.4.19 to define a topology on X. (Functions defined on planes are encountered in image processing problems.) ∎

3.4.23 Exercise. Refer to Proposition 3.4.19 b). Consider *any* topology $\mathcal{T}'$ on X with respect to which all the functions in $\mathcal{F}$ are continuous. Then $\mathcal{T}'$ contains the topology $\mathcal{T}$ defined in 3.4.19 as a subcollection. In other words, every subset of X that is open with respect to $\mathcal{T}$ is necessarily open with respect to $\mathcal{T}'$. In this sense, the topology $\mathcal{T}$ is *the smallest topology* in which the given functions in $\mathcal{F}$ are continuous. ∎

• **3.4.24 Exercise.** Let X and Y be as in Example 3.4.18, and let $\mathcal{F}$ consist of the two functions $f_{(1,0)}, f_{(0,1)}$ ($(1,0)$ and $(0,1)$ are two choices of $\mathbf{a} \in \mathbf{R}^2$). Show that the topology defined on X is still the usual topology on $\mathbf{R}^2$. Show also that the collection $f^{-1}(V)$, $f \in \mathcal{F}, V \in \mathcal{S}$ introduced in (3.4.19 b) may indeed only be a subbase: in the situation at hand, it is neither a topology nor a base for a topology. ∎

3.4.25 SUPPLEMENT: Product topologies

Variations on the theme of (3.4.19 b) are possible. For example, the domains of the functions $f \in \mathcal{F}$ need not lie in a single given topological space. We give one more illustration, to be used as another opportunity to practice thinking about topologies.

3.4.26 Definition. *Let $(X_1, \mathcal{T}_1)$ and $(X_2, \mathcal{T}_2)$ be topological spaces. Let $X_1 \times X_2$ be the Cartesian product, $\{(x_1, x_2) \mid x_1 \in X_1, x_2 \in X_2\}$. Let π_1, π_2 be the functions that project onto the two "coordinates:" $\pi_1((x_1, x_2)) = x_1$, $\pi_2((x_1, x_2)) = x_2$. Define $\mathcal{B}^0$ to be the collection of subsets $\pi_i^{-1}(V_i)$ of $X_1 \times X_2$, where $i = 1, 2$ and V_i ranges over the open subsets of X_i. This collection is a subbase for a topology, called the* ***product topology*** *on $X_1 \times X_2$.* ∎

Let Λ be an arbitrary set (the *index set*). For each $\lambda \in \Lambda$, let there be given a topological space $(X_\lambda, \mathcal{T}_\lambda)$. Let $\prod X_\lambda$ be the collection of all sets $\{x_\lambda\}$ indexed by Λ, such that $x_\lambda \in X_\lambda$ for each $\lambda \in \Lambda$. When $\Lambda = \{1, \ldots, n\}$, this is just the usual Cartesian product. For general index sets, the existence of $\prod X_\lambda$ becomes a delicate issue of set theory, and requires a tricky axiom called the *Axiom of Choice*.

One now defines the projection functions $\pi_\lambda : \prod X_\lambda \to X_\lambda$ in the obvious way; they pick out the λ^{th} "coordinate." They again induce a product topology on $\prod X_\lambda$.

When the index set is $\mathbf{R}$, and each X_λ is itself $\mathbf{R}$, the product space $\prod X_\lambda$ may be thought of as $\mathbf{R}^{\mathbf{R}}$; it is the set of all functions from $\mathbf{R}$ to $\mathbf{R}$. This is a pretty ugly object, with no obvious norm or metric, but it can at least be equipped with a (product) topology in what turns out to be a fairly natural and useful way.

• **3.4.27 Exercise.** Let X_1, X_2 be $\mathbf{R}$ with the usual topology. Sketch $X_1 \times X_2$, sketch the sets $\pi_i^{-1}(V_i)$ for V_i open in X_i, $i = 1, 2$. Notice that the product topology is just the usual topology on $\mathbf{R}^2$. ∎

3.4.28 Exercise. Show that the collection of subsets of $X_1 \times X_2$ of the form $U_1 \times U_2$, where U_i is open in X_i for $i = 1, 2$, is a base for the product topology on $X_1 \times X_2$. ∎

3.4.29 Exercise. Extend the concept of product topology to Cartesian products $X_1 \times \ldots \times X_n$ of a finite number of topological spaces $(X_i, \mathcal{T}_i)$. Explain how one recovers the usual topology on $\mathbf{R}^n$ as a product topology. ∎

Miscellaneous Exercises

• **3.4.30 Exercise.** Is the set of all balls $B(\mathbf{x}, \epsilon) = \{\mathbf{y} \mid \|\mathbf{y} - \mathbf{x}\|_2 < \epsilon\}$, as $\mathbf{x}$ varies over $l^2(\mathbf{R}, \mathbf{N})$ and ϵ ranges over all positive reals, a base of $l^2(\mathbf{R}, \mathbf{N})$? If not, is it a subbase? ∎

• **3.4.31 Exercise.** Let $\mathcal{B}^0$ be the collection of all subsets of $\mathbf{R}$ of the form $\mathbf{R} - \{x\}$, for some $x \in \mathbf{R}$. Show that $\mathcal{B}^0$ is a subbase for a topology on $\mathbf{R}$. Describe this topology as explicitly as possible. ∎

• **3.4.32 Exercise.** Let X be the set of spheres with center at the origin $\mathbf{0}$ of $\mathbf{R}^3$. There is the obvious projection $p : \mathbf{R}^3 \to X$ that maps $\mathbf{x} \in \mathbf{R}^3$ to the sphere on which $\mathbf{x}$ lies. Describe the resulting topology on X. Is it Hausdorff? ∎

3.4.33 Exercise*. Let Ω be the set of all continuous real-valued functions ω on $[0, \infty)$ that satisfy $\omega(0) = 0$. Let $\mathcal{F}$ be the set of functions

$$\mathcal{F} = \{f : \Omega \to \mathbf{R} \mid f(\omega) = \omega(t_0) \text{ for some } t_0 \in [0, \infty)\}.$$

Show that the sets $U(a, b, t_0)$ defined by

$$U(a, b, t_0) = \{\omega \mid \omega(t_0) \in (a, b)\}$$

are a subbase for the topology on Ω induced by the family $\mathcal{F}$ according to Proposition 3.4.19 b). ∎

Subsection 3.5 Vocabulary of Topology

We have now met: *topological spaces, topologies, Hausdorff topologies, bases.* There are many other standard notions which every analyst will either use, or perhaps encounter and dismiss. Much of the emphasis is on developing *a language* that will streamline arguments involving topology.

3.5.1 Definition. *Let $(X, \mathcal{T})$ be a topological space, and let $x \in X$. An open set U containing x is called a **neighborhood of** x.* ∎

The term "neighborhood" is supposed to convey not only "closeness", but also the idea that the point is "surrounded" by the set (that's the import of "open" [22]). Let us rewrite Definition 3.2.3 in the more customary language, using "neighborhood of x" instead of "open set containing x."

3.2.3 Definition. *Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces, and let $f : X \to Y$ be a function. One says that f is **continuous at** $x_0 \in X$ if to any neighborhood V of $f(x_0)$, there is a neighborhood U of x_0 such that $f(U) \subset V$.* ∎

This version says no more or less than the original one. It just says it in a more economical way. This whole subsection is devoted to the introduction of a vocabulary that turns out to be useful in discussions involving various standard situations.

Open sets are a fancy version of open intervals, as we have seen; the familiar *closed intervals* $[a, b] \subset \mathbf{R}$ also have a topological counterpart.

3.5.2 Definition. *A subset F of a topological space $(X, \mathcal{T})$ is said to be **closed** if its complement*

$$F^c \stackrel{\text{def}}{=} \{x \mid x \notin F\}$$

is open. ∎

---

[22] Some texts use "neighborhood" for every set C such that $x \in V \subset C$ for some open set V. It pays to check whether an author means neighborhoods to be open, as I do.

3.5.3 Proposition. *i) $\emptyset$ and X are closed; ii) if F, G are closed then so is $F \cup G$; iii) an arbitrary intersection of closed sets is closed.* ∎

Proof. I prove part ii); this should help you to figure out how to prove the other parts.

If F and G are closed, then F^c and G^c are open. By (3.0.4), $F^c \cap G^c$ is open. But deMorgan's law from set theory says that

$$F^c \cap G^c = (F \cup G)^c,$$

and so $(F \cup G)^c$ is open. Then $\left((F \cup G)^c\right)^c = F \cup G$ is closed. ∎

• **3.5.4 Exercise.** Prove parts i) and iii) of the Proposition. Part iii) will again require deMorgan's laws. ∎

3.5.5 Examples

A. In $\mathbb{R}$, $[a, b]$ is a closed set, being the complement of the open set $(-\infty, a) \cup (b, \infty)$. $[a, b)$ is neither open nor closed.

B. Arbitrary intersections of open sets are, in general, not open, see Remark 3.1.5. Likewise, arbitrary unions of closed sets are, in general, not closed. For example,

$$\bigcup_{n=1}^{\infty} [\frac{1}{n}, 1 - \frac{1}{n}] = (0, 1).$$

Only *finite* unions of closed sets are guaranteed to be closed.

C. Mathematicians often try to understand the implications of an axiom system by replacing it with a seemingly different, but in fact equivalent, set of axioms. In (3.0.1), we defined a topology on X as a collection of sets satisfying (3.0.2-5). We could equally well introduce the concept of a "cotopology" (not a standard term), a collection C of subsets of X, the *closed* subsets, defined by conditions i), ii), iii) in Proposition 3.5.3. The two concepts are equivalent: anything one can prove using open sets can be proved equally well with closed sets and deMorgan's laws, or vice versa. Logical hairsplitting is of course not the real point. It may be easier to understand a particular property if it is formulated in terms of open sets, while another result is more efficiently proved by use of closed sets. ∎

3.5.6 Definition. *Let A be a subset of $(X, \mathcal{T})$. Let $\mathcal{F}(A)$ be the set of all closed sets containing A. The intersection*

$$\overline{A} \stackrel{\text{def}}{=} \bigcap_{F \in \mathcal{F}(A)} F$$

contains A, and it is closed (since arbitrary intersections of closed sets are closed, see (3.5.3)). It is called the closure of A. ∎

3.5.7 Example. Definition 3.5.6 is not particularly easy to work with, since "all closed sets containing A" are hard to find. To get the idea, consider $A = (a, b)$. Then

$$[a - \frac{1}{n}, b + \frac{1}{n}], n = 1, 2, \ldots$$

are *some* of the closed sets containing A, and

$$\bigcap_{n=1}^{\infty} \left[a - \frac{1}{n}, b + \frac{1}{n}\right] = [a, b].$$

This is, it turns out, the closure of A. We did not intersect *all* the closed sets containing A, however, and the more sets we intersect, the smaller the intersection will be: all we can conclude at this stage is that $\overline{A} \subset [a, b]$. ∎

A more practical approach to closure is the following.

3.5.8 Definition. *Let A be a subset of the topological space $(X, \mathcal{T})$. A point $x \in X$ is called a point of closure of A if every open set U containing x has nonempty intersection with A:*

$$\text{if } x \in U \text{ and } U \text{ is open, then } U \cap A \neq \emptyset. \blacksquare$$

Note that every point of A is a point of closure of A. For, if $x \in A$ and U containing x is open, then $U \cap A$ is nonempty — it contains x itself, for example. The schematic picture in Figure 40 should be kept in mind; the set A is $\{1/n \mid n \geq 1\}$.

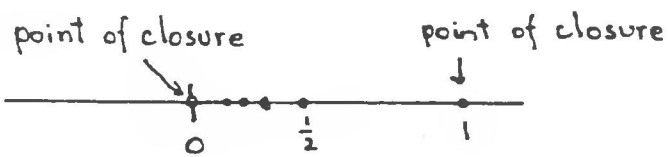
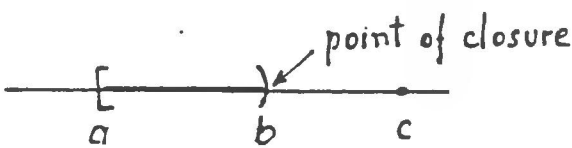


Figure 40. *Points of closure*

3.5.9 Example. Suppose that $c \notin [a, b]$. If $A = [a, b) \cup \{c\}$, the set of points of closure of A is $[a, b] \cup \{c\}$.



If you know about *accumulation points* (see below), note that in the figure above, c is a point of closure, but it is not an accumulation point. A point of closure of a set A may be isolated, meaning that a sufficiently "small" neighborhood of the point does not contain any other points of A. ∎

3.5.10 Proposition. *The set of points of closure of A is the closure of A as defined in (3.5.6).* ∎

Proof. We prove this in detail, as example of the style of proof common in general topology. Let

$$\tilde{A} = \{x \mid x \text{ is a point of closure of A}\},$$

and let $\overline{A}$ = closure of A. In order to show that $\tilde{A} = \overline{A}$, we prove:

$$i)\ \tilde{A} \subset \overline{A}$$
$$ii)\ \overline{A} \subset \tilde{A}.$$

(This is the standard approach to establishing equality of sets. See, e.g. the proof in (3.1.11).)

i) Let $x \in \tilde{A}$; we must show that it is in $\overline{A}$. Since $\overline{A} = \bigcap F$, the intersection of all closed sets F containing A, we must show that x is in every closed F containing A. (If x fails to be in just one such F, then x is not in the intersection). It is to be proved, then, that for any closed F containing A:

$$x \in \tilde{A} \quad \Rightarrow \quad x \in F.$$

Equivalently, one can prove the contrapositive,

(*) If there is one F containing A but not x, then $x \notin \tilde{A}$.[23]

I prefer to prove this contrapositive.

So, suppose that $A \subset F$ and $x \notin F$. Then $x \in F^c$. Since F is closed, F^c is open, and it contains x. But $F^c \cap A = \emptyset$, since $A \subset F$. Since there is an open set containing x which has empty intersection with A, x does not belong to $\tilde{A}$. This proves (*).

[To recapitulate:

$$\begin{aligned} x \in \tilde{A} &\Rightarrow x \in F \text{ whenever } A \subset F \\ &\Rightarrow x \in \bigcap_{\mathcal{F}(A)} F \\ &\Rightarrow x \in \overline{A}, \end{aligned}$$

so $\tilde{A} \subset \overline{A}$.]

ii) We next prove that $\overline{A} \subset \tilde{A}$. It must be shown that

$$x \in \overline{A} \Rightarrow x \in \tilde{A}.$$

---

[23] $p \Rightarrow q$ is equivalent to *not* $q \Rightarrow$ *not* p.

This, in turn, means:

$$x \in F \text{ whenever the closed set } F \text{ contains } A \;\Rightarrow\; x \in \tilde{A}.$$

Again, I prefer the contrapositive:

$$\text{not } \text{“}x \in \tilde{A}\text{”} \;\Rightarrow\; \text{not } \text{“}x \in \text{ every closed } F \text{ containing } A\text{”}$$

or

(**) $\qquad x \notin \tilde{A} \Rightarrow x$ does not belong to at least one F containing A.

[The negation of "blah happens for every" is "blah fails to happen for at least one"; the advantage of the contrapositive (**) is that one does not have to check something "for every F," which would be more difficult.]

To prove (**), observe that if $x \notin \tilde{A}$, then there is an open set U containing x but not meeting A: $U \cap A = \emptyset$ (see the definition of $\tilde{A}$ above.) Let F be the set U^c. This F is the complement of an open set, and is therefore closed, and it contains A (see Figure 41) — but $x \notin F$.

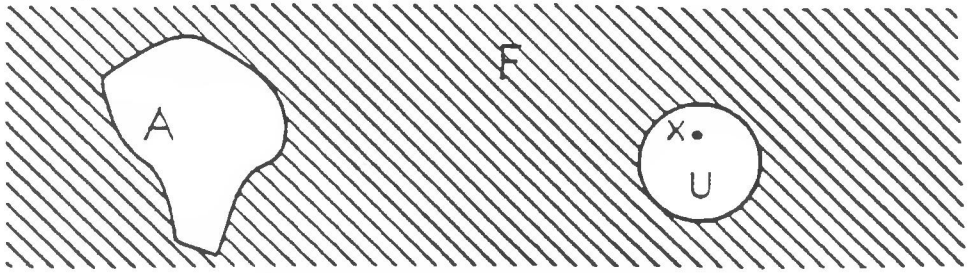


Figure 41. ***Picture for the proof of 3.5.10***

So, there is at least one closed set F containing A but not x. It follows that $x \notin \tilde{A}$.

[To recapitulate: $x \notin \tilde{A} \Rightarrow x \notin \overline{A}$, which is equivalent to $x \in \overline{A} \Rightarrow x \in \tilde{A}$ or $\overline{A} \subset \tilde{A}$.] ∎

A concept related to "point of closure" is "accumulation point."

3.5.11 Definition. *Let $(X, \mathcal{T})$ be a topological space. x is an* ***accumulation point*** *of A if for every open U containing x, $U \cap A$ contains points other than x.* ∎

The operation "closure" takes a set and throws in all boundary points. There is a kind of opposite, the operation "interior": it throws away all boundary points.

3.5.12 Definition. *Let A be any subset of $(X, \mathcal{T})$. A point $x \in A$ is called an* ***interior point*** *of A if there is an open U so that $x \in U \subset A$. Let $A^\circ = \{x \mid x$ is an interior point of $A\}$. The set A° is called the* ***interior*** *of A.* ∎

3.5.13 Example. The interior of $[a, b]$ is (a, b); the interior of $[a, b)$ is also (a, b). The set of rational numbers in $\mathbf{R}$ has no interior points. Every point of $B(x, \epsilon) = \{y \mid d(x, y) < \epsilon\}$ in a metric space (M, d) is an interior point of A. ∎

• **3.5.14 Exercise.** Find the interior of $\{(x,y) \in \mathbf{R}^2 \,|\, 0 < x \leq 1, 0 \leq y < 1\}$. The topology is the usual metric topology on $\mathbf{R}^2$. ∎

• **3.5.15 Exercise.** Refer to Exercise 3.0.7. What is the interior of $\{0,1\}$? ∎

• **3.5.16 Exercise.** Let $(X,\mathcal{T})$ be a topological space, and let $A, B \subset X$. Show that $A^\circ \cap B^\circ = (A \cap B)^\circ$. ∎

• **3.5.17 Exercise.** Let $f : M \to M'$ be a function from a metric space (M,d) to a metric space (M',d'). What is the *negation* of the following statement: For every $\epsilon > 0$, there is a $\delta > 0$, such that whenever $x \in M$, then $d(x,y) < \delta$ implies $d'(f(x),f(y)) < \epsilon$? ∎

• **3.5.18 Exercise.** A sequence $\{f_n\}$ of real-valued functions on a metric space (M,d) is said to *converge to the function* f if, for every $x \in M$ and $\epsilon > 0$ there is an integer $N_{\epsilon,x} > 0$ such that $n > N_{\epsilon,x}$ implies $|f_n(x) - f(x)| < \epsilon$. How is this condition *not* satisfied for a sequence $\{f_n\}$ that does *not* converge to f? ∎

• **3.5.19 Exercise.** Let A be the set of rational numbers in $[0,1]$. What is its closure? The intuitive answer is clear, but can you prove it? ∎

• **3.5.20 Exercise.** Refer to Exercise 3.0.7. What is the closure of the set $\{0,1\}$? ∎

• **3.5.21 Exercise.** Let (M,d) be a metric space with its metric topology, and let $\overline{A} = B(x,\epsilon)$. Show that if y is a point of closure of A then $d(x,y) \leq \epsilon$. Show by example that $d(x,y) = \epsilon$ need not imply that y is a point of closure of A. (Hint: discrete metric.) ∎

• **3.5.22 Exercise.** Show that a function $f : (X,\mathcal{T}) \longrightarrow (Y,\mathcal{S})$ is continuous if and only if for every closed subset C of Y, $f^{-1}(C)$ is a closed subset of X. ∎

• **3.5.23 Exercise.** Refer to Example 3.1.6. Show that the Cantor set is closed. ∎

3.5.24 Exercise*. Equip $\mathbf{R}$ with the Zariski topology. What are the closed sets in this topological space? Show that the closure of every nonempty open set U is the whole space, $\mathbf{R}$. ∎

3.5.25 Exercise. Refer to Exercise 3.3.5. In the orbit space X, what is the closure of the set of hyperbola-like orbits? ∎

• **3.5.26 Exercise.** Prove that every accumulation point of A is a point of closure of A. The converse is not true; see Example 3.5.9. ∎

• **3.5.27 Exercise.** Let $A = \{1, 1/2. 1/3, \ldots\} \subset \mathbf{R}$ (with the usual topology). Find the points of closure of A and the accumulation points of A. ∎

• **3.5.28 Exercise.** Let $(X,\mathcal{T})$ be a topological space, and let $A \subset X$ be closed. Show that every accumulation point of A is contained in A. (This is one of the most important properties of closed sets, probably familiar from undergraduate analysis.) ∎

• **3.5.29 Exercise.** Let $(X,\mathcal{T})$ be a topological space, and let $A \subset X$. Show that every point of $\overline{A} - A$ is an accumulation point of A. ∎

• **3.5.30 Exercise.** Let $(X, \mathcal{T})$ be a topological space, and let $A \subset X$. There are points that "divide" A from the rest of X.
Definition. *A point $x \in X$ is said to be a* ***boundary point of*** *A if $x \in \overline{A} \cap \overline{A^c}$.*
Prove that x is a boundary point of A if, and only if, $x \in (A^\circ \cup (A^c)^\circ)^c$. ∎

• **3.5.31 Exercise.** Using the definition in the preceding exercise, find the boundary points of the following subset of $\mathbf{R}^2$ (with the usual topology): $\{(x, y) \mid x^2 + y^2 < 1\} \cup \{(x, y) \mid x^2 + y^2 = 1 \text{ and } y > 0\}$. ∎

• **3.5.32 Exercise.** What are the boundary points of the ball $B(x, \epsilon)$ in a metric space (M, d)? ∎

3.5.33 Exercise. We have seen (in Subsection 3.2) that topological spaces can be classified according to separation properties. Another kind of classification looks at the "size" of the space. (This *is* in fact the same as Exercise 3.2.16)

Let $(X, \mathcal{T})$ be a topological space. It is said to be ***first countable*** if, for every $x \in X$, there is a countable set $\{U_n\}$ of neighborhoods of x with the following property: for every neighborhood U of x, there is a U_n such that $x \in U_n \subset U$. (This is not quite the same as the definition in Exercise 3.2.16, but it is equivalent.) Show that every metric space, with its metric topology, is first countable. ∎

3.5.34 Exercise. Is the space in Exercise 3.0.7 first countable? ∎

3.5.35 Exercise*. Equip $\mathbf{R}$ with the Zariski topology. Show that this topological space is *not* first countable. ∎

3.5.36 Exercise. Let $(X, \mathcal{T})$ be a topological space. It is said to be *second countable* if it has a countable base. Show that $\mathbf{R}$ with its usual metric is second countable. ∎

3.5.37 Exercise. Prove: a second countable topological space is automatically first countable. ∎

3.5.38 Exercise. A metric space (M, d) is said to be *separable* if there is a countable subset $\{x_n\} \subset M$ with the following property: given $x \in M$ and $\epsilon > 0$, there is an x_n such that $d(x, x_n) < \epsilon$. Prove that a metric space is second countable if, and only if, it is separable. ∎

3.5.39 Exercise. Show that $l^\infty(\mathbf{R}, \mathbf{N})$ is not second countable (but it is first countable, by Exercise 3.5.33.) ∎

Subsection 3.6 Application: Brownian Motion

In 1827, Robert Brown (1773-1858), a Scottish botanist, published a paper on the erratic motion of pollen particles suspended in liquid. This motion, now called *Brownian motion,* is caused by the impulses imparted to the pollen particle by the molecules making up the fluid in which it is suspended. The qualitative and quantitative explanation of Brown's observations[24] has concerned many mathematicians and physicists. For instance, one of Einstein's great papers *before* his work on the theory of relativity dealt with the computation of the physical parameters describing Brownian motion; apparently, he was unaware of Brown's observations, and arrived at his conclusions by purely theoretical considerations.

In these notes, we deal mostly with the mathematical setup. For us, Brownian motion is interesting because it illustrates concepts from topology, measure theory, and probability theory. A thorough study of Brownian motion would require *all* the ideas developed in this course. It is, in that sense, the perfect application, and the present subsection is only the first one to make use of this mathematical goldmine.[25]

This subsection treats a special case of Brownian motion. The "pollen" particle moves in one dimension only, and it starts its wanderings at a fixed reference point at time $t = 0$. Its distance from this reference point is therefore given by a function ω of time t, subject to the normalization $\omega(0) = 0$. It seems reasonable on physical grounds, and will be postulated for the mathematical theory, that the possible functions $\omega(\cdot)$ describing the motion of the diffusing particle should be *continuous* on $[0, \infty)$.

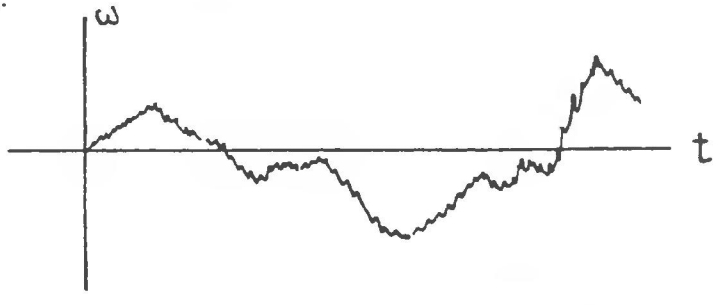


Figure 42. *A Brownian motion sample path*

In this introductory discussion, we define the space Ω of all the possible motions of the particle, and try to understand some topological properties of this space. The probabilistic aspect of the theory will be taken up later in this chapter.

---

[24] According to a story in *Science News,* **139**, p. 287, Brown probably did *not* observe Brownian motion! Oh well —

[25] For a general survey (much of which will be incomprehensible at this time), see Article 45 in EDM II.

3.6.1 Notation. *Let Ω be the set of continuous real-valued functions ω on $[0, \infty)$ satisfying $\omega(0) = 0$.* ∎

We wish to introduce a topology on Ω. It will be defined by a subbase $\mathcal{B}^0$ (see (3.4.9)). The basic open sets are

$$U(t_0, a, b) \stackrel{\text{def}}{=} \{\omega \mid \omega(t_0) \in (a, b)\},$$

where $t_0 > 0$ and $-\infty \leq a < b \leq \infty$.

• **3.6.2 Exercise.** Verify that the collection of all sets $U(t_0, a, b)$ satisfies the condition in Proposition 3.4.10, and is therefore a subbase for a unique topology on Ω. ∎

The set Ω is not at all like the normed linear spaces L^p we met in Section 2. The functions in Ω need not have finite integrals over $[0, \infty)$; indeed, one must allow for the possibility that the diffusing pollen particle moves arbitrarily far from its starting point, which means that $\omega(t)$ for $\omega \in \Omega$ can become arbitrarily large as $t \to \infty$. There is no natural measure of size of such an ω, and there is no natural measure of distance between two different possible ω's. The choice of topology is determined by the following considerations. We should be able to measure the distance of the diffusing particle from the starting point, at any time t_0; this distance, for a motion ω, is $\omega(t_0)$. Two different possible motions should be "close" if (but *not* only if) they are close at a particular time t_0. Hence, we follow the strategy of Example 3.4.18 and Proposition 3.4.19 b), and ask that the so-called *evaluation functions* $f_{t_0} : \Omega \to \mathbf{R}$, defined by

$$f_{t_0}(\omega) = \omega(t_0),$$

be continuous. This insures that our topology will consider ω_1 and ω_2 to be "close" if the values $\omega_1(t_0)$ and $\omega_2(t_0)$ are close. (We put quotation marks around *close* when the adjective is applied to elements of Ω, because, as we now know, topologies measure "distance" in a peculiar and non-quantitative way.)

One should understand precisely how the sets $U(t_0, a, b)$ fit into the framework of Example 3.4.18 and Proposition 3.4.19 b). Refer to Figure 38. The function f_{t_0} has domain Ω and range $\mathbf{R}$. The interval (a, b) is open in $\mathbf{R}$, and we want its pre-image to be open. That pre-image is the set $\{\omega \mid f_{t_0}(\omega) \in (a, b)\}$, which is just $U(t_0, a, b)$.

The collection $\mathcal{B}^0$ of sets $U(t_0, a, b)$ is only a *subbase* for the topology on Ω. According to Proposition 3.4.10, a *base* for the topology is provided by the collection of all finite intersections of sets from $\mathcal{B}^0$. The typical such set has the form

$$U(t_1, a_1, b_1) \cap \ldots \cap U(t_n, a_n, b_n) = \{\omega \mid \omega(t_1) \in (a_1, b_1), \ldots, \omega(t_n) \in (a_n, b_n)\}. \tag{3.6.3}$$

The integer n is arbitrary (≥ 1) and $-\infty \leq a_i < b_i \leq \infty$ for $i = 1, \ldots, n$. One pictures such sets as collections of continuous functions that pass through prescribed "windows" (a_i, b_i) at prescribed times t_i.

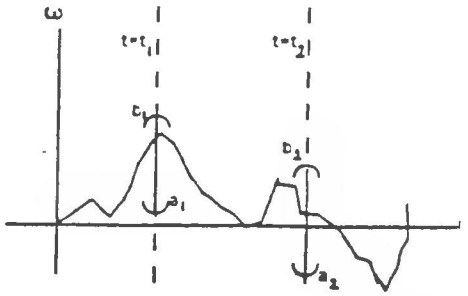


Figure 43. *Pollen passing through windows*

We denote the base of the topology on Ω by B; this is the collection of all sets of the form (3.6.3). We will call a set of this form a ***window set.***

From now on, we refer to a function ω, representing a possible motion of the diffusing particle, as a *sample path*, or more simply as a *path*. In the analysis of Brownian motion, one is interested in various subsets of the set Ω of all possible paths, such as: the set of paths that do not return to the origin (=starting point) before some time T, the set of paths for which ω is differentiable, and so forth. We will look at a few examples of this type, partly in exercises.

Brownian motion theory is only peripherally concerned with topology; that is just a means to an end. Later in this chapter, we will introduce the *probability* that a randomly chosen sample path belongs to $U(t_0, a, b)$, i.e., passes between a and b at time t_0. This probability will be

$$\frac{1}{\sqrt{2\pi t_0}} \int_a^b e^{-x^2/2t_0}\, dx.$$

The formula is motivated by physical arguments about diffusion, or by mathematical arguments involving random walks and coin tosses. At any rate, a probability is then assigned to sets (3.6.3) belonging to the basis of the topology $\mathcal{T}$ on Ω, and is extended, by appeal to a general theory, to the complicated sets described below. One then deduces a number of surprising results: the typical path ω is continuous but nowhere differentiable; the typical path ω has a Cantor set of zeros, and so on. The presentation of these theorems is perhaps the mathematical highlight of the chapter. For now, however, we concentrate on the more mundane topological aspects of the path-space Ω.

3.6.4 Example. Let A_{t_0} be the set $\{\omega \mid \omega(t_0) \neq 0\}$. Is A_t open? The answer is *yes*. This set is just $U(t_0, 0, \infty) \cup U(t_0, -\infty, 0)$, which is the union of two basic open sets and is therefore open. ∎

3.6.5 Example. Let B_T be the set $\{\omega \mid \omega(t) \neq 0 \text{ for } 0 < t < T\}$. Is B_T open? Note that B_T is the intersection, over $0 < t < T$, of the sets A_t from the preceding

example. This is an intersection of uncountably many open sets, however, and is therefore not necessarily open. In fact, B_T is *not* open. One can see this as follows.

Suppose B_T were open, and let $\omega_0 \in B_T$. Then by Proposition 3.4.2, in particular by property (3.4.5), there must be an open set V belonging to the base of the topology on Ω such that $\omega \in V \subset B_T$. Such a V is a window set (3.6.3). It is clear, however, that this V will not be contained in B_T: if $\omega \in V$, then ω is constrained only at certain points $t_1, \ldots, t_n$. Even if ω is not zero at these distinguished times t_i, it can, of course, be zero at some other $t \in (0, T)$; Figure 44 shows how to construct such an ω. Therefore, condition (3.4.5) cannot hold, and B_T is not open. ∎

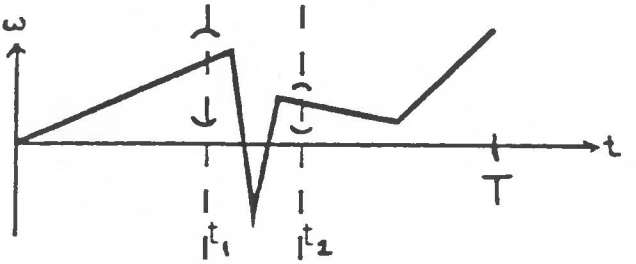


Figure 44. *A path that returns to the origin before time T*

3.6.6 Example. Let C_T be the set $\{\omega \mid \omega(t) = 0 \text{ for some } t \in (0, T)\}$. Is C_T open, closed, or neither?

It is neither open nor closed. If it were closed, its complement would be open (by definition of "closed"), but that complement is the set B_T from the preceding example which, as we have just seen, is not open. The set C_T is not open, either — this can be seen by an argument similar to the one in Example 3.6.5. ∎

• **3.6.7 Exercise.** Prove that the set C_T in the preceding example is not open. ∎

• **3.6.8 Exercise.** The set B_T in Example 3.6.5 is not open. Does it have any interior points (see (3.5.12))? ∎

• **3.6.9 Exercise.** Is the set B_T in (3.6.5) closed? ∎

It is obviously not easy to invent interesting open or closed subsets of Ω. One therefore introduces two more general classes of sets, both of great importance in measure theory (to be described later in this chapter.)

3.6.10 Definition. *Let (X, T) be a topological space. A G_δ set is a countable intersection of open sets, and an F_σ set is a countable union of closed sets. A countable intersection of F_σ sets is denoted by $F_{\sigma\delta}$, and a countable union of G_δ sets is denoted by $G_{\delta\sigma}$. (Read from the inside out: $G_{\delta\sigma} = (G_\delta)_\sigma$, etc., where "$\delta$" stands for intersection and "σ" stands for union, as explained in the next remark.) Likewise, $G_{\delta\sigma\delta}$ is a countable intersection of a countable union of G_δ sets, and so forth.* ∎

Remarks

A. Remember that *finite* intersections of open sets are open, and *finite* unions of closed sets are closed. So, a G_δ set is usually not open (see (3.1.5)), and an F_σ set is usually not closed (see (3.5.5.B)). The (standard) notation shows that mathematics knows no boundaries of language. In German, *Gebiet* is one of the terms that connotes *open set*; the "G" in G_δ is the first letter of Gebiet. The Greek δ represents the first letter of *Durchschnitt*, German for *intersection*. *Closed*, in French, is *fermé* — hence the "F"; the Greek σ is the first letter of *Summe*, or *sum*, an older term for *union*. Aren't you glad you asked?

B. Sets of the type G_δ, $G_{\delta\sigma}$, and so on, come up in the study of Brownian motion because they allow one to build complicated sets by performing countably many operations on simple sets. This is essential once probabilities are included in the setup. ∎

• **3.6.11 Exercise.** Prove: the complement of a G_δ set is an F_σ set, and vice versa. ∎

3.6.12 Example. The set B_T in (3.6.5) is a $G_{\delta\sigma\delta}$ set.

This is a subtle and rather long construction; first I try for a verbal description, without messy formulas. Let $\omega \in B_T$, so that $\omega(t) \neq 0$ for $t \in (0,T)$. In particular, $\omega(t) \neq 0$ for $1/m \leq t \leq T - 1/m$, whenever m is an integer and is large enough so that $[1/m, T-1/m] \subset (0,T)$. At this stage we once again appeal to a basic theorem about continuous functions (which will be proved in II.1.3.8):

A continuous function defined on a closed, bounded interval $[a,b]$ assumes its maximum value and its minimum value on the interval $[a,b]$.

Since the continuous function $|\omega|$ is not zero on $[1/m, T-1/m]$, its minimum value must be strictly greater than zero: $|\omega(t)| > 1/n$ for some integer n.

The last step is to realize that if $|\omega(t)| > 1/n$ for all $t \in [1/m, T-1/m]$, then in particular $|\omega(k/N)| > 1/n$ for every rational number $k/N \in [1/m, T-1/m]$. All these integers m, n, k, N are needed in order to reduce the *uncountable* number of conditions "$|\omega(t)| > 0$ for $t \in (0,T)$" to a *countable* number of conditions. We summarize the discussion so far in a figure. For sake of definiteness, we draw the case $\omega(t) > 0$ for $t \in (0,T)$ (the other possibility would be $\omega(t) < 0$.)

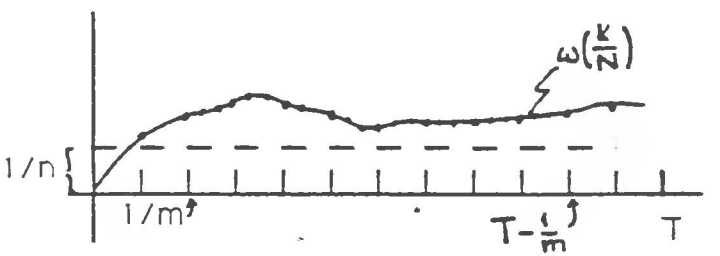


Figure 45. *Illustrating Example 3.6.12*

Now we write a formula that captures Figure 45 and all the words that precede it:

$$B_T = \bigcap_{m=m_0}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{N=N_0}^{\infty} \{\omega \,|\, |\omega(k/N)| > 1/n \text{ for all rational } k/N \in [1/m, T-1/m]\}. \tag{3.6.13}$$

The starting index m_0 in (3.6.13) is chosen so that $[1/m, T-1/m] \subset (0,T)$ for $m \geq m_0$; its value depends on the size of T. The starting index N_0 is chosen so that the intervals $[1/m, T-1/m]$ (for $m \geq m_0$) contain rational points of the form k/N; N_0 again depends on the size of T. (For example, if $T = 1$, one should take $m_0 = 3$ and $N_0 = 2$.)

We now unravel (3.6.13), and show (as usual) that each side is a subset of the other side. One half of this argument has essentially been done in the discussion at the start of this example. We repeat it, to illustrate the technique.

Suppose $\omega \in B_T$. We need to show that ω belongs to the $\cap\cup\cap$ on the right. One works through $\cap$, $\cup$, $\cap$ from left to right (outside to inside.) Recall that $\cap$ can be read "for every index" and $\cup$ can be read "there is an index." Thus, ω will lie in the set on the right side, if

> for every $m \geq m_0$, there is an $n \geq 1$ (n will depend on the m chosen) such that $|\omega(k/N)| > 1/n$ for all $N \geq N_0$ and all k for which $k/N \in [1/m, T-1/m]$.

This is the content of the figure above. On every subinterval $[1/m, T-1/m]$, the function $|\omega|$ is bounded away from 0 by some $1/n$, and in particular, this holds true at the rational points k/N.

Now suppose that ω is in the right side of (3.6.13). We must show that $\omega(t_0) = 0$ cannot happen for any $t_0 \in (0,T)$. Suppose it did happen. Then t_0 lies in some interval $[1/m, T-1/m]$. There is an n such that $|\omega(k/N)| > 1/n$ for all rationals k/N in $[1/m, T-1/m]$. This, however, is a contradiction: by continuity, there is a δ such that

$$|\omega(t)| = |\omega(t) - \omega(t_0)| < \frac{1}{2n}$$

whenever $|t - t_0| < \delta$. Since there must be a rational k/N satisfying $|k/N - t_0| < \delta$, we could not have $|\omega(k/N)| > 1/n$ for *all* rationals k/N. Hence, ω cannot vanish on $(0,T)$.

We have shown that B_T is a countable intersection of countable unions of countable intersections of sets of the form

$$X_{mnN} \stackrel{\text{def}}{=} \{\omega \,|\, |\omega(k/N)| > 1/n \text{ for all rational } k/N \in [1/m, T-1/m]\}.$$

To argue that B_T is a $G_{\delta\sigma\delta}$ set, we must prove that X_{mnN} is open. This, at least, is easy. Fix m, n and N. There are only finitely many rational numbers k/N in $[1/m, T-1/m]$. Call them $t_1, \ldots, t_r$. Then rewrite the definition of X_{mnN}:

$$X_{mnN} = \{\omega \,|\, \omega(t_j) \in (-\infty, -1/n) \cup (1/n, \infty),\ j = 1, \ldots, r\}.$$

This new description exhibits X_{mnN} as intersection of two basis sets of the form (3.6.3). Hence, X_{mnN} is open, and the assertion of the example is (at last) completely proved.

Examples of this type are not, strictly speaking, the concern of topology, which worries about open sets and closed sets — its axioms do not extend beyond such simple sets. The example really straddles the boundary between topology and measure theory (as will be discussed in Section 5); when one tries to do integration and probability on topological spaces, one must admit sets got from open ones by arbitrary countable operations. ∎

• **3.6.14 Exercise.** What kind of set is C_T in (3.6.6): G_δ, F_σ, $G_{\delta\sigma\delta}$, etc.? ∎

3.6.15 Exercise. Let $D_{T,\alpha}$ be the set $\{\omega \mid |\omega(t)| < \alpha \text{ for } t \in (0,T)\}$. What kind of set is $D_{T,\alpha}$: open, closed, G_δ, F_σ, etc.? ∎

3.6.16 Exercise. Work the last exercise with $|\omega(t)| \leq \alpha$ instead of $|\omega(t)| < \alpha$. ∎

3.6.17 Exercise. Let E_T be the set $\{\omega \mid \omega(t) \neq 0 \text{ for } t \in (0,T) \text{ and } \omega(T) = 0\}$. (In words: the first return to the starting point happens at time T.) What kind of set is E_T: open, closed, G_δ, F_σ, $G_{\delta\sigma\delta}$, $F_{\sigma\delta\sigma}$, etc.? ∎

3.6.18 Exercise*. The true Brownian motion takes place in three dimensions; one may, however, assume, with some justification, that the motion of a pollen particle in a flat drop of liquid is confined to two dimensions. This leads to a new, more realistic and more complicated, setup. Let $\vec{\Omega}$ be the set of continuous functions $\vec{\omega}(\cdot) = (x_1(\cdot), x_2(\cdot))$ with domain $[0,\infty)$ and range in $\mathbf{R}^2$, such that $\vec{\omega}(0) = (0,0)$. A function $\vec{\omega}$ represents the path of the diffusing particle in the (x_1, x_2) plane, and $\vec{\omega}(0) = \vec{0}$ fixes the starting point at the origin.

a) Define the appropriate topology on $\vec{\Omega}$. Hint: let $\pi_{t_0,\imath}$ be the projection function taking $\vec{\omega}$ to $x_\imath(t_0)$ (for $\imath = 1, 2$.) For $U \subset \mathbf{R}$ open, the inverse image $\pi^{-1}_{t_0,\imath}$ is declared to be open in $\vec{\Omega}$.

b) Is the set of $\vec{\omega}$ for which $\vec{\omega}(t)$ leaves the unit disk $x_1^2 + x_2^2 \leq 1$ before a fixed time $T > 0$ open, closed, G_δ, F_σ, $G_{\delta\sigma}$, etc. etc.?

c) What kind of set is the collection of $\vec{\omega}$ for which $\vec{\omega}(t) = \vec{0}$ for some $t \leq T$? (I.e., the particle returns to the origin before time T.) ∎

Subsection 3.7 Application: Weak Topology

A "weak topology" is a topology on a linear space that is defined by a set of linear functions. Example 3.4.18 provides a simple illustration, but it is untypical because the resulting topology is in fact identical with the standard one derived from the norm $||\cdot||_2$. In infinite dimensions, weak topologies give new structures on function spaces, and these have some interesting and useful concrete interpretations. It would take too long to describe all the details, but since the construction fits so naturally into this section, a short summary and preview is perhaps appropriate.

The term "weak," in connection with topology, has two meanings: a generic one and a limited one. The generic use is more basic.

3.7.1 Definition. *Let X be a set, and let $\mathcal{T}$, $\mathcal{S}$ be two topologies on X. If $\mathcal{T} \subset \mathcal{S}$, one says that $\mathcal{T}$ is weaker than (or coarser than) $\mathcal{S}$, and that $\mathcal{S}$ is stronger than (or finer than) $\mathcal{T}$.* ∎

I will use the terms "coarser", respectively "finer", to indicate that one topology on a set X is a subset of, resp. contains, another. The words "weak" and "strong" are reserved for the special topologies described in this subsection.

The coarser of two topologies has fewer open sets, and the finer one has more. For instance, Exercise 3.3.17 asserts that the Zariski topology on $\mathbf{R}$ is coarser than the metric topology on $\mathbf{R}$.

We will now consider the normed linear spaces $l^2(\mathbf{R}, \mathbf{N})$ and L^2_0. They already have one topology, given by the metric associated to the norm. The open balls $B(x, \epsilon)$ form a basis for this topology. We will define a coarser topology: it is called *the weak topology on* $l^2(\mathbf{R}, \mathbf{N})$, respectively L^2_0. The metric or norm topology is also called *the strong topology*. All the other normed linear spaces introduced in Section 2 also have weak topologies, but those will not be discussed.

3.7.2 Notation. *For $\mathbf{x} = (x_0, x_1, \ldots)$, $\mathbf{y} = (y_0, y_1, \ldots) \in l^2(\mathbf{R}, \mathbf{N})$, set*

$$(\mathbf{x}, \mathbf{y}) = \sum_{n=0}^{\infty} x_n y_n. \tag{3.7.3}$$

For $\mathbf{a} = (a_0, a_1, \ldots) \in l^2(\mathbf{R}, \mathbf{N})$, let $f_{\mathbf{a}}$ be the (linear) function defined by

$$f_{\mathbf{a}}(x) = (\mathbf{x}, \mathbf{a}) = \sum_{n=0}^{\infty} x_n a_n. \quad ∎$$

The notation $(\mathbf{x}, \mathbf{y})$ is standard; the expression in (3.7.3) is called the *inner product* of x and y, and later on it will play an important role in our study of the space $l^2(\mathbf{R}, \mathbf{N})$. ∎

3.7.4 Definition. *The weak topology $\mathcal{W}$ on $l^2(\mathbf{R}, \mathbf{N})$ is the topology induced by the family $\mathcal{F}$ of functions as in Proposition 3.4.19 b).* ∎

Remark. Recall from Proposition 3.4.19 b) and Exercise 3.4.23 that the topology $\mathcal{W}$ defined by the collection $\mathcal{F}$ of linear functions is the smallest topology for which all the sets $f_{\mathbf{a}}^{-1}(a, b)$ are open, where (a, b) is an arbitrary open interval in $\mathbf{R}$. One then has $f_{\mathbf{a}}^{-1}(\text{open set}) = $ open set, and all linear functions are continuous. Many more functions would be continuous with respect to the strong (metric) topology on $l^2(\mathbf{R}, \mathbf{N})$; for example, the function $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$ is continuous in the metric topology, but not in the weak topology. ∎

In Example 3.4.18, the construction in Definition 3.7.4 was described for $\mathbf{R}^2$. Since one cannot really visualize the situation in $l^2(\mathbf{R}, \mathbf{N})$, it is essential to

understand the simpler setup in $\mathbf{R}^2$. Take a function $f_{\mathbf{a}}(x) = a_1x_1 + a_2x_2$. For $V = (\alpha, \beta)$ open in R, the set

$$f_{\mathbf{a}}^{-1}((\alpha, \beta)) = \{(x_1, x_2) \,|\, \alpha < a_1x_1 + a_2x_2 < \beta\}$$

is, by definition, open in the topology induced by the collection $\mathcal{F}$ of (linear) functions. It is an unbounded open strip, drawn in Figure 46.

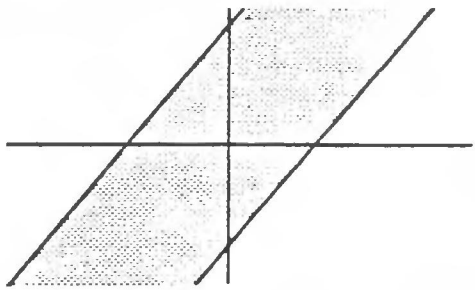


Figure 46. $\alpha < a_1x_1 + a_2x_2 < \beta$

The intersection of two suitable such strips will be a square, and it can be made as small as desired.

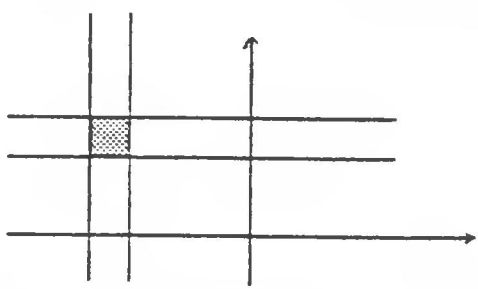


Figure 47. *A teeny intersection of two open strips*

We know that the balls $B(x, \epsilon)$ of the l^∞ metric on $\mathbf{R}^2$ are squares. So, the topology generated by the family $\mathcal{F}$ is just the metric topology of the l^∞ metric, but we know ((3.1.12)) that all the metric topologies on $\mathbf{R}^2$ are equivalent. Therefore, the weak topology, in this case, gives nothing new.

If this construction is generalized to $\mathbf{R}^3$, the strips are replaced by the "slices" between two parallel planes. One must intersect three of them to produce a cube; but this is still open, and the weak topology is again the metric topology on $\mathbf{R}^3$. By the time one gets to $l^2(\mathbf{R}, \mathbf{N})$, however, one would have to intersect infinitely many open "slices" of the form

(3.7.5) $$\alpha < a_1x_1 + a_2x_2 + \ldots < \beta$$

to produce a cube (with infinitely many sides, now). The intersection of infinitely many open sets is usually not open, and there is no reason to expect the topology defined in (3.7.5) to be the same as the metric topology. In fact, it is *not* the same any more.

• **3.7.6 Exercise.** A base for the metric topology on $l^2(\mathbf{R}, \mathbf{N})$ is given by the open balls $B(\mathbf{x}, \epsilon) = \{\mathbf{y} \mid ||\mathbf{x} - \mathbf{y}||_2 < \epsilon\}$. Let U be a finite intersection of slices of the form (3.7.5), and let $\mathbf{x} \in U$. There is a ball $B(\mathbf{x}, \epsilon)$ that is contained in U. Conversely, however, a ball $B(\mathbf{x}, \epsilon)$ does not contain a finite intersection of slices (3.7.5). Hence, every set that is open in the weak topology on $l^2(\mathbf{R}, \mathbf{N})$ is also open in the metric toplogy, but the converse is not true. Prove these assertions in detail. ∎

What is the point of "weak topology"? The weak topology has fewer open sets than the strong topology. Therefore, any property that requires something to hold "for every open set" will be more easily satisfied in the weak topology — there are fewer open sets to check. A property that requires one to find "an open set such that" will be more easily satisfied in the strong topology — there are more open sets to choose from. To illustrate these points, we will study some examples of sequences in topological spaces. A sequence $\mathbf{x}^{(n)}$ converges to $\mathbf{x}$ if *every* open set containing x eventually contains all elements of the sequence. It is easier, therefore, for a sequence to converge if the topology has fewer open sets. The important question, of course, is "why would anyone want to make it easier for sequences to converge?" Answer: if you try to solve an equation by iteration, getting approximate solutions $\mathbf{x}^{(n)}$ that, you hope, converge to a solution $\mathbf{x}$, you will have more chance of success when you make it easier for sequences to converge. Of course, when sequences need not do much to converge, they may well converge to the wrong limit. This sounds like a mathematical confidence game, but the ends justify the means: often (e.g., in the theory of elliptic partial differential equations), one first establishes existence of a *weak solution*, and then proves that *weak = strong*, i.e. that the weak solution (got by a weak convergence argument) is actually the solution one expected to find.

Weak convergence also appears in the following setting. To solve some nasty partial differential equations, one sometimes adds ϵ times a term that makes the nasty equation nice. The equation

$$u_t(t, x) + u(t, x)u_x(t, x) = 0$$

is "nasty"; its solutions may not be continuous, and, worse, may not be unique. The solutions of

(3.7.7) $$u_t + uu_x = \epsilon u_{xx}$$

or

(3.7.8) $$u_t + uu_x + \epsilon u_{xxx} = 0$$

are most pleasant: differentiable and unique. One now lets $\epsilon \to 0$, and hopes that the solutions of the nice equations tend to an acceptable solution of the nasty equation. For Nice Equation 3.7.7, this is true, in a kind of strong-convergence sense. For Nice Equation 3.7.8, matters are much more complicated; the limit as $\epsilon \to 0$ exists only in the sense of weak convergence, and it is in fact not true that this weak limit solves $u_t + uu_x = 0$ (see Exercise 3.7.32). This is not just a matter of concern for stupid theoreticians who happened to add the wrong ϵ term to $u_t + uu_x = 0$. Schemes for the numerical solution of $u_t + uu_x = 0$ (and related equations) often act as if they are trying to approximate the equation (3.7.8), and their convergence, as the step-size of the discretization tends to 0, is weak convergence to the weak limit of solutions of (3.7.8). See Figure 54 below.

Definition 3.2.10 of convergence of a sequence is repeated for the reader's convenience.

3.7.9 Definition. *Let (X, T) be a topological space, and let $\{x_n \mid n \geq 1\}$ be a subset of X. We call such a set a* ***sequence in*** *X. The sequence is said to* ***converge to*** *$x \in X$ if for every neighborhood V of x there is an integer N_V (usually depending on V), such that $x_n \in V$ for every index $n \geq N_U$.* ∎

This definition requires something to happen "for every open set V containing x," so, as pointed out already, it is more likely to be satisfied in a topological space with fewer open sets. We will see examples below.

3.7.10 Definition. *A sequence $\{x^{(n)}\}$ in $l^2(\mathbf{R}, \mathbf{N})$* ***converges weakly*** *to x if it converges in the weak topology on $l^2(\mathbf{R}, \mathbf{N})$. A sequence $x^{(n)}$* ***converges strongly*** *to x if it converges in the metric topology on $l^2(\mathbf{R}, \mathbf{N})$. We use the notation $x^{(n)} \rightharpoonup x$ to indicate that $x^{(n)}$ converges* ***weakly*** *to x, and the notation $x^{(n)} \longrightarrow x$ to indicate that $x^{(n)}$ converges* ***strongly*** *to x.* ∎

The logical next step would be to prove a series of lemmas and propositions, which would then be illustrated by a bunch of examples. It seems better, however, to state the results, work out the examples, and only then, after the reader has (perhaps) become convinced of the value of the concepts, run through the proofs.

The next proposition provides the most important, *computable*, criterion for weak convergence of a sequence in $l^2(\mathbf{R}, \mathbf{N})$. One should understand it before wasting too much time on the official, but uninformative, Definition 3.7.10.

3.7.11 Proposition. *A sequence $\mathbf{x}^{(n)}$ in $l^2(\mathbf{R}, \mathbf{N})$ converges weakly to $\mathbf{x}$ if, and only if, the sequence of inner products $(\mathbf{x}^{(n)} - \mathbf{x}, \mathbf{a})$ converges to 0 in $\mathbf{R}$, for every $\mathbf{a} \in l^2(\mathbf{R}, \mathbf{N})$.* ∎

Remark. The weak topology is designed to make all the linear functions $f_{\mathbf{a}}$ continuous. For a continuous function $f_{\mathbf{a}}$, we expect that when $\mathbf{x}^{(n)}$ approaches

x, then $f_a(x^{(n)})$ approaches $f_a(x)$. That idea is turned upside-down in this proposition. When the numbers $f_a(x^{(n)})$ approach the number $f_a(x)$ for every linear function, we *declare* that the vectors $x^{(n)}$ converge to the vector x. This is a new and, as the name suggests, weaker (=less restrictive) notion of convergence. ∎

The concept of *weak topology*, and the criterion (3.7.11), generalize to function spaces. We take the interval of integration to be $[0, 1]$ for sake of definiteness, but it could be any interval whatsoever.

3.7.12 Definition. *For $a \in L_0^2([0,1])$, let F_a be the linear function from L_0^2 to $\mathbb{R}$ defined by*

$$F_a(f) \stackrel{\text{def}}{=} \int_0^1 f(x)a(x)\,dx.$$

The weak topology on L^2 is the topology induced, according to Proposition 3.4.19 b), on L_0^2 by the family of function $\mathcal{F} = \{F_a \mid a \in L_0^2\}$. ∎

3.7.13 Proposition. *A sequence f_n in $L_0^2([0,1])$ converges weakly to f if, and only if,*

$$\lim_{n\to\infty} \int_0^1 (f_n(x) - f(x))a(x)\,dx = 0$$

for all $a \in L_0^2([0,1])$. ∎

3.7.14 Example. Let $e^{(n)}$ be the vector $(0, \ldots, 0, 1, 0, \ldots) \in l^2(\mathbf{R}, \mathbf{N})$, which has a single 1 in the n^{th} slot and zeros everywhere else. The sequence converges weakly to the vector 0. To see this, let $a = (a_0, a_1, \ldots)$; then

$$(e^{(n)}, a) = a_n.$$

Since

$$\sum_{j=0}^{\infty} |a_j|^2 < \infty,$$

$a_n \to 0$, and then by Proposition 3.7.11,

$$(e^{(n)}, a) = a_n \to 0$$

implies weak convergence of $e^{(n)}$ to 0.

Evidently, $\|e^{(n)} - 0\|_2 = 1$ for all n, so that the sequence $e^{(n)}$ does *not* approach zero in the strong (metric) topology. ∎

The preceding example shows that it is indeed easier for a sequence to converge weakly than to converge strongly. On the other hand, strong convergence does imply weak convergence, see Exercise 3.7.32.

3.7.15 Exampie. To understand intuitively what "weak convergence" means, one should look at function spaces rather than sequence spaces.
Let $f_n(x) = \sqrt{2}\sin n\pi x$ for $n \geq 1$. Clearly, $f_n \in L^2_0([0,1])$ and $||f_n||_2 = 1$. We will show later that $f_n \to 0$ (where now "0" stands for the zero function.) The proof will use the criterion in Proposition 3.7.13. The goal for the moment is to understand that

$$\lim_{n\to\infty} \sqrt{2}\int_0^1 a(x)\sin n\pi x\, dx = 0$$

is a) believable, and b) helps one to appreciate the concept of weak convergence.

Let us take, for the arbitrary function a, a step function, as in Figure 48. More general functions may be thought of as limits of step functions, Figure 49. (We will not need to make this statement more precise.)

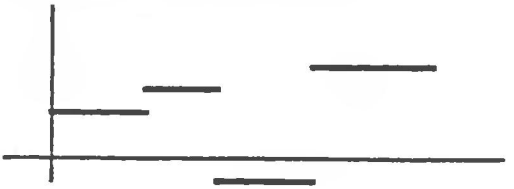


Figure 48. *A step function*

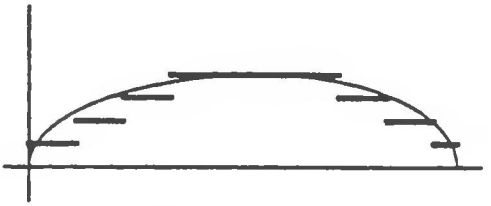


Figure 49. *Approximating a function by a step-function*

Now, imagine multiplying a step function a by $\sqrt{2}\sin n\pi x$, taking a large n, as shown in Figure 50. Integrate the product. Evidently, over most of the intervals of constancy of a, the rapid positive oscillations of the sin will cancel the negative oscillations, and the value of the integral will be very small. As $n \to \infty$, the part of sin that does *not* cancel (at the very ends of the intervals of constancy of a) contributes less and less to the integral, and the limit of (f_n, a) is zero.

Intuitively speaking: weak convergence averages out rapid oscillations of the functions f_n in the sequence. The positive parts tend to cancel the negative parts. Note how very differently one proceeds to check whether or not a sequence f_n converges strongly — to the zero function, say. *First*, one squares all the f_n, to *keep the positive values from cancelling the negative ones*, and *then* one integrates. Strong convergence to 0 can only happen if, somehow, the total area under the graph of the squared functions tends to zero. These interpretations should also make clear why it is easier for a sequence to converge weakly than strongly. ∎

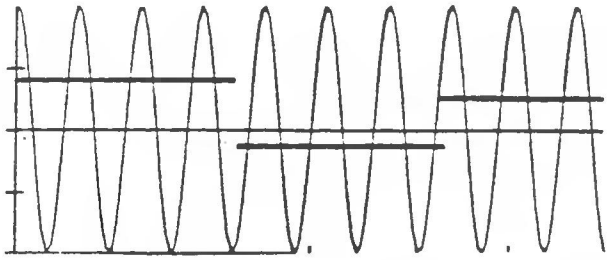


Figure 50. *How to converge weakly by oscillating strongly*

Figure 51. *A breaking wave*

3.7.16 Example. Weak convergence, interpreted as just described, is encountered in the theory of partial differential equations. A typical illustration is afforded by the so-called *small dispersion limit of the Korteweg - deVries equation*, which at this point can only be described in the most general terms.

The Korteweg - deVries equation is

$$u_t + uu_x + \epsilon^2 u_{xxx} = 0. \tag{3.7.17}$$

The unknown function $u(t,x)$ represents a disturbance propagating along the x-axis. Solutions of the partial differential equation $u_t + uu_x = 0$ tend to steepen and topple over, as do water waves running up onto a beach (Figure 51).

One can understand Figure 51 as follows. The partial differential equation $u_t + au_x = 0$, with constant a, has the general solution $u(t,x) = f(x - at)$ (just plug in and check). This solution describes a shape f moving to the right at speed a. By analogy, the solution of $u_t + uu_x = 0$ should be $u = f(x - ut)$. When $t = 0$, this gives $u(0,x) = f(x)$, so that the function f is the *initial condition*. The relation $u = f(x - ut)$ determines u as function of (t,x), at least as long as the equation can be solved for u, and it says that the shape f moves to the right (or left, if $u < 0$) at a speed = to the height u. In other words, the higher parts of the initial shape f move faster than do the lower parts. Figure 51 shows the result.

The term $\epsilon^2 u_{xxx}$ in (3.7.17) models a physical process called *dispersion*: waves of different lengths propagate at different speeds.[26] Dispersion counteracts

[26] As when white light passes through a prism: components of different wavelengths travel at different speeds, so that the colors are separated.

steepening of the wave, causing ripples to form on the bending crest. (figure 52.)

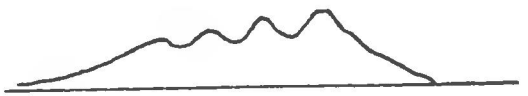


Figure 52. *Dispersion versus breaking*

As the (physical) parameter ϵ tends to 0, the number of ripples increases. One now wants to know: in the limit $\epsilon \to 0$ of small dispersion, how is the rippling wave an approximation to the breaking wave? It will be clear from our earlier discussion that one should not expect the limit to exist in any strong sense. Only if the oscillations of the ripples somehow average each other away can one hope to see the shape on the right in Figure 53 as limit of the shapes on the left.

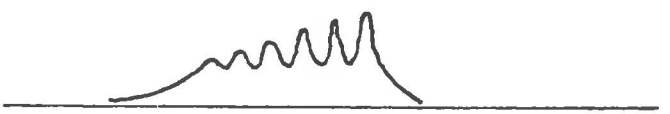


Figure 53. *What happens as the rippling increases?*

The details are much more complicated than this superficial discussion might lead one to believe. I only wanted to argue that physical processes can lead, in a natural way, to weak convergence. One also finds that schemes for the numerical solution of an equation like $u_t + uu_x = 0$ can, because of discretization of all the variables, introduce artificial small dispersion and cause spurious oscillations in the computed solution. Figure 54, taken from the preprint *Dispersive Approximations in Fluid Dynamics* by T. Hou and P. D. Lax, (New York University, 1990) shows the result of one computer simulation. The analysis of such numerical instabilities again relies on the theory of weak convergence in function spaces. ∎

3.7.18 Example. I want to emphasize the intuition behind *weak convergence*: integrals tend to zero because ever more rapid oscillations of the integrands tend to cancel each other. This is a powerful idea with many applications. The mechanism

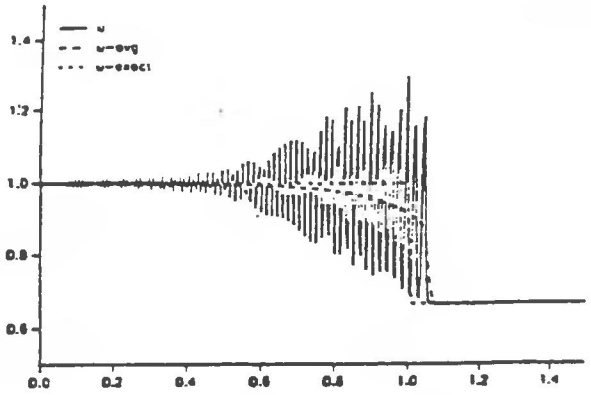


Figure 54. *Numerical solution for an IVP for a 1-dimensional compressible gas*

of cancellation may survive even if the official setting of weak convergence does not. Here is one more example: it is a special case of the *Riemann-Lebesgue lemma*, which is important in Fourier theory.

Let φ be a continuously differentiable function, suppose that

$$\int_{-\infty}^{\infty} |\varphi(x)|\,dx < \infty, \qquad \int_{-\infty}^{\infty} |\varphi'(x)|\,dx < \infty$$

and that

$$\lim_{|x|\to\infty} |\varphi(x)| = 0. \tag{3.7.19}$$

Then, I claim,

$$\lim_{k\to\infty} \int_{-\infty}^{\infty} \sin kx\, \varphi(x)\,dx = 0. \tag{3.7.20}$$

To prove this, write

$$\sin kx = -\Big(\frac{\cos kx}{k}\Big)'$$

and integrate by parts:

$$\int_{-\infty}^{\infty} \sin kx\, \varphi(x)\,dx = -\frac{\cos kx}{k}\varphi(x)\Big|_{-\infty}^{\infty} + \frac{1}{k}\int_{-\infty}^{\infty} \cos kx\, \varphi'(x)\,dx. \tag{3.7.21}$$

The first term vanishes by (3.7.19). The second integral is finite,

$$\left| \int_{-\infty}^{\infty} \cos kx\, \varphi'(x)\, dx \right| \leq \int_{-\infty}^{\infty} |\varphi'(x)|\, dx < \infty.$$

by assumption. As $k \to \infty$, the right side of (3.7.21) therefore tends to zero.

The result has the following physical interpretation. A fairly general function φ can be written as a continuous linear superposition of sines and cosines,

$$\varphi(x) = \int_{-\infty}^{\infty} (\exp ikx)\hat{\varphi}(k)\, dk, \tag{3.7.22}$$

where

$$\hat{\varphi}(k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (\exp(-ikx))\varphi(x)\, dx. \tag{3.7.23}$$

The argument just given above shows that the amplitude of the $\exp ikx$ component, assuming smallness at infinity of φ, decreases to zero as $|k| \to \infty$. In words: the contribution of highly oscillatory trigonometric functions to the integral representation (3.7.22) is small. In applied words: smooth, rapidly decaying signals have negligible short-wavelength components.

This is a basic result of Fourier analysis, and will appear again in Chapters II and III. I include it to illustrate the mechanism of cancellation by oscillation, rather than the definition of weak topology or weak convergence. ∎

It is now time to fill in the missing details and prove the weak-convergence criteria (3.7.11) and (3.7.13).

Proof of Proposition 3.7.11. First, suppose that $\mathbf{x}^{(n)} \to 0$ in the sense of Definition 3.7.10. Then, for every weakly open set V containing $\mathbf{0}$, there is an $N_V > 0$ such that $\mathbf{x}^{(n)} \in V$ for all $n \geq N_V$. This will be true, in particular, for a set V of the form

$$V = \{\mathbf{y} \mid -\epsilon < (\mathbf{y}, \mathbf{a}) < \epsilon\}. \tag{3.7.24}$$

Such a V is weakly open according to Definition 3.7.4; indeed, if we define the linear function $f_{\mathbf{a}}$ as in (3.7.2), we find that

$$V = f_{\mathbf{a}}^{-1}((-\epsilon, \epsilon)).$$

Now choose $\mathbf{a}$, and pick an arbitrary $\epsilon > 0$. There is an N (depending on a and ϵ), such that for $n \geq N$, we have $\mathbf{x}^{(n)} \in V$. But, according to (3.7.24), $\mathbf{x}^{(n)} \in V$ means $-\epsilon < (\mathbf{x}^{(n)}, \mathbf{a}) < \epsilon$, or $|(\mathbf{x}^{(n)}, \mathbf{a})| < \epsilon$. Hence $(\mathbf{x}^{(n)}, \mathbf{a}) \to 0$ in $\mathbf{R}$.

Conversely, suppose that for a sequence $\mathbf{x}^{(n)}$ in $l^2(\mathbf{R}, \mathbf{N})$, the sequence of numbers $(\mathbf{x}^{(n)}, \mathbf{a})$ converges to $0 \in \mathbf{R}$ for every $\mathbf{a} \in l^2(\mathbf{R}, \mathbf{N})$. We are to prove that $\mathbf{x}^{(n)}$ must then converge weakly to the zero sequence in the sense of Definition 3.7.10.

Let V be a weakly open set containing 0. By definition of the weak topology, there is a basis set B between 0 and V: $0 \in B \subset V$. A basis set, according to (3.4.10), has the form

$$B = \bigcap_{i=1}^{M} f_{\mathbf{a}_i}^{-1}(U_i),$$

where $f_{\mathbf{a}_i}$ is defined in (3.7.2), and U_i is open in $\mathbf{R}$. One may choose these U_i to have a very simple form. Since the zero sequence 0 lies in B, we must have $0 \in f_{\mathbf{a}_i}^{-1}(U_i)$ for every i. Since $f_{\mathbf{a}_i}(0) = 0$, it must be that $0 \in U_i$. But U_i is open in $\mathbf{R}$, and so it contains an open interval $(-\epsilon_i, \epsilon_i)$. Therefore

$$0 \in C \stackrel{\text{def}}{=} \bigcap_{i=1}^{M} f_{\mathbf{a}_i}^{-1}((-\epsilon_i, \epsilon_i)) \subset \bigcap_{i=1}^{M} f_{\mathbf{a}_i}^{-1}(U_i).$$

Figure 55 attempts to depict the situation: the weakly open set V contains 0, and a basis set B (everything that's shaded) fits between 0 and V. The basis set B contains an even more special (and simple) set C.

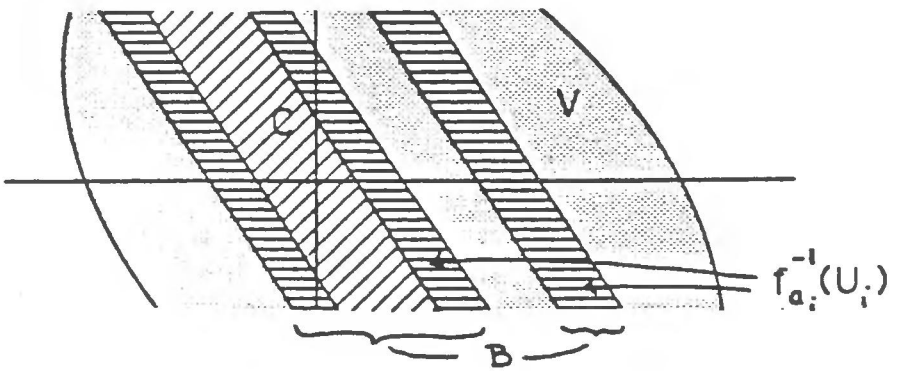


Figure 55. *Criterion for weak convergence*

Now suppose that the condition in Proposition 3.7.11 is satisfied. Then, for each $i = 1, \ldots, M$, there is an integer N_i such that $|(\mathbf{x}^{(n)}, \mathbf{a}_i)| < \epsilon_i$ for $n \geq N_i$. When $n \geq \max\{N_1, \ldots, N_M\}$, then $|(\mathbf{x}^{(n)}, \mathbf{a}_i)| < \epsilon_i$ for all $i = 1, \ldots, M$. In other words, for such indices n, $\mathbf{x}^{(n)}$ will belong to

$$\bigcap_{i=1}^{M} f_{\mathbf{a}_i}^{-1}((-\epsilon_i, \epsilon_i)) \subset V.$$

Thus, $x^{(n)} \in V$ for n larger than a certain N, and since V was arbitrary, we have proved that $x^{(n)} \to 0$. ∎

We will now prove the result claimed in Example 3.7.15:

3.7.25 Proposition. *The sequence $f_n(x) = \sqrt{2}\sin n\pi x$ converges weakly to zero in $L_0^2([0,1])$.* ∎

The proof relies on a lemma called *Bessel's inequality*, which is very important in the theory of Fourier series. To formulate it, we introduce some notation:

$$\begin{aligned} f_n(x) &= \sqrt{2}\sin n\pi x; \\ (f,g) &= \int_0^1 f(x)g(x)\,dx; \\ (f_n, a) &= c_n. \end{aligned} \tag{3.7.26}$$

Some elementary properties of the inner product (doubtless familiar from vector analysis) are also required, and are left to the reader.

- **3.7.27 Exercise.** Let $f \in L_0^2$. Then $(f,f) = \|f\|_2^2$. ∎
- **3.7.28 Exercise.** If $(e,f) = 0$, then $(e+f, e+f) = \|e\|_2^2 + \|f\|_2^2$. ∎
- **3.7.29 Exercise.** Let f_n, f_m be as in (3.7.26). Then $(f_n, f_m) = 0$. ∎

3.7.30 Lemma. *Bessel's Inequality. Let $f \in L_0^2$. Then for every $a \in L_0^2$ and every integer $N > 0$,*

$$\sum_{n=1}^{N} |(f_n, a)|^2 = \sum_{n=1}^{N} |c_n|^2 \le \|a\|_2^2. \quad ∎$$

Proof of Bessel's inequality. We have

$$\begin{aligned} \|a - \sum_{n=1}^{N} c_n f_n\|_2^2 &= (a - \sum_{n=1}^{N} c_n f_n, a - \sum_{n=1}^{N} c_n f_n) \\ &= (a,a) - 2(a, \sum_{n=1}^{N} c_n f_n) + (\sum_{n=1}^{N} c_n f_n, \sum_{n=1}^{N} c_n f_n) \\ &= \|a\|_2^2 - 2\sum_{n=1}^{N} |c_n|^2 + \sum_{n=1}^{N} |c_n|^2. \end{aligned} \tag{3.7.31}$$

To get the last line, we used

$$(a, \sum_{n=1}^{N} c_n f_n) = \sum_{n=1}^{N} (a, f_n) c_n = \sum_{n=1}^{N} |c_n|^2$$

and

$$(\sum_{n=1}^{N} c_n f_n, \sum_{m=1}^{N} c_m f_m) = \sum_{n=1}^{N}\sum_{m=1}^{N} c_n c_m (f_n, f_m) = \sum_{n=1}^{N} |c_n|^2.$$

The first and last line of (3.7.31) give

$$0 \le ||a - \sum_{n=1}^{N} c_n f_n||_2^2 = ||a||_2^2 - \sum_{n=1}^{N} |c_n|^2.$$

This proves Bessel's inequality. ∎

Proof of Proposition 3.7.25. Inasmuch as the N in Bessel's inequality is arbitrary, we can let it tend to ∞. Thus,

$$\sum_{n=0}^{\infty} |(f_n, a)|^2 \le ||a||_2^2 < \infty.$$

The terms of a convergent series must approach zero, so that $(f_n, a) \to 0$. Weak convergence now follows from Proposition 3.7.13, which is proved just as in the sequence space case. ∎

• **3.7.32 Exercise.** Prove: if a sequence $\mathbf{x}^{(n)}$ converges strongly (in the norm topology) to $\mathbf{0}$ in $l^2(\mathbf{R}, \mathbf{N})$, then it converges weakly to $\mathbf{0}$. (You can prove this by verifying the condition in Proposition 3.7.11, or by comparing the weak and strong topologies.) ∎

• **3.7.33 Exercise.** Does the sequence $\{x^n\}$ converge weakly to zero in $L_0^2([0,1])$? ∎

• **3.7.34 Exercise.** Set $f_n(x) = \sin n\pi x$ on $[0,1]$. Show that f_n^2 converges weakly to the function $1/2$ in L^2. (Hint: $f_n^2(x) = (1 - \cos 2n\pi x)/2$. Now show that $\cos 2n\pi x$ converges weakly to zero.)

This example turns out to be important: $f_n \to 0$ does not imply $f_n^2 \to 0$, and this can cause difficulties. With reference to Example 3.7.16, let $u^\epsilon(t,x)$ be a family of solutions of the Korteweg - deVries equation 3.7.17, and suppose that u^ϵ converges weakly to a function u as $\epsilon \to 0$. (One could let ϵ run over the numbers $1/n$ if one prefers a sequence.) Now, as in this exercise, it need not be the case that the term $u^\epsilon u_x^\epsilon$ has weak limit uu_x just because u^ϵ has weak limit u. Hence, one cannot assert that the limit u solves the $\epsilon = 0$ equation, $u_t - uu_x = 0$. Such difficulties (among others) make the theory of Example 3.7.16 very intricate. ∎

3.7.35 Exercise. Is the weak topology on $l^2(\mathbf{R}, \mathbf{N})$ Hausdorff? ∎

3.7.36 Exercise. Let $f \in L_0^2(\mathbf{R})$, and define $f_n(x) = f(x - n)$. Show that $f_n \to 0$. (Hint: Use the criterion in Proposition 3.7.13. Break the integral $\int g f_n \, dx$ into two pieces, and use Hölder's inequality on each piece.) Show that the sequence does not converge strongly to 0. ∎

3.7.37 Exercise. For $x \in [0,1]$, let $.x_1x_2\ldots$ be the unique base 2 expansion that has infinitely many zeros (compare (3.1.7)). Define functions f_n by

$$f_n(x) = \begin{cases} +1. & \text{if } x_n = 1 \\ -1, & \text{if } x_n = 0. \end{cases}$$

Explain (heuristically) why $f_n \to 0$ as $n \to \infty$. Then *prove* that the sequence converges weakly to the zero function.

These f_n are called *Rademacher functions;* they arise in probability theory. Toss a fair coininfinitely many times; record a 1 if the toss results in "heads" and a 0 if the toss results in "tails." The result of one such experiment is then a sequence $(x_1, x_2, \ldots)$ of 0's and 1's, or, equivalently, a point $.x_1x_2\ldots$ in $[0,1]$. Then define a function X_n by $X_n = 1$ if the n^{th} toss is heads, and $X_n = -1$ if the n^{th} toss is tails. These functions are essentially the Rademacher functions f_n; there is a slight (and, as it turns out, unimportant) ambiguity arising from the existence of two distinct binary expansions for dyadic rationals. The statistical properties of coin-tossing are analyzed in terms of the X_n. For instance, $X_1 + \ldots + X_N$ is the excess of heads over tails, and so on. ∎

3.7.38 Exercise*. Let $\mathbf{x}^{(m,n)} \in l^2(\mathbf{R}, \mathbf{N})$ be the sequence whose m^{th} entry is 1, n^{th} entry is m, and all other entries are zero. Show that 0 is in the closure in the weak topology of $A = \{\mathbf{x}^{(m,n)} \mid 1 \le m < n\}$. Show that no sequence of elements in A converges weakly to 0.

This example shows that one cannot always define "closure" of a set by taking limits of sequences. ∎

Section 4 Application: Theory of Distributions

This entire section is devoted to one application of the ideas developed so far. The *theory of distributions*, also called the *theory of generalized functions*, deals with certain kinds of singular functions which arise as limits of ordinary, well-behaved, calculus-type functions. Topology is present, but not obtrusive. A "generalized function" is defined to be a certain kind of function on a topological space whose elements are themselves functions. It helps to know this definition when one works with generalized functions, and so a little bit of topology language is useful. One can, however, get quite far by emphasizing the intuitive and calculational aspects of the theory of generalized functions, and I will try to do so.

We begin with the most important and popular generalized function, the so-called *Dirac delta function*. The discussion only suggests what "generalized function" might mean. Everything will start to make sense eventually.

Consider the (one-dimensional) heat equation (2.4.43) for the unknown function $u(t,x)$, $x \in \mathbf{R}$. It is easy to verify that the positive function

$$E(t,x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t} \tag{4.0.1}$$

solves $u_t = u_{xx}$ for $x \in \mathbb{R}, t > 0$. This E should therefore describe the temperature distribution in an infinite rod. What is the temperature at time $t = 0$, as function of x?

Unfortunately, one cannot set $t = 0$ in (4.0.1), so look at the graph of $E(t, x)$ versus x for a couple of values of t:

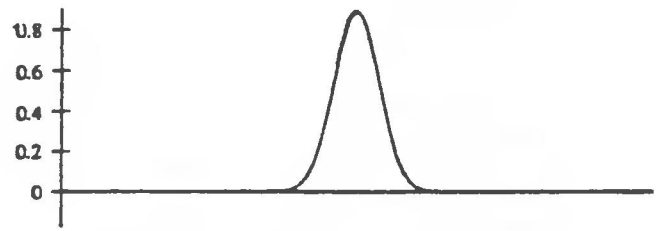


Figure 56. *Temperature created by a point source*

As $t \downarrow 0$, the temperature concentrates around $x = 0$. For a fixed $x \neq 0$, the limit of $E(t, x)$ as $t \downarrow 0$ is zero, while for $x = 0$, the limit of $E(t, 0) = 1/\sqrt{4\pi t}$ is ∞. The temperature therefore becomes infinite at $x = 0$. The total *heat* in the rod, however, remains finite.[27] Indeed, using the familiar result

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi},$$

one finds that

$$\int_{-\infty}^{\infty} E(t, x)\, dx = 1.$$

The physical picture is the following. A finite amount of heat is introduced into the rod at one point, $x = 0$, at time zero, and this heat diffuses as t becomes positive. One speaks of a *point source of heat*. This is only an idealized version of a physical situation, of course, but it is a very useful fiction.

Point sources are popular in physics: electrical charges concentrated at points, sources and sinks in fluid flows, vorticity concentrated along two-dimensional surfaces in three dimensional fluids — such abstractions allow one to model complicated phenomena. Mathematics should provide a calculus of point sources,

---

[27] E is the temperature, measured in degrees. When it is multiplied by the conversion factor ρc, where ρ is density and c is specific heat, it becomes *heat*, measured in units of energy. Being mathematicians, we take all physical constants to be unity.

and that is precisely what the theory of distributions does. The point heat source $\lim_{t\downarrow 0} E(t,x)$, for example, is a mathematical object called the *Dirac delta function*, $\delta(x)$,[28] which is supposed to have the following curious properties: it is 0 for $x \neq 0$ and ∞ for $x = 0$, and its integral over R is 1. This description is of little use for calculations: what is the x-derivative of such a function (the gradient of the initial temperature distribution)? Therefore, the mathematical theory takes a different path.

To illustrate the underlying idea, we consider a more general solution of the heat equation. Let

$$u(t,x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} e^{-y^2/4t} f(x-y)\,dy. \tag{4.0.2}$$

Here, f is a smooth function defined on R; assume that it decreases to zero at $\pm\infty$, so that the integral converges. We verify, first of all, that u really does satisfy $u_t = u_{xx}$.

Write

$$u(t,x) = \int_{-\infty}^{\infty} E(t,y) f(x-y)\,dy,$$

where E is the abbreviation introduced in (4.0.1). Now compute u_t, u_x, and u_{xx} by differentiating under the integral sign and integrating by parts; more information on the validity of this operation can be found in II.1.4.37. As in Section 2.4, subscripts denote partial derivatives.

$$\begin{aligned}
u_t(t,x) &= \int_{-\infty}^{\infty} E_t(t,y) f(x-y)\,dy; \\
u_x(t,x) &= \int_{-\infty}^{\infty} E(t,y) f'(x-y)\,dy \\
&= -E(t,y) f(x-y)\Big|_{y=-\infty}^{y=+\infty} + \int_{-\infty}^{\infty} E_y(t,y) f(x-y)\,dy \\
&= \int_{-\infty}^{\infty} E_y(t,y) f(x-y)\,dy
\end{aligned}$$

($E(t,y)f(x-y)$ vanishes at $y = \pm\infty$ because we assume that f approaches zero as its argument tends to ∞);

$$u_{xx}(t,x) = \int_{-\infty}^{\infty} E_{yy}(t,y) f(x-y)\,dy$$

---

[28] Paul A. M. Dirac, 1902-1984, introduced this "function" in his work on quantum electrodynamics. The mathematical underpinning for Dirac's delta and other highly singular functions was provided in the 1950's by Laurent Schwartz (1915 -). Schwartz was awarded the Fields medal, the mathematics version of the Nobel prize, for his work.

by a similar computation. Hence,

$$u_t(t,x) - u_{xx}(t,x) = \int_{-\infty}^{\infty} (E_t(t,y) - E_{yy}(t,y)) f(x-y)\,dy,$$

but $E_t - E_{yy} = 0$, and so u does satisfy $u_t - u_{xx} = 0$.

What is the initial temperature distribution, $u(0,x)$? Again, one cannot just set $t = 0$ in (4.0.2); rather, one must take the limit $t \downarrow 0$ in that fairly complicated expression. We state the result as a proposition, which will be proved at the end of the introduction. Before giving the (standard, typical, and important) proof, however, we will explain the meaning of the result, and indicate its relevance to the general definition of *distribution* that will be the subject of this section.

4.0.3 Proposition. *Let φ be continuous on* R, *and suppose that*

$$\sup_{-\infty<y<\infty} |\varphi(y)| < \infty.$$

Then

$$\lim_{t\downarrow 0} \int_{-\infty}^{\infty} E(t,y)\varphi(y)\,dy = \varphi(0). \blacksquare$$

4.0.4 Corollary. *If f satisfies the same conditions as φ in the proposition, then*

$$\lim_{t\downarrow 0} \int_{-\infty}^{\infty} E(t,y) f(x-y)\,dy = f(x). \blacksquare$$

Proof. For fixed x, let $\varphi(\cdot) = f(x - \cdot)$, and apply the proposition. ∎

The intuitive interpretation of formula (4.0.2) for the solution of the heat equation is interesting. Change the variable of integration to rewrite $u(t,x)$ as

$$\int_{-\infty}^{\infty} E(t, x-y) f(y)\,dy.$$

Now $E(t, x-y)$ is the temperature distribution produced by a unit point source of heat located at the point y. Likewise, $E(t,x-y)f(y)\,dy$ is the temperature distribution produced by a point source at y of (infinitesimal) strength $f(y)\,dy$. The solution (4.0.2) is a superposition of the temperature distributions due to a continuum of point sources whose strengths are described by the function f. A "continuum of point sources of strength $f(y)\,dy$" is just another way of saying "initial temperature distribution."

Proposition 4.0.3 introduces the rigorous mathematical definition of the Dirac delta function. Write $\delta = \lim_{t\downarrow 0} E(t,\cdot)$. δ is an object that takes a function φ,

as in the proposition, and associates to it the number $\varphi(0)$: we express this by writing $\delta(\varphi) = \varphi(0)$. In other words, δ is a *function of functions*. In this section, we will introduce, first of all, a topological space containing these functions φ; such functions are called *test functions* because the effect of δ and other distributions is "tested" on the φ's. A *distribution* will then be defined to be a continuous, real- (or complex-) valued linear function defined on the space of test functions. This definition will allow one to manipulate distributions pretty much as one does ordinary functions, but that part of the theory is only taken up, gradually, in later chapters.

To finish with the introductory remarks, we prove Proposition 4.0.3 in detail. You will see many more arguments of this sort, and should understand the steps.

The change of variables $y = \sqrt{4t}u$ turns the integral in (4.0.3) into

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} \varphi(\sqrt{4t}u)\, du,$$

and we are to prove that

$$\lim_{t \downarrow 0} \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} \varphi(\sqrt{4t}u)\, du \; - \varphi(0) \; = \; 0. \tag{4.0.5}$$

Since $\int_{-\infty}^{\infty} e^{-u^2}\, du = \sqrt{\pi}$, we can rewrite (4.0.5) as

$$\lim_{t \downarrow 0} \left(\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} \varphi(\sqrt{4t}u)\, du - \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-u^2} \varphi(0)\, du \right) \; = \; 0.$$

It suffices to prove that

$$\lim_{t \downarrow 0} \int_{-\infty}^{\infty} e^{-u^2} \left| \varphi(\sqrt{4t}u) - \varphi(0) \right| du \; = \; 0. \tag{4.0.6}$$

At this point, it is important to understand why (4.0.6) should be true. The integral splits into three parts, each of which can be made small, for different reasons.

i,ii) $\int_{-\infty}^{-M} \ldots$ and $\int_{M}^{\infty} \ldots$ are small if M is large, because e^{-u^2} is integrable and φ is bounded (by hypothesis);

iii) $\int_{-M}^{M} \ldots$ is small for small t because $\varphi(\sqrt{4t}u)$ and $\varphi(0)$ are close to each other.

In detail: Let $\epsilon > 0$ be given. To establish the limit in (4.0.6), we must find a $\delta > 0$ such that $0 < t < \delta$ implies that the integral in (4.0.6) is $< \epsilon$.

Set $K = \sup_{y \in \mathbf{R}} |\varphi(y)|$; this is finite by hypothesis. We then have

$$|\varphi(\sqrt{4t}u) - \varphi(0)| \leq |\varphi(\sqrt{4t}u)| + |\varphi(0)| \leq 2K.$$

Choose M so large that

$$\int_M^\infty e^{-u^2}|\varphi(\sqrt{4t}u) - \varphi(0)|\,du \le \int_M^\infty e^{-u^2}\cdot 2K\,du < \frac{\epsilon}{3};$$

then, likewise,

$$\int_{-\infty}^{-M} e^{-u^2}\,du < \frac{\epsilon}{6K}.$$

This is possible, since $\lim_{M\to\infty}\int_M^\infty \exp(-u^2)\,du = 0$.

Having fixed M, note now that if $-M \le u \le M$, then $-M\sqrt{4t} \le \sqrt{4t}u \le M\sqrt{4t}$. Since φ is continuous, there is an $\eta > 0$ such that $|y| < \eta$ implies $|\varphi(y) - \varphi(0)| < \epsilon/3\sqrt{\pi}$. Choose t so small that $M\sqrt{4t} < \eta$ (i.e., choose $\delta = \eta^2/4M^2$.) Then $|\varphi(\sqrt{4t}u) - \varphi(0)| < \epsilon/3\sqrt{\pi}$ as long as $-M\sqrt{4t} \le \sqrt{4t}u \le M\sqrt{4t}$, and so for $0 < t < \delta$,

$$\int_{-M}^{M} e^{-u^2}|\varphi(\sqrt{4t}u) - \varphi(0)|\,du < \int_{-M}^{M} e^{-u^2}\cdot\epsilon/3\sqrt{\pi}\,du < \int_{-\infty}^{\infty} e^{-u^2}\cdot\epsilon/3\sqrt{\pi}\,du = \epsilon/3.$$

Finally, then, the integral in (4.0.6) is estimated by

$$\int_{-\infty}^{-M} + \int_{-M}^{M} + \int_M^\infty < \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon.$$

This establishes (4.0.6). ∎

• **4.0.7 Exercise.** Adapt this argument to prove that

$$\lim_{a\downarrow 0}\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{a}{x^2+a^2}\varphi(x)\,dx = \varphi(0).$$

First, draw some graphs and explain why the result should be expected to be true. ∎

• **4.0.8 Exercise.** For $a > 0$, define the function δ_a by

$$\delta_a(x) = \begin{cases} 1/2a. & \text{if } -a < x < a \\ 0 & \text{otherwise.} \end{cases}$$

Prove that for φ satisfying the condition in Proposition 4.0.3,

$$\lim_{a\downarrow 0}\int_{-\infty}^{\infty}\delta_a(x)\varphi(x)\,dx = \varphi(0). \text{ ∎}$$

4.0.9 Exercise. Find conditions on φ that guarantee:

$$\lim_{a\downarrow 0}\int_{-\infty}^{\infty}\left(\frac{1}{\pi}\tan^{-1}\left(\frac{x}{a}\right) + \frac{1}{2}\right)\varphi(x)\,dx = \int_0^\infty \varphi(x)\,dx. \text{ ∎}$$

This section is organized as follows. Subsection 1 lists a few more examples of distributions; these will help you get acquainted with the idea, and will provide a reservoir of illustrations for the more theoretical discussion that is to follow. Subsection 2 introduces several new function spaces that are important in the theory of distributions: infinitely differentiable functions, infinitely differentiable functions with compact support, and the Schwartz space of rapidly decreasing functions. This is where topology comes in. Subsection 3 explains how "generalized functions" get their name and the official definition of *distribution* is given, but interesting examples and applications only come much later.

Subsection 4.1 Examples of Distributions

This subsection lists examples of generalized functions. The theory that unifies the examples is not presented here. The goal is only to show that elementary calculus leads very quickly to subtle problems: something diverges — how can I understand the divergence?

Vague Words. *If you do a (correct) calculation and come up with the answer $\pm\infty$ or, even worse, $+\infty - \infty$, and if I find a way to fudge your calculation to get a well-defined, finite number, then I have invented a* ***regularization procedure.*** ∎

If this means nothing now, it might make sense after you study the next few examples.

4.1.1 Example. We all know that the harmonic series $1 + \frac{1}{2} + \frac{1}{3} + \ldots$ diverges, but that is not the end of the matter. The limit

$$\lim_{n\to\infty} 1 + \frac{1}{2} + \ldots + \frac{1}{n} - \ln n$$

exists; it is called *Euler's constant*, its value is .57721566490135 . . . to a few decimal places, and it appears all over the place in the theory of special functions (such as the Gamma function.) Put differently, it is not only important to know *that* the harmonic series diverges, it is equally important to know *at what rate* it diverges.

Subtracting $\ln n$ from the partial sums of the harmonic series is one method of *regularization*. It is important because this *regularized* harmonic series is encountered in a variety of problems. ∎

4.1.2 Example. A more familiar example of regularization is afforded by the *improper Riemann integral*. Recall that one defines, for a piecewise continuous function f on $[0, \infty)$,

$$\int_0^\infty f(x)\,dx \stackrel{\text{def}}{=} \lim_{N\to\infty} \int_0^N f(x)\,dx, \tag{4.1.3}$$

if this limit exists. This kind of integral is *not* necessarily a measure of the total area under the graph of f, as is the ordinary integral we learn about in beginning calculus. For example, we could define a piecewise continuous function f as follows:

$$f(x) = \begin{cases} 1, & \text{for } x \in [0,1), \\ -1/2, & \text{for } x \in [1,2), \\ +1/3, & \text{for } x \in [2,3), \\ -1/4, & \text{for } x \in [3,4), \\ & \text{etc.} \end{cases}$$

The improper Riemann integral of f converges, because the alternating harmonic series $1 - \frac{1}{2} + \frac{1}{3} - \ldots$ converges. The total area under the graph of f is, however, infinite.

In what sense is this "regularization"? If you have learned that "integral = positive area minus negative area", you might be inclined to add up all the positive areas first,

$$1 + \frac{1}{3} + \frac{1}{5} + \ldots.$$

and then try to subtract the total negative area,

$$-\left(\frac{1}{2} + \frac{1}{4} + \ldots\right).$$

The result would be the meaningless difference $\infty - \infty$.

This way of assigning a meaning to $\int_0^\infty f(x)\,dx$ does not lead to a sensible answer. We will learn in Chapter II that this function f does *not* have an integral in the context of Lebesgue integration; that type of integral takes "area" seriously, and computes "total area" as "area above minus area below" When that prescription. leads to $\infty - \infty$, the integral is considered to be undefined. People who do not yet possess Advanced Knowledge are conditioned, by their sophomore Calculus courses, to interpret

$$\int_0^\infty f(x)\,dx$$

as

$$\lim_{N \to \infty} \int_0^N f(x)\,dx.$$

The improper integral is then assigned a value:

$$\int_0^\infty f(x)\,dx = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} \pm \ldots = \ln 2.$$

This regularization will be useful if the answer, $\ln 2$, can be put to use. A regularization is *never* correct or incorrect, only *useful or useless*, and *consistent with* or *inconsistent with* other theories. ∎

4.1.4 Example. Perhaps more interesting is the fact that

$$\int_0^\infty \sin(x^2)\,dx$$

is finite in the sense described in the preceding example; this integral is the *Fresnel integral* that arises in diffraction theory. The integral is finite as an improper Riemann integral, because the positive parts are cancelled by the negative parts, and vice versa, and neither has a chance to overwhelm the other.

Other integration theories (Lebesgue integration, for example) would consider the Fresnel integral to be nonexistent. Different applications require different strategies for dealing with singular or divergent objects and processes. Lebesgue, perhaps, did not care about optics; he looked at the total area above the x-axis, subtracted the total area below the x-axis, and arrived at $\infty - \infty$, which he considered to be undefined. Fresnel needed to know the limit

$$\lim_{N\to\infty} \int_0^N \sin(x^2)\,dx. \tag{4.1.5}$$

was not concerned about the total area, and was happy to be able to compute the limit to be $\sqrt{\pi/8}$.

The proof of the existence of the limit (4.1.5) is probably not familiar, so I present it in outline.

The change of variables $u = x^2$ converts

$$\int_0^N \sin x^2\,dx$$

into

$$\frac{1}{2}\int_0^{N^2} \frac{\sin u}{\sqrt{u}}\,du. \tag{4.1.6}$$

The limit $N \to \infty$ of (4.1.6) is evaluated by contour integration. Integrate (setting $z = u + iv$)

$$\frac{e^{iz}}{2\sqrt{z}}$$

over the contour shown in Figure 57. There are three different terms.

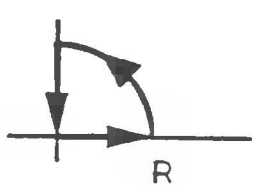


Figure 57.

The imaginary part of

$$\frac{1}{2}\int_0^R \frac{e^{iu}}{\sqrt{u}}\,du.$$

in the limit $R \to \infty$, is the number we want to find. The integral over the circular arc goes to zero as $R \to \infty$. The integral along the imaginary axis,

$$\frac{1}{2}\int_R^0 \frac{e^{-v}}{\sqrt{iv}}\,d(iv)$$

can be evaluated in the limit $R \to \infty$ (set $v = y^2$ and use $\int_0^\infty \exp(-y^2)\,dy = \sqrt{\pi}/2$). ∎

• **4.1.7 Exercise.** Fill in the details of the calculation just outlined. ∎

• **4.1.8 Exercise.** Find

$$\lim_{N\to\infty}\int_0^N \cos x^2\,dx.$$ ∎

4.1.9 Exercise*. The *Gamma function* $\Gamma(s)$ is defined for $s > 0$ by

$$\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}\,dt.$$

For what α does contour integration give a finite value of

$$\lim_{N\to\infty}\int_0^N \sin x^\alpha\,dx?$$

Find that limit in terms of the Gamma function. ∎

The next example describes an important, but probably less familiar regularization, process. A physical application is sketched in Supplement 4.1.74.

4.1.10 Example. *The Cauchy Principal Value.*[29] This is only one of several ways of assigning a finite value to the divergent integral

$$\int_{-\infty}^{\infty} \frac{\varphi(x)}{x}\,dx.$$

(A different, but related, method is presented in Supplement 4.1.36.) We assume that the function φ decreases rapidly to zero at $\pm\infty$; the divergence of the integral

[29] Augustin Louis Cauchy (1789-1857). One of the most prolific and creative mathematicians of "modern" times.

then occurs at $x = 0$, because of the denominator x. Cauchy defines ("PV" stands for "principal value")

$$\text{PV}\,\frac{1}{x}(\varphi) \stackrel{\text{def}}{=} \lim_{\epsilon \downarrow 0} \left(\int_{-\infty}^{-\epsilon} \frac{\varphi(x)}{x}\,dx + \int_{\epsilon}^{\infty} \frac{\varphi(x)}{x}\,dx \right). \tag{4.1.11}$$

We now show that this limit is always finite, but in addition, the proof is arranged to give somewhat more precise information, which will be useful later on. Obviously, one cannot prove anything without using properties of the function φ, so in the course of the argument we shall draw attention to the hypotheses to be imposed on φ. In the end, these assumptions will be codified in a precise definition of the space of test functions.

For a start, we require:

$$\varphi \text{ is continuous on } \mathbf{R}. \tag{4.1.12}$$

There are now two steps in the argument. *First*, we prove that

$$\int_{-\infty}^{-\epsilon} \frac{\varphi(x)}{x}\,dx + \int_{\epsilon}^{\infty} \frac{\varphi(x)}{x}\,dx \tag{4.1.13}$$

remains bounded as $\epsilon \downarrow 0$. This step shows that the integrals do not tend to $\pm\infty$. *Second*, we must show that the limit actually exists. This step is carried out by guessing the formula for the limit, and then proving rigorously that the guess is correct. The steps are quite elementary once you understand them, but since this is the first of several examples of this type, all details are given. They should be studied with care.

Notation. We write $\int_{|x|>\epsilon}$ for $\int_{-\infty}^{-\epsilon} + \int_{\epsilon}^{\infty}$. ∎

There is, first of all, a useful reduction to a simpler case. Note that

$$\begin{aligned} \int_{|x|>\epsilon} \frac{\varphi(x)}{x}\,dx &= \int_{-\infty}^{-1} \frac{\varphi(x)}{x}\,dx + \int_{-1}^{-\epsilon} \frac{\varphi(x)-\varphi(0)}{x}\,dx + \int_{-1}^{-\epsilon} \frac{\varphi(0)}{x}\,dx \\ &\quad + \int_{\epsilon}^{1} \frac{\varphi(x)-\varphi(0)}{x}\,dx + \int_{\epsilon}^{1} \frac{\varphi(0)}{x}\,dx + \int_{1}^{\infty} \frac{\varphi(x)}{x}\,dx. \end{aligned} \tag{4.1.14}$$

The sum of the third and fifth integrals on the right side of this identity can be evaluated explicitly:

$$\int_{-1}^{-\epsilon} \frac{\varphi(0)}{x}\,dx + \int_{\epsilon}^{1} \frac{\varphi(0)}{x}\,dx = \varphi(0) \cdot \left\{ \ln|x| \Big|_{-1}^{-\epsilon} + \ln|x| \Big|_{\epsilon}^{1} \right\} = 0.$$

The first and sixth integrals do not involve ϵ, so their limit as $\epsilon \downarrow 0$ obviously exists, *if the individual integrals converge*. To guarantee this, we assume that

$$K \stackrel{\text{def}}{=} \sup_{x \in \mathbf{R}} |x\varphi(x)| < \infty. \tag{4.1.15}$$

Then (for the first integral)

$$\left|\int_{-\infty}^{-1}\frac{\varphi(x)}{x}\,dx\right| \le \int_{-\infty}^{-1}\left|\frac{\varphi(x)}{x}\right|dx = \int_{-\infty}^{-1}\left|\frac{x\varphi(x)}{x^2}\right|dx \le \int_{-\infty}^{-1}\frac{K}{x^2}\,dx = K < \infty, \tag{4.1.16}$$

and similarly for the sixth integral. If we impose condition (4.1.15), then those two integrals converge.

We therefore need only prove that the sum of the second and fourth integrals,

$$\int_{-1}^{-\epsilon}\frac{\varphi(x)-\varphi(0)}{x}\,dx + \int_{\epsilon}^{1}\frac{\varphi(x)-\varphi(0)}{x}\,dx$$

has a limit. The function $\tilde{\varphi}(x) \stackrel{\text{def}}{=} \varphi(x) - \varphi(0)$ in these integrals vanishes at $x = 0$. So, to prove that the limit (4.1.11) exists, we need only show that

$$\lim_{\epsilon \downarrow 0}\left(\int_{-1}^{-\epsilon}\frac{\tilde{\varphi}(x)}{x}\,dx + \int_{\epsilon}^{1}\frac{\tilde{\varphi}(x)}{x}\,dx\right) \tag{4.1.17}$$

exists, whenever $\tilde{\varphi}(0) = 0$ (and $\tilde{\varphi}$ is subject to certain other conditions determined below).

Change the variable of integration in the first integral in (4.1.17) from x to $-x$. The expression (4.1.17) then becomes

$$\lim_{\epsilon \downarrow 0}\int_{\epsilon}^{1}\frac{\tilde{\varphi}(x)-\tilde{\varphi}(-x)}{x}\,dx. \tag{4.1.18}$$

We must prove that this limit exists. To do this, we appeal to the Mean Value Theorem from Calculus:

4.1.19 Mean Value Theorem. *Let f be a continuously differentiable real-valued function on an interval $[a, b]$. Then there exists a $c \in [a, b]$ such that*

$$f(b) - f(a) = f'(c)(b - a). \blacksquare$$

I cannot resist reproducing the pictorial proof we all saw in calculus:

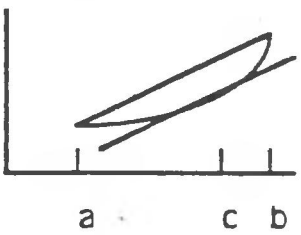


Figure 58. *The mean value theorem*

(And I challenge the reader to *prove* this most basic of analysis theorems without looking at a text!)

To be able to apply the Mean Value Theorem, we must assume that the function $\tilde{\varphi}$ is continuously differentiable on R. Then (set $b = x$ and $a = -x$, so that $b - a = 2x$) we can conclude that for every $x \in [0.1]$,

$$\tilde{\varphi}(x) - \tilde{\varphi}(-x) = 2x \cdot \tilde{\varphi}'(c_x)$$

for some $c_x \in [-x, x]$. Hence

$$\left| \int_\epsilon^1 \frac{\tilde{\varphi}(x) - \tilde{\varphi}(-x)}{x}\, dx \right| \leq \int_\epsilon^1 \left| \frac{\tilde{\varphi}(x) - \tilde{\varphi}(-x)}{x} \right| dx \leq \int_\epsilon^1 \left| \frac{2x\tilde{\varphi}'(c_x)}{x} \right| dx. \tag{4.1.20}$$

Enter the last assumption on φ:

$$\varphi' \text{ is continuous and } K_1 \stackrel{\text{def}}{=} \sup_{y \in \mathbf{R}} |\varphi'(y)| < \infty. \tag{4.1.21}$$

Note that since φ and φ' differ by a constant, the bound (4.1.21) also holds for $\tilde{\varphi}$, with the same K_1. If φ is differentiable, then it is continuous, and so assumption (4.1.21) subsumes (4.1.12).

Using (4.1.21), we continue the estimates in (4.1.20):

$$\leq \int_\epsilon^1 2 \sup_{c_x \in \mathbf{R}} \left| \tilde{\varphi}(c_x) \right| dx \leq \int_\epsilon^1 2K_1\, dx < 2K_1 < \infty. \tag{4.1.22}$$

We know now that the expression

$$\int_{|x|>\epsilon} \frac{\varphi(x)}{x}\, dx$$

that appears in the definition of Cauchy's principal value is bounded, independently of ϵ. We know, furthermore, that the limiting value, if it exists, is

$$\int_{-\infty}^{-1} \frac{\varphi(x)}{x}\, dx + \int_1^\infty \frac{\varphi(x)}{x}\, dx + \lim_{\epsilon \downarrow 0} \left(\int_{-1}^{-\epsilon} \frac{\tilde{\varphi}(x)}{x}\, dx + \int_\epsilon^1 \frac{\tilde{\varphi}(x)}{x}\, dx \right). \tag{4.1.23}$$

To show that the PV limit exists, and to find a formula for it, it suffices to prove existence, and find a formula for, the limit in (4.1.23).

The idea is simple. Set $\tilde{\varphi}(x)/x = \psi(x)$. By the Mean Value Theorem, because $\tilde{\varphi}$ is continuously differentiable, $\lim_{x \to 0} \psi(x)$ exists and equals $\tilde{\varphi}'(0)$. It stands to reason that the limit

$$\lim_{\epsilon \downarrow 0} \left(\int_{-1}^{-\epsilon} \frac{\tilde{\varphi}(x)}{x}\, dx + \int_\epsilon^1 \frac{\tilde{\varphi}(x)}{x}\, dx \right) \tag{4.1.24}$$

is nothing more than the integral of the continuous function ψ,

$$\int_{-1}^{1} \psi(x)\,dx.$$

The proof is harder to write down than to understand. To get $\lim_{\epsilon \downarrow 0} f(\epsilon) = A$, one shows that $|f(\epsilon) - A|$ has limit zero. This general rule applies to our situation, as follows. Set

$$f(\epsilon) = \int_{-1}^{-\epsilon} \frac{\tilde{\varphi}(x)}{x}\,dx + \int_{\epsilon}^{1} \frac{\tilde{\varphi}(x)}{x}\,dx$$

and

$$A = \int_{-1}^{1} \psi(x)\,dx.$$

The difference $|f(\epsilon) - A|$ is

$$\left| \int_{-\epsilon}^{\epsilon} \psi(x)\,dx \right|. \tag{4.1.25}$$

Since ψ is continuous, it is bounded, say $|\psi(x)| \leq R$ for some $R < \infty$. Then the integral in (4.1.25) is $\leq 2\epsilon R$, and this can certainly be made as small as you please if ϵ is taken small enough. This shows that the obvious guess for the limit (4.1.24) is correct.

We have now proved existence of, and found a formula for, the Cauchy Principal Value regularization. The result is summarized in a proposition.

4.1.26 Proposition. *Suppose that φ satisfies*

a) $\sup_{x \in \mathbf{R}} |x\varphi(x)| \stackrel{\text{def}}{=} K < \infty$.

b) φ' *is continuous, and* $\sup_{x \in \mathbf{R}} |\varphi'(x)| \stackrel{\text{def}}{=} K_1 < \infty$.

Then

$$\begin{aligned} \mathrm{PV}\frac{1}{x}(\varphi) &\stackrel{\text{def}}{=} \lim_{\epsilon \downarrow 0} \int_{|x|>\epsilon} \frac{\varphi(x)}{x}\,dx \\ &= \int_{-\infty}^{-1} \frac{\varphi(x)}{x}\,dx + \int_{-1}^{1} \frac{\varphi(x) - \varphi(0)}{x}\,dx + \int_{1}^{\infty} \frac{\varphi(x)}{x}\,dx. \end{aligned} \tag{4.1.27}$$

∎

It is time to step back and describe, in more general terms, the features common to the examples introduced so far. In each case, we take a function φ, and do something (peculiar?) to it in order to come up with a number:

1) $\delta(\varphi) = \varphi(0)$;
2) $\int_0^\infty \varphi(x)\,dx = \lim_{N \to -\infty} \int_0^N \varphi(x)\,dx$;
3) $\mathrm{PV}\frac{1}{x}(\varphi) = \lim_{\epsilon \downarrow 0} \int_{|x|>\epsilon} \varphi(x)/x\,dx$.

On the left side is the name of the operation done on φ, and on the right side is the prescription for getting a number. This is completely analogous to the calculus definition of functions of a real variable, such as

$$\exp x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots$$

or

$$\ln x = \int_1^x \frac{1}{s}\, ds.$$

The formulas 1), 2), 3) above therefore define *functions*, but they are now *functions of functions*, rather than functions of numbers. Once one adopts this point of view, it becomes clear what must be done to formulate a sensible theory.

First, one must specify a *domain*. The properties of φ in each of examples 1), 2), 3) were designed to fit the particular definition; it is convenient to impose somewhat more severe restrictions on φ which will permit application of many different "function rules" to a single class of φ's. The resulting domains, i.e., the set of φ for which functions like 1), 2), or 3) make sense, are called *test function spaces*. The most important test function spaces are introduced in Subsection 2.

Second, it is usually desirable to work with "nice" functions, even if their arguments are themselves functions. For example, one wants $\mathrm{PV}\frac{1}{x}(\varphi)$ to depend continuously on φ. This calls for the introduction of a *topology* on the test function space. The topologies encountered here are not derived from a norm, and sometimes not even from a metric: that is the excuse for spending time on these examples now rather than later.

Third, it turns out that many of the most useful operations done to functions φ to get numbers out of them are *linear* in φ. It is easy to check, for example, that

$$\delta(\alpha\varphi + \beta\psi) = \alpha\delta(\varphi) + \beta\delta(\psi),$$

with similar properties holding for examples 2) and 3) above. In Subsection 3, we will see how these functions of functions generalize the ordinary functions from calculus, and it will then become clear why linearity is a natural requirement.

The main definition of the theory can now be formulated, in outline.

4.1.28 Pseudo-Definition. *Let $\mathcal{X}$ be a test function space, i.e., a linear space of functions φ that satisfy some conditions yet to be spelled out. Let T be a continuous, complex-valued linear function on $\mathcal{X}$. Then T is called a* **distribution** *on $\mathcal{X}$.* ∎

The exercises define more distributions. Later chapters will refer to these examples, so you should become familiar with them. There are also three Supplements. The first presents another regularization of $\int_{-\infty}^{\infty} \varphi(x)/x\, dx$, the second introduces a delta "function" that describes a three-dimensional temperature distribution concentrated on an infinitely thin spherical shell, and the third explains how a principal value regularization arises in a fluid dynamics problem.

• **4.1.29 Exercise.** In the proof of existence of $\mathrm{PV}\frac{1}{x}$, why were the integrals split up over $-\infty$ to -1 to $-\epsilon$, *etc.*? Why don't we work with

$$\int_{|x|>\epsilon} \frac{\varphi(x)}{x}\,dx = \int_{|x|>\epsilon} \frac{\varphi(x)-\varphi(0)}{x}\,dx + \int_{|x|>\epsilon} \frac{\varphi(0)}{x}\,dx\,?$$ ∎

• **4.1.30 Exercise.** In the proof of existence of $\mathrm{PV}\frac{1}{x}$, there was no particular reason to split the integral from $-\infty$ to $-\epsilon$ at -1 (and the $[\epsilon,\infty)$ integral at $x=1$). Those were just nice round numbers. Suppose I split those ranges of integration at $x=\pm 2$. That will change formula (4.1.27). Do I still get the same number for $\mathrm{PV}\frac{1}{x}(\varphi)$? ∎

• **4.1.31 Exercise.** Evaluate $\mathrm{PV}\frac{1}{x}$ on the test function $1/(1+x^2)$. ∎

• **4.1.32 Exercise.** Define

$$\mathrm{PV}\frac{1}{x^2}(\varphi) = \lim_{\epsilon\downarrow 0} \int_{|x|>\epsilon} \frac{\varphi(x)-\varphi(0)}{x^2}\,dx.$$

Prove that this limit exists and is finite. You will need to impose additional hypotheses on φ! ∎

• **4.1.33 Exercise.** Define

$$\mathrm{PV}\frac{1}{x^3}(\varphi) = \lim_{\epsilon\downarrow 0} \int_{|x|>\epsilon} \frac{\varphi(x)-\varphi(0)-x\varphi'(0)}{x^2}\,dx.$$

Prove that this limit exists and is finite. You will need to impose additional hypotheses on φ! ∎

4.1.34 Exercise. Extrapolate from the preceding two exercises: define

$$\mathrm{PV}\frac{1}{x^n}(\varphi).$$

What hypotheses must be imposed on φ? ∎

4.1.35 Exercise. Set

$$E(t,x) = \frac{1}{4\pi t} e^{-\|x\|^2/4t},$$

with $x=(x_1,x_2)\in\mathbf{R}^2$, $t>0$, and $\|\cdot\|$ the Euclidean norm on $\mathbf{R}^2$. Show that E solves the 2-dimensional heat equation for $t>0$. Show that

$$\lim_{t\downarrow 0} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(t,x_1,x_2)\varphi(x_1,x_2)\,dx_1dx_2 = \varphi(0,0).$$

This is the *Dirac delta distribution* in $\mathbf{R}^2$. ∎

4.1.36 SUPPLEMENT: $(x+i0)^{-1}$

Another regularization of the divergent integral

$$\int_{-\infty}^{\infty} \frac{\varphi(x)}{x}\, dx$$

is suggested by complex function theory, specifically by Cauchy's formula:

$$\oint_{\Gamma} \frac{f(\zeta)}{\zeta - z}\, d\zeta = \begin{cases} 2\pi i f(z), & \text{if } z \text{ is inside ,} \\ 0, & \text{if } z \text{ is outside } Gamma. \end{cases} \tag{4.1.37}$$

The direction of integration is counterclockwise. Suppose we now let $z \to \zeta_0$, where ζ_0 is a point *on* Γ. Formula (4.1.37) does not make sense when $z = \zeta_0$, but the limit

$$\lim_{z \to \zeta_0} \oint_{\Gamma} \frac{f(\zeta)}{\zeta - z}\, d\zeta$$

will exist (under mild hypotheses on f and Γ):

$$\lim_{z \to \zeta_0} \oint_{\Gamma} \frac{f(\zeta)}{\zeta - z}\, d\zeta = \pm i\pi f(\zeta_0) + \lim_{\epsilon \downarrow 0} \oint_{\substack{\zeta \in \Gamma \\ |\zeta - \zeta_0| > \epsilon}} \frac{f(\zeta)}{\zeta - \zeta_0}\, d\zeta. \tag{4.1.38}$$

The + (resp. −) sign is used if $z \to \zeta_0$ from the inside (resp. outside) of Γ. This is *Plemelj's formula.*

4.1.39 Exercise. Verify Plemelj's formula for Γ = the unit circle and $f \equiv 1$. The limits can then be computed exactly. ∎

The regularization to be described looks like Plemelj's formula, but does not rely on the theory of analytic functions. It is related to Cauchy's Principal Value: this emerges in the final formula (4.1.51). A physical application is sketched in Supplement 4.1.74.

Define:

$$\frac{1}{x + i0}(\varphi) \stackrel{\text{def}}{=} \lim_{y \downarrow 0} \int_{-\infty}^{\infty} \frac{\varphi(x)}{x + iy}\, dx. \tag{4.1.40}$$

In other words, the distribution $(x + i0)^{-1}$ associates a number to each test function φ; that number is the value of the limit in (4.1.40). This number depends *continuously* on φ; we discuss this point later, after a topology has been introduced on the test function space (the space of suitable φ's). The assumptions imposed on φ will be *the same as assumptions* (4.1.15) *and* (4.1.21) *in Example* (4.1.10).

It is important to note that φ need *not* be analytic in the variable $x + iy$. The function φ may be complex-valued; this only means that it has the form $\varphi = \varphi_1 + i\varphi_2$, where φ_1 and φ_2 are *real-valued* functions of the *real* variable x, but φ_1 and φ_2 need not satisfy the Cauchy – Riemann equations. To find the complex integral in (4.1.40), just separate it

into real and imaginary parts, and do each of the resulting integrals separately. The key to the matter is: the complex variable $x + iy$ is *never* substituted into the function φ, and so there is no need to invoke analyticity of φ.

• **4.1.41 Exercise.** Let $\varphi = 1/(x^2+1)$. Find $\int_{-\infty}^{\infty} \varphi(x)/(x+iy)dx$ for $y > 0$. Does this have a limit as $y \downarrow 0$? ∎

We will prove that the limit in (4.1.40) exists and is finite. To this end, fix $\epsilon > 0$. Then

$$\int_{-\infty}^{\infty} \frac{\varphi(x)}{x+iy}\,dx = \int_{-\epsilon}^{\epsilon} \frac{\varphi(x)-\varphi(0)}{x+iy}\,dx + \int_{-\epsilon}^{\epsilon} \frac{\varphi(0)}{x+iy}\,dx + \int_{|x|>\epsilon} \frac{\varphi(x)}{x+iy}\,dx. \tag{4.1.42}$$

Call the three integrals I, II, and III. We will prove that they all remain finite as $y \downarrow 0$, and in fact relate the limit (4.1.40) to the Dirac δ and the Cauchy PV $\frac{1}{x}$.

Integral II. This is the easiest one: it can be evaluated explicitly, and the result is, with the $\varphi(0)$ factored out:

$$\int_{-\epsilon}^{\epsilon} \frac{x}{x^2+y^2}\,dx - i\int_{-\epsilon}^{\epsilon} \frac{y}{x^2+y^2}\,dx = \frac{1}{2}\ln(x^2+y^2)\Big|_{-\epsilon}^{\epsilon} - i\tan^{-1}\frac{x}{y}\Big|_{-\epsilon}^{\epsilon} = 0 - 2i\tan^{-1}\frac{\epsilon}{y}.$$

Hence, we get

$$\lim_{y\downarrow 0} \text{Integral II} = -i\pi\varphi(0). \tag{4.1.43}$$

The other two integrals require rather more words.

Integral I. Set $y = 0$ in this integral. The result is

$$\int_{-\epsilon}^{\epsilon} \frac{\varphi(x)-\varphi(0)}{x}\,dx, \tag{4.1.44}$$

which is finite, as explained in connection with equation (4.1.18) in Example 4.1.10. But be careful: *it does does not follow* that the limit $y \downarrow 0$ of Integral I exists, or is equal to (4.1.44)!

The last italicized statement warrants explanation. You may have a formula, call it Form(y), which makes sense for every number y. But it need *not* be the case that Form(0) = $\lim_{y\to 0}$ Form(y). Artificial examples are easy to construct. Define

$$\text{Form}(y) = \begin{cases} 1, & \text{if } y > 0 \\ 0, & \text{if } y = 0 \\ 1, & \text{if } y < 0. \end{cases}$$

Obviously, $\lim_{y\to 0}$ Form$(y) \neq$ Form(0). Of course, we don't like our formulas broken into pieces to illustrate an obscure point — so consider this next one. If you know about Fourier series, you will be able to derive the Fourier expansion of the constant function $\equiv 1$ on the interval $[0, \pi]$:

$$1 = 4\sum_{n=0}^{\infty} \frac{\sin(2n+1)y}{2n+1}.$$

Engineering students learn to compute this early in their career. The formula is correct for most values of y, but it is patently false for $y = 0$. If you graph partial sums (on a computer), it will seem that the limit as $y \downarrow 0$ is not the same as the value at zero (try it). The moral is: in elementary calculus, we construct artificial examples to show that "functions" can be discontinuous at some point. In advanced analysis, we meet very complicated formulas, and we should not *assume* that they represent continuous functions. In our Integral I, we have a parameter y that tends to 0. When we set $y = 0$, Integral I still makes sense. It *does not* follow that the value at $y = 0$ is the limit of the values at nearby y's.

Rigor now demands that we *prove*: the difference

$$\int_{-\epsilon}^{\epsilon} \frac{\varphi(x) - \varphi(0)}{x + iy}\, dx - \int_{-\epsilon}^{\epsilon} \frac{\varphi(x) - \varphi(0)}{x}\, dx \tag{4.1.45}$$

can be made as small as desired, if only $y > 0$ is small enough. Again, set $\varphi(x) - \varphi(0) = \tilde{\varphi}(x)$. The absolute value of the difference (4.1.45) is $\leq$ the integral of the absolute value of the difference of the difference of the integrands, which is

$$\begin{aligned} \int_{-\epsilon}^{\epsilon} |\tilde{\varphi}(x)| \left| \frac{1}{x+iy} - \frac{1}{x} \right| dx &= \int_{-\epsilon}^{\epsilon} \frac{|\tilde{\varphi}(x)|}{|x|} \left| \frac{-iy}{x+iy} \right| dx \\ &= \int_{-\epsilon}^{\epsilon} \frac{|\tilde{\varphi}(x)|}{|x|} \left| \frac{y}{\sqrt{x^2+y^2}} \right| dx. \end{aligned} \tag{4.1.46}$$

As in (4.1.21), we assume that $|\tilde{\varphi}'(y)|$ is bounded on $\mathbf{R}$ by a finite constant K_1. The Mean Value Theorem shows, as before, that $|\tilde{\varphi}(x)/x| \leq K_1$. The last integral in (4.1.46) is then

$$\leq K_1 \int_{-\epsilon}^{\epsilon} \frac{|y|}{\sqrt{x^2+y^2}}\, dx = 2K_1 \int_0^{\epsilon} \frac{y}{\sqrt{x^2+y^2}}\, dx = 2K_1 y \left[\ln(\sqrt{\epsilon^2+y^2}+\epsilon) - \ln y \right]; \tag{4.1.47}$$

hence, that integral goes to zero when y does.

• **4.1.48 Exercise.** This last paragraph is written in typical *analysis proof style*. I showed you once how the Mean Value Theorem is used, and from now on I don't bother to repeat those details in a similar type of problem. Write a careful proof of the claim that Integral I has limit

$$\int_{-\epsilon}^{\epsilon} \frac{\varphi(x) - \varphi(0)}{x}\, dx$$

as $y \downarrow 0$. ∎

Integral III. Again, Integral III remains defined as $y \downarrow 0$, so one *expects* that

$$\lim_{y \downarrow 0} \text{Integral III} = \int_{|x|>\epsilon} \frac{\varphi(x)}{x}\, dx. \tag{4.1.49}$$

We prove that this is indeed true. As for Integral I, we have

$$\begin{aligned}\left| \int_{|x|>\epsilon} \frac{\varphi(x)}{x+iy}\, dx - \int_{|x|>\epsilon} \frac{\varphi(x)}{x}\, dx \right| &\leq \int_{|x|>\epsilon} |\varphi(x)| \left| \frac{1}{x+iy} - \frac{1}{x} \right| dx \\ &\leq \int_{|x|>\epsilon} (\sup |\varphi(x)|) \frac{y}{|x|\sqrt{x^2+y^2}}\, dx.\end{aligned}$$

With $K = \sup_{x \in \mathbf{R}} |\varphi(x)|$, the last integral is bounded by

$$Ky \int_{|x|>\epsilon} \frac{1}{x^2}\, dx = \frac{2Ky}{\epsilon} < \infty.$$

This approaches 0 as $y \downarrow 0$, and the claim (4.1.49) is established. (Throughout this argument, ϵ has been a *fixed* positive number. Its only purpose in life was to separate the divergence – causing point $x = 0$ from the convergence – supporting points. You should not (yet) think of it as going to 0; if you did, you would become confused about that last constant, $2Ky/\epsilon$, which might then seem capable of becoming infinite.

We have established the following result. For arbitrary (but fixed) $\epsilon > 0$,

$$\lim_{y \downarrow 0} \int_{-\infty}^{\infty} \frac{\varphi(x)}{x+iy}\, dx = \int_{-\epsilon}^{\epsilon} \frac{\varphi(x) - \varphi(0)}{x}\, dx - i\pi\varphi(0) + \int_{|x|>\epsilon} \frac{\varphi(x)}{x}\, dx. \tag{4.1.50}$$

This proves that the limit (4.1.40) intended to define the distribution $(x + i0)^{-1}$ exists and is finite, but one can do more. As stated at the outset, one can express this limit in terms of the Dirac δ and the Cauchy PV $\frac{1}{x}$. The simplest way is to set $\epsilon = 1$, and to compare formula (4.1.50) with (4.1.27). More commonly, one now lets $\epsilon \downarrow 0$. To se what happens, write

$$\int_{-\epsilon}^{\epsilon} \frac{\varphi(x) - \varphi(0)}{x}\, dx = \int_{-\epsilon}^{\epsilon} \frac{\tilde{\varphi}(x)}{x}\, dx = \int_{-\epsilon}^{\epsilon} \psi(x)\, dx,$$

where ψ is the function introduced just before equation (4.1.24). Recall that ψ is continuous. Since the integral over $(-\epsilon, \epsilon)$ of a continuous function tends to zero as $\epsilon \downarrow 0$ (prove this!), the first integral in (4.1.50) tends to 0. This leaves:

$$\begin{aligned}\lim_{y \downarrow 0} \int_{-\infty}^{\infty} \frac{\varphi(x)}{x+iy}\, dx &= -i\pi\varphi(0) + \lim_{\epsilon \downarrow 0} \int_{|x|>\epsilon} \frac{\varphi(x)}{x}\, dx \\ &= -i\pi\delta(\varphi) + \text{PV}\, \frac{1}{x}(\varphi).\end{aligned} \tag{4.1.51}$$

(Compare this result with Plemelj's formula (4.1.38) from complex function theory.) (4.1.51) is usually abbreviated as follows:

$$\frac{1}{x+i0} = -i\pi\delta + \mathrm{PV}\,\frac{1}{x}. \tag{4.1.52}$$

This final version shows clearly that PV $\frac{1}{x}$ and $(x+i0)^{-1}$ provide different regularizations of the divergent integral $\int_{-\infty}^{\infty} \varphi(x)/x\,dx$. Neither is more *correct* than the other, but in a particular problem, one of the two may be more *appropriate*. Note, for example, that if φ is real-valued, then PV $\frac{1}{x}(\varphi)$ is a real number, while $(x+i0)^{-1}(\varphi)$ is in general complex. The latter distribution would certainly not be used in a problem in which complex numbers are (physically) meaningless. ∎

• **4.1.53 Exercise.** The distribution $(x-i0)^{-1}$ is defined by

$$\frac{1}{x-i0}(\varphi) \stackrel{\text{def}}{=} \lim_{y \downarrow 0} \int_{-\infty}^{\infty} \frac{\varphi(x)}{x-iy}\,dx.$$

Prove that this limit exists, and express it in terms of δ and PV $\frac{1}{x}$. ∎

• **4.1.54 Exercise.** Refer to 4.1.41 and to the preceding exercise. Evaluate $(x \pm i0)^{-1}$ on $\varphi(x) = 1/(1+x^2)$. ∎

• **4.1.55 Exercise.** Formula (4.1.52) and Exercise 4.1.53 together prove that

$$\frac{1}{x+i0} - \frac{1}{x-i0} = -2\pi i\delta.$$

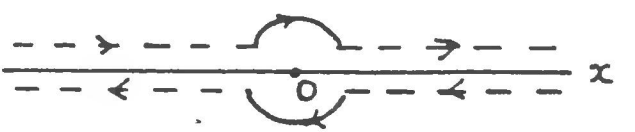
Prove this result directly, by evaluating the limit

$$\lim_{y \downarrow 0} \int_{-\infty}^{\infty} \varphi(x) \left(\frac{1}{x+iy} - \frac{1}{x-iy} \right) dx. \tag{4.1.56}$$

If you know complex function theory, you should try to understand the relation between (4.1.56) and a certain contour integral. Suppose that φ is holomorphic in a strip containing the real axis, and find the integral

$$\oint_{\Gamma} \frac{\varphi(z)}{z}\,dz,$$

where the contour Γ is



The circle has a radius r which tends to 0; outside of the circle, the contour is *on* the real axis, but it has been drawn slightly raised and lowered to indicate that these segments of the real axis are traversed twice, in opposite directions. Assume whatever you want to about φ; the point here is not rigor, but intuitive understanding. ∎

4.1.57 SUPPLEMENT: The delta distribution on a sphere.

The delta distribution studied at the beginning of the section picked out the value of a test function at the origin; it was interpreted as representing a point source of heat. The 2-dimensional delta distribution in Exercise (4.1.35) is again to be thought of as a point source of heat, this time in an infinite 2-dimensional conduction plate. There are other possibilities. We could define a distribution δ_a on $\mathbf{R}$ by the property

$$\delta_a(\varphi) = \varphi(a). \tag{4.1.58}$$

This picks out the value of the test function at $x = a$. Because the notation δ_a is so often used to denote an approximate delta function, as in Exercise 4.0.8, one should be clear about the intended meaning. Sometimes, it will be convenient to employ functional notation (illogical though it may be), and write $\delta(x - a)$ for this distribution. The motivation comes from formal manipulations:

$$\int_{-\infty}^{\infty} \delta(x)\varphi(x)\,dx = \varphi(0),$$

because "$\delta(x)$ is 0 except when $x = 0$, and then $\delta(0) = \infty$, with integral = 1" (physicists' heuristics). By the same token,

$$\int_{-\infty}^{\infty} \delta(x - a)\varphi(x)\,dx = \int_{-\infty}^{\infty} \delta(x)\varphi(x + a)\,dx = \varphi(a).$$

There are rigorous rules governing such calculations, but for the moment you might think of $\delta(x - a)$ as a point heat source of unit strength, placed at $x = a$ rather than at $x = 0$.

In one dimension, one can do little more. There are the distributions defined by

$$\sum_{j=1}^{n} c_j \delta(x - a_j)(\varphi) = \sum_{j=1}^{n} c_j \varphi(a_j),$$

which are quite useful, but these distributions just pick out values of test functions at specified points. They represent heat sources of strength c_j placed at the points a_j. The situation is much more interesting in two or more dimensions.

A concentrated heat source in one dimension can only be applied at a set of points (otherwise, it should not be called "concentrated"). In two or three dimensions, however, one can imagine not only point sources but also "filaments." Those are

infinitely thin "wires" embedded in an infinite plate ($\mathbf{R}^2$), carrying a (not necessarily uniform) heat distribution. One then asks about the development of the temperature of the plate with time. There could also be "sheets" of heat (in $\mathbf{R}^3$), infinitely thin two-dimensional surfaces. One encounters many other physically relevant examples: a wire or surface carrying an electric charge (as opposed to a point charge), a vortex line or vortex sheet in a fluid (as opposed to a point vortex, such as the perfect bathtub drain), and so on. We want to give just one example of such a distribution.

The example is relatively simple: we imagine a unit amount of heat evenly distributed over the sphere with center $\mathbf{0} \in \mathbf{R}^3$.[30] One can try several approaches to get formulas suitable for calculation. Two of them are illustrated. The first one is more intuitive, perhaps: the heat is thought of as being distributed over a shell of small but finite width ϵ; the desired distribution arises as the limit $\epsilon \downarrow 0$. The second way is to define this singular "function" directly by its action on test functions. Of course, the two methods are related, and in fact we arrive at the rigorous definition after understanding the intuitive approach.

Introduce the standard spherical polar coordinates ρ, θ, ϕ ($\theta \in [0, 2\pi)$ is the angle of longtitude (azimuth angle), and $\phi \in [0, \pi)$ is the angle of latitude (zenith angle)). Consider the spherical shell S_ϵ defined by $a - \epsilon \leq \rho \leq a + \epsilon$. Its volume is

$$V_\epsilon = \frac{8}{3}\pi\epsilon(3a^2 + \epsilon^2).$$

A heat source of total strength A, distributed uniformly over the shell, is therefore described by

$$\delta_{A,\epsilon}(x) = \begin{cases} A/V_\epsilon, & \text{if } x \in S_\epsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{4.1.59}$$

The definition is arranged to guarantee that

$$\int_0^{2\pi}\int_0^{\pi}\int_{a-\epsilon}^{a+\epsilon} \delta_{A,\epsilon}(x)\,\rho^2 \sin\phi\, d\rho d\phi d\theta = A. \tag{4.1.60}$$

It is clear that $\delta_{A,\epsilon}$ tends to zero (as $\epsilon \downarrow 0$) everywhere outside the shell, while it approaches ∞ on the shell. This is the behavior we have, by now, come to expect from a delta distribution. We also expect the limit to "pick out a value" of a test function φ; the meaning of this statement will be clarified presently. Let φ be a test function (depending on three variables, of course). For present purposes, it is sufficient to assume that φ is bounded and continuous. We want to evaluate

$$\lim_{\epsilon \downarrow 0} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta_{A,\epsilon}(x_1, x_2, x_3)\varphi(x_1, x_2, x_3)\, dx_1 dx_2 dx_3.$$

[30] *Sphere* means *surface of the sphere*; the sphere together with the volume inside is called a *ball*.

In polar coordinates, the triple integral becomes

$$\frac{3A}{8\pi\epsilon(3a^2+\epsilon^2)}\int_0^{2\pi}\int_0^{\pi}\int_{a-\epsilon}^{a+\epsilon}\varphi(\rho,\phi,\theta)\rho^2\sin\phi\,d\rho d\phi d\theta. \tag{4.1.61}$$

We simplify the formula by putting

$$\tilde{\varphi}(\rho)\overset{\text{def}}{=}\int_0^{2\pi}\int_0^{\pi}\varphi(\rho,\phi,\theta)\sin\phi\,d\rho d\phi d\theta; \tag{4.1.62}$$

$\tilde{\varphi}$ is a continuous function of $\rho\in[0,\infty)$ (you might try to prove this). With the notation (4.1.62), the integral (4.1.61) assumes the form

$$\frac{3A}{4\pi(3a^2+\epsilon^2)}\left(\frac{1}{2\epsilon}\int_{a-\epsilon}^{a+\epsilon}\tilde{\varphi}(\rho)\rho^2\,d\rho\right). \tag{4.1.63}$$

It is now quite easy to find the limit as $\epsilon\downarrow 0$. The messy numerical factor in (4.1.63) tends to $A/4\pi a^2$. The integral is just like the one in Exercise 4.0.8; its limit is $\tilde{\varphi}(\rho)\rho^2$ evaluated at $\rho=a$, namely $a^2\tilde{\varphi}(a)$. So finally, replacing $\tilde{\varphi}$ by the double integral in the definition (4.1.62), we find, for the limit in (4.1.61):

$$\frac{A}{4\pi}\int_0^{2\pi}\int_0^{\pi}\varphi(a,\phi,\theta)\sin\phi\,d\phi d\theta. \tag{4.1.64}$$

This formula has a very natural interpretation. The average of a function φ over the surface of the sphere of radius a is defined to be: the integral of φ over the sphere, divided by the surface area $4\pi a^2$ of the sphere:

$$\frac{1}{4\pi a^2}\int_0^{2\pi}\int_0^{\pi}\varphi(a,\phi,\theta)a^2\sin\phi\,d\phi d\theta.$$

Our limit (4.1.64), the "value" of the test function φ picked out by this delta distribution, is just *a multiple of the average of the test function over the surface of the sphere.* We now *define the distribution* $\delta(||x||-a)$ by choosing a particular value of A, $A=4\pi a^2$:

$$\delta(||x||-a)(\varphi)\overset{\text{def}}{=}a^2\int_0^{2\pi}\int_0^{\pi}\varphi(a,\phi,\theta)\sin\phi\,d\phi d\theta. \tag{4.1.65}$$

If the test function φ is equal to 1 in a neighborhood of the sphere of radius a, Definition 4.1.65 produces $4\pi a^2$, the surface area of the sphere. This should seem reasonable.

• **4.1.66 Exercise.** Define $\delta(||x||-a)$ in $\mathbf{R}^2$. Follow the outline of the preceding example. Start with a heat source distributed uniformly over an annulus $a-\epsilon\le r\le a+\epsilon$. Let $\epsilon\downarrow 0$, and compute the appropriate limit. Then define the distribution directly, by its action on a test function φ; when $\varphi\equiv 1$ near the circle, you should get the circumference $2\pi a$. ∎

4.1.67 Exercise°. The same question for the ellipse $x_1^2 + x_2^2/4 = 1$ in $\mathbb{R}^2$. ∎

Two remarks to conclude this example (about which more will be said, after the requisite techniques are developed).

Remark 1. For a consistent calculus of distributions, coordinates must be kept track of carefully. It turns out, for example, that a distribution called $\delta(\|x\|^2 - a^2)$ differs from the $\delta(\|x\| - a)$ defined above. We won't go into this now, but a simpler one-dimensional example should convey the idea.

How would one define the one-dimensional distribution $\delta(e^{50x} - 1)$? You should not think that just because $e^{50x} - 1 = 0$ only when $x = 0$, the distribution picks out the value $\varphi(0)$ again. To see this, consider the heuristic (but useful!) expression

$$\int_{-\infty}^{\infty} \delta(y)\varphi(y)\,dy = \varphi(0),$$

where φ has compact support in the interval $(-1, \infty)$, and set $y = e^{50x} - 1$, $dy = 50e^{50x}dx$:

$$\varphi(0) = \int_{-\infty}^{\infty} \delta(y)\varphi(y)\,dy = \varphi(0) = \int_{-\infty}^{\infty} \delta(e^{50x} - 1)\varphi(e^{50x} - 1)50e^{50x}\,dx. \tag{4.1.68}$$

Now let $\psi(x) = 50e^{50x}\varphi(e^{50x} - 1)$, so that

$$\varphi(0) = \int_{-\infty}^{\infty} \delta(e^{50x} - 1)\psi(x)\,dx. \tag{4.1.69}$$

The right side of (4.1.69) represents the action of the distribution $\delta(e^{50x} - 1)$ on the test function ψ. It is necessary to compute this number, which according to (4.1.68) equals $\varphi(0)$, in terms of ψ. Since

$$\psi(0) = 50\varphi(0),$$

we have

$$\int_{-\infty}^{\infty} \delta(e^{50x} - 1)\psi(x)\,dx = \frac{\psi(0)}{50},$$

or, put another way,

$$\delta(e^{50x} - 1) = \frac{1}{50}\delta(x). \tag{4.1.70}$$

You might, by now, believe that $\delta(\|x\| - a)$ and $\delta(\|y\|^2 - a^2)$ are different distributions. After all, if we think of them as "functions" appearing inside an integral, the change of variable that converts one to the other will carry along a Jacobian (or conversion factor from dx to dy), which changes the result. The moral is: the calculus of distributions is not insensitive to choice of coordinates.

• **4.1.71 Exercise.** Compute the action of $\delta(x^2 - 1)$ on a test function (in R). ∎

• **4.1.72 Exercise.** In one dimension, define the following distributions (by giving their action on a test function φ):

i) $\delta(ax)$, where $a > 0$ is fixed.
ii) $\delta(-x)$.
iii) $\delta(a - x)$, for given a.

How can you convince people that your definition is reasonable? Perhaps by presenting a sequence of functions that "converges" to the distribution. For example, we know that $\lim_{t \downarrow 0} E(t,x)$ is $\delta(x)$ ($E(t,x)$ is defined in (4.0.1)). Now replace x by ax. This method will work in all three cases. ∎

Remark 2. The distribution $\delta(||x|| - a)$ was introduced above as a possible initial heat distribution for the heat equation. It arises in other contexts as well. For example, it can be shown that the distribution

$$\frac{1}{4\pi} \frac{\delta(||x|| - t)}{t}, \quad t > 0. \tag{4.1.73}$$

solves the three-dimensional wave equation, $u_{tt}(t,x) = \Delta u(t,x)$. The concept of *distribution solution* of a partial differential equation has not yet been introduced, but the idea is fairly simple, in this example at least: (4.1.73) represents a spherical wavefront propagating outward from the origin. The $\delta(||x|| - t)$ part says that the disturbance is concentrated on the sphere $||x|| = t$, whose radius grows with time t. The factor $1/4\pi t$ models the loss of energy per unit of surface area of the expanding sphere. Distributions do not really arise as solutions of the heat equation because — as we shall see later — solutions of that equation do not show any singular behavior. Put differently, localized heat diffuses and gives rise to a smooth temperature distribution. If I heat one end of a needle until it is red hot, you will feel the warmth at the other end for some time. The wave equation, however, allows for localized solutions: if you clap your hand, the resulting sound wave propagates away, and if I stand at a little distance from you, I will hear the clap once, and then it is gone. The solution (4.0.1) of the heat equation is a model for the heated needle, while the solution (4.1.73) of the wave equation is a model for the handclap. ∎

4.1.74 SUPPLEMENT: Vortex lines

Regularizations of divergent integrals appear in many theoretical and applied problems. In this supplement, I describe one of them: an example taken from fluid dynamics. Certain facts from the theory of potential flow must be taken for granted.

We consider first the (very idealized) flow of an incompressible fluid in two dimensions, shown in Figure 59. The coordinates (x, y) are written in complex form as $\zeta = x + iy$. At point ζ, the velocity of a fluid particle passing through that point is a vector $(u(\zeta), v(\zeta))$. This velocity vector is assumed to be independent of time; one says that the flow is "stationary".

The circles shown in Figure 59 are the paths of fluid particles as they spiral around a point vortex located at ζ_0. No fluid enters or leaves this ideal vortex. The velocity vector is determined, in complex variable notation, from

$$u(\zeta) - iv(\zeta) = \frac{\Gamma}{2\pi i} \frac{1}{\zeta - \zeta_0}. \tag{4.1.75}$$

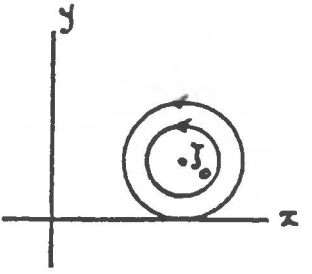


Figure 59. *A point vortex in two dimensions*

The constant Γ is the strength of the vortex; its sign determines the sense of rotation of the fluid. The figure is drawn for $\Gamma > 0$. When the vortex ζ_0 is located at the origin, one has

$$u = -\frac{\Gamma}{2\pi}\frac{y}{x^2+y^2}, \quad v = \frac{\Gamma}{2\pi}\frac{x}{x^2+y^2}.$$

These equations shows that the velocity is tangent to circles centered at the origin, confirming that particles rotate about the vortex. Near the origin, the speed increases; this agrees with experience also.

We now consider a continuous distribution of point vortices, as shown in Figure 60. The vortices lie along the x-axis, and the point vortex located at $(x', 0)$ has infinitesimal strength $\Gamma(x')\,dx'$. The velocity induced at point ζ by this infinitesimal vortex is, in accordance with (4.1.75),

$$\frac{1}{2\pi i}\frac{\Gamma(x')}{\zeta - x'}\,dx'. \tag{4.1.76}$$

The contributions (4.1.76) of the individual vortices are summed to give the total velocity at the point $\zeta = x + iy$, where first y is assumed to be positive:

$$u(\zeta) - iv(\zeta) = \frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{\Gamma(x')}{(x-x')+iy}\,dx'. \tag{4.1.77}$$

What about the induced velocity field *on* the vortex line itself? Obviously, if one sets $\zeta = x$ in (4.1.76) and then sums the infinitesimal contributions, one arrives directly at the divergent integral

$$\frac{1}{2\pi i}\int_{-\infty}^{\infty}\frac{\Gamma(x')}{x-x'}\,dx'.$$

The correct procedure, of course, is to let $y \downarrow 0$ in (4.1.77). The $(x \pm i0)^{-1}$ regularizations arise completely naturally in this problem.

One can equally well take $y < 0$. The limiting value of the velocity $u - iv$ on the $x-axis$ will differ from the limit coming from $y > 0$. In fact, Exercise 4.1.55 implies that the difference of the two limits is

$$(u_+(x,0) - iv_+(x,0)) - (u_-(x,0) - iv_-(x,0)) = -\Gamma(x), \tag{4.1.78}$$

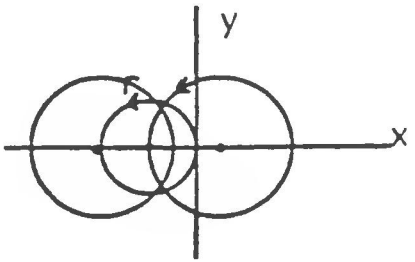


Figure 60. *A line of vortices on the x-axis*

where subscripts + and − denote the limits from $y > 0$ and $y < 0$ of (4.1.77). Taking real and imaginary parts of (4.1.78), one has

$$u_+(x,0) - u_+(x,0) = -\Gamma(x), \quad v_+(x,0) = v_-(x,0). \tag{4.1.79}$$

Formula (4.1.79) has an interesting physical interpretation. It says that the *normal* component v of the velocity is continuous across the vortex line, but that the tangential component u has a jump $-\Gamma$. This jump in velocity may be thought of as producing the vortex distribution along the x-axis.

The vortex line is often turned into a *vortex sheet;* this is obtained by lifting Figure 60 out of the page and creating a planar distribution of point vortices in the $x - z$ plane. You can find more information in fluid dynamics texts under the heading "vortex sheet".

• **4.1.80 Exercise.** Derive formula (4.1.78). ∎

Subsection 4.2 Test Function Spaces

In the Examples of the preceding Subsection, the hypotheses on the test functions φ were kept as unrestrictive as possible. For instance, (4.1.15) and (4.1.21) just suffice for proving existence of PV $\frac{1}{x}$, but they are not stringent enough to guarantee existence of the limit defining PV $\frac{1}{x^2}$ (see Exercise 4.1.32). One often needs to use several different distributions in one problem, and so one designs hypotheses on test functions that guarantee existence of all those distributions. There are several basic types of test functions, which we will describe, and others (more exotic) invented for special purposes; those will not be discussed. Look at EDM II, Article 125, for a quick survey of what you are not learning.

4.2.1 Notation. *We write* $(x_1, \ldots, x_n)$, *but mostly just* x, *for points in* $\mathbb{R}^n$. *Let* α *denote an n-tuple* $(\alpha_1, \ldots, \alpha_n)$ *of non - negative integers, and let* $|\alpha| = \alpha_1 + \ldots + \alpha_n$. *Abbreviate*

$$D_i = \frac{\partial}{\partial x_i}, \text{ for } i = 1, \ldots, n,$$

and set

$$D^\alpha \varphi(x) = D_1^{\alpha_1} \dots D_n^{\alpha_n} \varphi(x) = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \dots \partial x_n^{\alpha_n}} \varphi(x).$$

Let $|x| = \sqrt{x_1^2 + \dots + x_n^2}$. *The notations* $\|x\|_2$ *or* $\|x\|$ *may sometimes be used for emphasis, but* $|x|$ *is usually simpler and more convenient.*

Finally, multiple integrals are, for simplicity, written as single integrals:

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \varphi(x_1, \dots, x_n)\, dx_1 \dots dx_n = \int_{-\infty}^{\infty} \varphi(x)\, dx.$$

It will always be clear from the context whether there are several variables or just one. ∎

We begin by defining the three most commonly used test function spaces. Note that at this point, we only describe them as sets, not as topological spaces. That refinement comes later in the subsection.

4.2.2 Definition. *Let* $C^\infty(\mathbf{R}^n)$ *be the linear space of complex-valued, infinitely differentiable functions on* $\mathbf{R}^n$. *(I.e., if* $\varphi \in C^\infty(\mathbf{R}^n)$, *all partial derivatives* $D^\alpha \varphi$ *exist, and* φ *and all its partial derivatives are continuous).* ∎

Warning. The functions in $C^\infty(\mathbf{R}^n)$ may have complex *values*, but they are usually not functions of a complex *argument*. No analyticity is implied. For example, $\varphi(x, y) = x^2 + iy^2$ is a perfectly good $C^\infty(\mathbf{R}^2)$ function, but its real and imaginary parts do not satisfy the Cauchy-Riemann equations. ∎

4.2.3 Examples

A. The functions e^x, $\sin x$, and $\exp(-x^2)$ all belong to $C^\infty(\mathbf{R})$. The function

$$\exp(-x_1^2 - x_2^2)$$

belongs to $C^\infty(\mathbf{R}^2)$. The function $\exp|x|$ is not differentiable at $x = 0$ and does not belong to $C^\infty(\mathbf{R})$ (or $C^\infty(\mathbf{R}^n)$, if you think of x as a vector in $\mathbf{R}^n$).

B. The test functions in $C^\infty(\mathbf{R}^n)$ are easy to understand, but are a little too general. For instance, the function

$$(e^{x^2})^{\exp x^2}$$

is in $C^\infty(\mathbf{R})$, but it grows at such a rate at $\pm\infty$ that integrals like (4.1.11) cannot possibly converge. If we admit this function as a test function, we will not be able to define PV $\frac{1}{x}$ and $(x + i0)^{-1}$.

C. It is possible for a function φ to have partial derivatives of all orders and still not be continuous. An example is given in (4.2.21) below, after some preliminary material has been explained. ∎

4.2.4 Definition. *$\mathcal{S}(\mathbf{R}^n)$ is the linear space of $\varphi \in C^\infty(\mathbf{R}^n)$ satisfying*

$$\lim_{|x|\to\infty} |x|^k |D^\alpha \varphi(x)| = 0 \tag{4.2.5}$$

for all $k = 0, 1, 2, \ldots$ and all α. $\mathcal{S}(\mathbf{R}^n)$ is called the Schwartz space.[31] *One often expresses this definition as follows: a Schwartz function and all its derivatives decrease faster than any power of $|x|$.* ∎

4.2.6 Examples

A. The function $\sin x$ is not in $\mathcal{S}(\mathbf{R})$, since

$$\lim_{|x|\to\infty} |\sin x| \neq 0.$$

So, (4.2.5) fails for $k = 0$ and $\alpha = (0, \ldots, 0)$. Note that for φ to belong to a Schwartz space $\mathcal{S}(\mathbf{R}^n)$, *all* the limits in (4.2.5) must be zero. If just one of them is not zero, the function does not belong to the Schwartz space.

B. The function $\varphi(x) = \exp(-x^2)$ belongs to $\mathcal{S}(\mathbf{R})$. Let us prove this. I claim, first of all, that

$$\lim_{|x|\to\infty} |x|^k e^{-x^2} = 0 \tag{4.2.7}$$

for all $k = 0, 1, 2, \ldots$. Let us take (4.2.7) for granted. Next, I claim that

$$D^\alpha \exp(-x^2) \text{ has the form } p_\alpha(x) \exp(-x^2), \tag{4.2.8}$$

where p_α is a polynomial in x. (Since we are working in $\mathbf{R}$, α is just a non-negative integer). This is pretty obvious, after you take the first couple of derivatives of $\exp(-x^2)$. I just want to mention that the officially sanctioned proof method is *mathematical induction*. Here is a sketch.

First, the assertion is true for $\alpha = 0$; the zeroth derivative is just the function itself, so the polynomial p_0 is the constant 1.

Now, suppose that $D^k \exp(-x^2) = p_k(x) \exp(-x^2)$ with some polynomial p_k. Then differentiate again and use the chain rule to get

$$D^{k+1} \exp(-x^2) = [p_k'(x) - 2x p_k(x)] \exp(-x^2) \stackrel{\text{def}}{=} p_{k+1}(x) \exp(-x^2).$$

where p_{k+1} is the polynomial $p_k'(x) - 2x p_k(x)$.

Results (4.2.7) and (4.2.8) together prove that all limits (4.2.5) are zero, and hence that the function $\exp(-x^2)$ belongs to $\mathcal{S}(\mathbf{R})$. ∎

• **4.2.9 Exercise.** Prove (4.2.7). ∎

• **4.2.10 Exercise.** Prove that the function $\exp(-\|x\|_2^2)$ belongs to $\mathcal{S}(\mathbf{R}^n)$. ∎

• **4.2.11 Exercise.** Prove that sech x belongs to $\mathcal{S}(\mathbf{R})$. ∎

The last test function space to be introduced is extremely useful, but also requires some new ideas.

---

[31] After Laurent Schwartz, b. 1915, one of the creators of the mathematical theory of distributions. See EDM II, Article 125, for more history.

4.2.12 Definition. *Let φ be a continuous function on $\mathbf{R}^n$. Define the* ***support*** *of φ to be the set*

$$\operatorname{supp}\varphi = \overline{\{x \in \mathbf{R}^n \mid \varphi(x) \neq 0\}}. \blacksquare$$

The overbar indicates *closure* in the usual topology of $\mathbf{R}^n$. By definition, $\operatorname{supp}\varphi$ is a closed subset of $\mathbf{R}^n$ (it might be all of $\mathbf{R}^n$). It may be thought of as the *smallest* closed set outside which φ vanishes.

• **4.2.13 Exercise.** The complement of $\operatorname{supp}\varphi$ is an open set on which φ is identically zero. Prove that it is the *largest* open set on which φ is zero. ∎

4.2.14 Examples

A. Let $\varphi(x) = \sin x$ on $\mathbf{R}$. There is no open set on which sin vanishes, which implies that $\operatorname{supp}\sin = \mathbf{R}$.

B. Define

$$\varphi(x) = \begin{cases} 1+x, & \text{if } -1 \le x \le 0, \\ 1-x, & \text{if } 0 \le x \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(φ is *not* infinitely differentiable.) The support of φ is $[-1,1]$. Note that this is a closed set. The support is *not* the set $(-1,1)$ on which φ is nonzero, but rather the *closure* of that set. ∎

4.2.15 Definition. *We say that the function φ on $\mathbf{R}^n$ has* ***compact support*** *if $\operatorname{supp}\varphi$ is contained in some ball $\{x \mid |x| \le R\}$.*[32] *If φ has compact support, we also say that it is* ***compactly supported.*** ∎

4.2.16 Definition. *The linear space of compactly supported functions $\varphi \in C^\infty(\mathbf{R}^n)$ is denoted by $C_0^\infty(\mathbf{R}^n)$. (Read: "C zero infinity.")*[33] ∎

Terminology. *An infinitely differentiable function is also called* ***smooth****. At other times, I say "φ is C^∞"; such a function may or may not be in Schwartz space or in $C_0^\infty(\mathbf{R}^n)$.* ∎

Unless you have studied this subject already, you will probably not be able to write the formula for a single function in $C_0^\infty(\mathbf{R})$. If you think about it, you will realize that all the differentiable functions you know well are secretly analytic functions of a complex variable: sin, cos, exp, and so on. But when an analytic

---

[32] A subset of $\mathbf{R}^n$ is called *compact* if it is closed and bounded, i.e. contained in some ball. We study compactness in the next chapter.

[33] One sometimes meets other notations for this space: $C_c^\infty(\mathbf{R}^n)$, with subscript c standing for *compact*, or $\mathcal{D}(\mathbf{R}^n)$, with $\mathcal{D}$ standing for who knows what.

function vanishes on an interval, it vanishes identically. Therefore, infinitely differentiable functions of compact support can *never* be analytic (or rather, be the restrictions of analytic functions to the real axis). The next few pages deal with the construction of those strange functions.

We begin with the most important example. Let

$$\varphi(x) = \begin{cases} e^{-1/x^2}, & x \neq 0 \\ 0, & x = 0. \end{cases} \tag{4.2.17}$$

Note that φ is continuous at $x = 0$. Also, $\varphi'(x)$ clearly exists when $x \neq 0$. At $x = 0$,

$$\lim_{x \to 0} \frac{e^{-1/x^2} - 0}{x} = 0$$

by application of l'Hospital's rule. It is therefore natural to extend φ' to $x = 0$ by defining $\varphi'(0) = 0$. The resulting function φ' is then continuous on all of $\mathbf{R}$. One has, more generally:

4.2.18 Proposition. *For each k,*

$$\lim_{x \to 0} \varphi^{(k)}(x) = 0. \blacksquare$$

Proof. By induction. We have seen that the statement is true for $k = 1$. Now, I claim that $\varphi^{(k)}(x)$ has the form

$$p_n(\frac{1}{x})e^{-1/x^2},$$

where p_n is a polynomial. This is again proved by induction:

$$\text{assume that } \varphi^{(k)}(x) = p_k(\frac{1}{x})e^{-1/x^2};$$

then

$$\varphi^{(k+1)}(x) = -\frac{1}{x^2}p_k'(\frac{1}{x})e^{-1/x^2} + \frac{2}{x^3}p_k(\frac{1}{x})e^{-1/x^2}$$

which has the form $p_{k+1}(\frac{1}{x})e^{-1/x^2}$.

But now, as in (4.2.7),

$$\lim_{x \to 0} (\frac{1}{x})^m e^{-1/x^2} = 0$$

for all $m \geq 0$ (just replace x by $1/x$ in (4.2.7)). Hence $\varphi^{(k)}(x) \to 0$ as $x \to 0$, for all k. We *define* each derivative to be 0 at 0; with this definition, φ has continuous derivatives of all orders. $\blacksquare$

This function e^{-1/x^2} cannot be expanded in a Taylor series about $x = 0$:

$$\sum_{n=0}^{\infty} \frac{1}{n!}\varphi^{(n)}(0)x^n \equiv 0, \text{ since } f^{(n)}(0) = 0 \text{ for all } n.$$

Complex function theory sheds more light on the situation. $z = 0$ is an essential singular point of e^{-1/z^2}; therefore, there is no Taylor series about $z = 0$. It is somewhat accidental that the function is so well behaved for real z. If $z = iy$ is pure imaginary, then the function is e^{1/y^2}, which becomes infinite at a horrendous rate as $y \to 0$.

The function in (4.2.17) belongs to $C^\infty(\mathrm{R})$, but not to $\mathcal{S}(\mathrm{R})$ or $C_0^\infty(\mathrm{R})$. I now show how to build a $C_0^\infty(\mathrm{R})$ function.

4.2.19 Proposition. *Whenever $a < b$, there is a smooth function f satisfying*

$$f(x) = \begin{cases} 0, & x \le a, \\ \in (0,1), & a < x < b, \\ 0, & x \ge b. \end{cases}$$

For obvious reasons, such a function is called a ***bump function.*** ∎

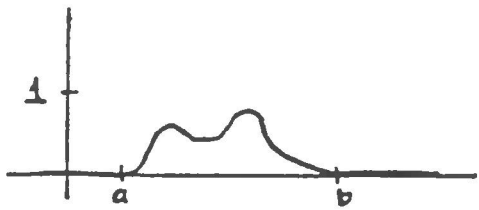


Figure 61. *A bump function*

Proof. There are many bump functions; I give an explicit formula for just one of them. Begin by defining $\varphi(x)$ for $a < x < b$ by

$$\varphi(x) = e^{-1/(x-a)(b-x)}.$$

Note that the exponent, $-1/(a-x)(b-x)$, is negative for $a < x < b$, so $0 < |f(x)| < 1$ there.

To see what happens at $x = a$ or $x = b$, it is best to write the exponent as

$$-\frac{1}{(x-a)(b-x)} = \frac{1/(a-b)}{x-a} + \frac{1/(b-a)}{x-b}$$

so that

$$\varphi(x) = \exp \frac{1/(a-b)}{x-a} \exp \frac{1/(b-a)}{x-b}.$$

As $x \downarrow a$, the first factor (and all its derivatives) become zero; the proof is just as in (4.2.18). The second factor remains finite, so as $x \downarrow a$, φ and all its derivatives tend to zero.

Likewise, all derivatives of φ have limit 0 as $x \uparrow b$.

We now extend the definition of φ to $x < a$ and $x > b$:

$$\varphi(x) = \begin{cases} 0, & \text{if } x \le a, \\ e^{-1/(x-a)(b-x)} & \text{if } a < x < b, \\ 0, & \text{if } x \ge b. \end{cases}$$

This function is smooth (C^∞) for $x < a$. $a < x < b$, and $x > b$. Let us look at the behavior at the point $x = a$. For $x < a$, $\varphi(x) \equiv 0$, and so $\varphi^{(k)}(x) \equiv 0$ for all $k \ge 0$. In particular,

$$\lim_{x \uparrow a} \varphi^{(k)}(x) = 0.$$

We have already argued that

$$\lim_{x \downarrow a} \varphi^{(k)}(x) = 0.$$

Hence, we extend $\varphi^{(k)}(x)$ to $x = a$ by defining $\varphi^{(k)}(a) = 0$. The function φ therefore has derivatives of all order at $x = a$. The point $x = b$ is dealt with in the same manner.

Therefore, $\varphi \in C^\infty(\mathbf{R})$. More is true: since $\operatorname{supp} \varphi = [a, b]$, $\varphi \in C_0^\infty(\mathbf{R})$, and we have *the first example of an infinitely differentiable function with compact support.* ∎

4.2.20 Examples

A. Once you have *one* $C_0^\infty(\mathbf{R}^n)$ function φ, as constructed in the proposition, you can build a huge supply of other such functions by the following simple device. Let ψ be any smooth function, not necessarily compactly supported (e.g., $\psi(x) = x^{49}$). Then $\psi\varphi$ is smooth and has compact support.

B. Indiscriminate use of bump functions can produce truly bizarre results. For example, take a harmless function like $\sin x$, and put smooth height 1 bumps all over it (Figure 62). Moral: smooth functions need not be "nice".

C. One is sometimes interested in C^∞ functions defined not on all of $\mathbf{R}^n$ but on some open subset $D \subset \mathbf{R}^n$. That linear space of functions is denoted by $C^\infty(D)$. If $\operatorname{supp} \varphi$ is a (closed) bounded subset of the open set D, one says that φ belongs to $C_0^\infty(D)$. For instance, if $D = (-1.1)$, the function $\varphi = 1$ is in $C^\infty(D)$, while the bump function in Figure 63 is in $C_0^\infty(D)$. ∎

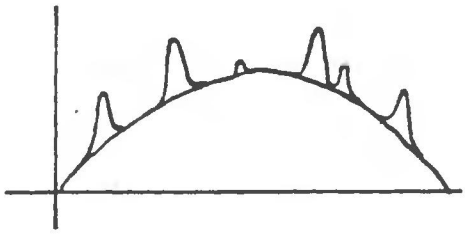


Figure 62. *A bizarre bumpy sine*

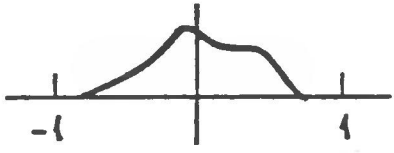


Figure 63. *A function in* $C_0^\infty((-1,1))$

4.2.21 Example. In Example 4.2.3 C., it was asserted that a discontinuous function could have partial derivatives of all orders. The construction of such a function relies on the trick used to prove that (4.2.17) is infinitely differentiable; we can now outline the argument.

Define

$$u(x,y) = \begin{cases} 0, & \text{if } x = y = 0, \\ \Re \exp(-1/z^4), & \text{otherwise,} \end{cases} \tag{4.2.22}$$

where $z = x + iy$ and $\Re$ denotes the real part of a complex-valued function. Let us now check existence and continuity of the partial derivatives. Set $x = 0$; then

$$u(0,y) = \begin{cases} 0, & \text{if } y = 0, \\ \exp(-1/y^4), & \text{otherwise.} \end{cases} \tag{4.2.23}$$

When $y = 0$, then

$$u(x,0) = \begin{cases} 0, & \text{if } x = 0, \\ \exp(-1/x^4), & \text{otherwise.} \end{cases} \tag{4.2.24}$$

By the kind of argument used on the function (4.2.17), one finds that the two functions (4.2.23) and (4.2.24) are infinitely differentiable. It follows that the partials

$$\frac{\partial^k u(x,y)}{\partial x^k}, \quad \frac{\partial^k u(x,y)}{\partial y^k}$$

all exist at $(x, y) = (0, 0)$ (which is the only questionable point). With more pain, but no new ideas, one checks existence of all the mixed partials as well.

On the other hand, the function u is really very badly behaved at the origin. Obviously, nothing ugly happens when z tends to 0 along the axes, but set $z = r\exp(i\pi/4)$: $z^4 = -r^4$, and $u = \exp(1/r^4)$, which tends to infinity.

This example shows that mere *existence* of partial derivatives, without some continuity requirement, is not enough to define a space of "smooth" functions of several variables. ∎

• **4.2.25 Exercise.** Prove that if $\varphi_1, \varphi_2 \in C_0^\infty(\mathbf{R}^n)$, then $\varphi_1 + \varphi_2 \in C_0^\infty(\mathbf{R}^n)$. Prove that $C_0^\infty(\mathbf{R}^n)$ is a linear space. ∎

• **4.2.26 Exercise.** Show that $C_0^\infty(\mathbf{R}^n) \subset \mathcal{S}(\mathbf{R}^n) \subset C^\infty(\mathbf{R}^n)$. ∎

• **4.2.27 Exercise.** Show that $\sin x$ and $e^{-1/x^2} + \sin x$ have the same Taylor series. ∎

• **4.2.28 Exercise.** Define

$$\varphi(x) = \begin{cases} e^{1/x}, & \text{if } x \le 0, \\ 0, & \text{if } x > 0. \end{cases}$$

Interpret this as a $C^\infty(\mathbf{R})$ function. Note that $\operatorname{supp}\varphi = (-\infty, 0]$. ∎

• **4.2.29 Exercise.** For the function φ in Example 4.2.28, prove in detail that

$$\lim_{x \downarrow a} \varphi^{(k)}(x) = 0. \quad ∎$$

4.2.30 Exercise. Sometimes a refinement of the construction in Proposition 4.2.19 is useful. Let $0 < r < R$. Find a C^∞ function φ which is

$$\begin{cases} \equiv 1, & \text{if } |x| < r, \\ \in [0, 1], & \text{if } r < |x| < R, \\ \equiv 0, & \text{if } R < |x|. \end{cases}$$

(Hint: let

$$\psi_1(x) = \begin{cases} e^{-1/(x-r)}e^{-1/(R-x)}, & \text{if } r < x < R, \\ 0. & \text{otherwise.} \end{cases}$$

Then set

$$\psi_2(x) = \left(\int_x^R \psi_1(t)\,dt\right)\left(\int_r^R \psi_1(t)\,dt\right)^{-1}.$$

Show that ψ_2 is smooth, is $\equiv 1$ for $x < r$ and $\equiv 0$ for $x > R$. Finally, define

$$\varphi(x) = \psi_2(|x|).$$

Check that φ has the required properties. What is $\operatorname{supp}\varphi$?) ∎

4.2.31 Exercise. Refer to the preceding exercise. Let ψ_2 be the function defined there, and set

$$\varphi(x_1, \ldots, x_n) = \psi_2(\sqrt{x_1^2 + \ldots + x_n^2}).$$

Show that $\varphi \in C_0^\infty(\mathbf{R}^n)$, and sketch its graph for $n = 2$. ∎

4.2.32 Exercise. Modify the construction in Exercise 4.2.30 to produce a $C_0^\infty(\mathbf{R})$ function which is $\equiv 1$ for $a < x < b$, takes values in $[0,1]$ for $a - \epsilon \le x \le a$ and $b \le x \le b + \epsilon$, and $\equiv 0$ for $x < a - \epsilon$ and $x > b + \epsilon$. Here a, b and ϵ (not necessarily small) are fixed. ∎

• **4.2.33 Exercise.** Let φ be the "bump" function constructed in Proposition 4.2.19. Define

$$\Phi(x) = \left(\int_a^b \varphi(t)\, dt \right)^{-1} \int_{-\infty}^{x} \varphi(t)\, dt.$$

Sketch Φ. Show that it is smooth. ∎

4.2.34 Exercise. In various problems of physical origin, one wants to construct a *smooth transition layer between two constant states*. This is a function $\varphi \in C^\infty(\mathbf{R})$ satisfying:

$$\varphi(x) = \begin{cases} \alpha, & \text{if } x \le a - \epsilon, \\ \in [\alpha, \beta], & \text{if } a - \epsilon \le x \le a + \epsilon, \\ \beta, & \text{if } x \ge a + \epsilon. \end{cases}$$

See Figure 64 for a sketch. Find such a function. (Hint: use the function Φ from the preceding exercise.) ∎

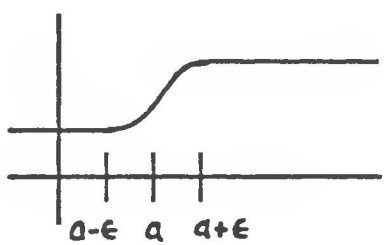


Figure 64. ***A smooth transition layer***

4.2.35 Exercise. Construct a $C^\infty(\mathbf{R})$ function φ with the following properties:

$$\varphi(x) = \begin{cases} x, & \text{if } x \le -1, \\ \in [-1, 1], & \text{if } -1 \le x \le 1, \\ 1, & \text{if } x = 0, \\ x, & \text{if } x \ge 1. \end{cases}$$ ∎

4.2.36 Exercise. Construct a $C_0^\infty(\mathbf{R})$ function φ with $\varphi^{(n)}(0) = 0$ for $n = 0, \ldots, 100$, but $\varphi^{(101)}(0) \neq 0$. ∎

• **4.2.37 Exercise.** Give an example of a complex-valued $C_0^\infty(\mathbf{R})$ function. ∎

We have now introduced certain spaces of test functions, and have seen some (not so elementary) examples of infinitely differentiable functions with compact support. For the moment, let $\mathcal{X}$ denote one of the test function spaces. A

distribution will be defined to be a *function on* $\mathcal{X}$. In words: a distribution associates a number to each test function. This assignment cannot be completely arbitrary if a useful theory is to result. The space $\mathcal{X}$ is a linear space, and the distribution should be a *linear function*. More than that, the distribution should be a *continuous* function on $\mathcal{X}$. Otherwise, one might encounter situations in which almost identical functions belonging to $\mathcal{X}$ are sent to very different numbers by a certain distribution (Figure 65).

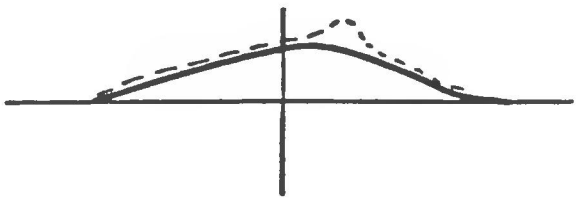


Figure 65. *Close, but not close enough*

In order to talk about continuity, we need a topology on $\mathcal{X}$. The standard topologies will be defined and discussed in the remainder of this subsection.

The topologies to be described are probably the most complicated and least easy to picture of all the ones presented in this chapter. Considerable functional analysis technology can be brought to bear on the theory of distributions, and it is not possible (or desirable, at this time) to set down all the official language. When things seem to be getting very complicated, just keep a few obvious points in mind:

Functions in $C^\infty(\mathbf{R}^n)$ have infinitely many derivatives on $\mathbf{R}^n$, and the topology should take account of their size. Functions in $C_0^\infty(\mathbf{R}^n)$ have infinitely many derivatives, and the topology should take note of that, but it also should use the compactness of the support. The functions in $\mathcal{S}(\mathbf{R}^n)$ have infinitely many derivatives, which tend to zero faster than any power of $|x|$. The topology again should measure the size of the derivatives, but this time in relation to the size of $|x|^k$. In each case, the key items in the definition of the test function space will be incorporated into the definition of the topology.

We begin with the topology on $C^\infty(\mathbf{R})$. It was already introduced, without any of the special symbols, in Example 3.1.19B. There, we used the language of local bases; below, the topology is introduced via a subbase. The difference is miniscule.

4.2.38 Definition. *Let $\varphi \in C^\infty(\mathbf{R})$. For each pair of integers $k \geq 0, m > 0$, set*

$$p_{m,k}(\varphi) \stackrel{\text{def}}{=} \max_{0 \leq j \leq k} \sup_{-m \leq x \leq m} |\varphi^{(j)}(x)|. \blacksquare$$

The quantity $p_{m,k}$ is a measure of size of the function φ. It looks at φ on the interval $[-m, m]$, and measures the largest value amongst the first k derivatives of φ on this interval. This is always a finite number, but it is certainly possible that $p_{m,k}(\varphi) \to \infty$ as $m \to \infty$ (for example, if $\varphi(x) = \exp x$). The quantity $p_{m,k}$ has the following three properties:

$$\begin{aligned} &p_{m,k}(\varphi) \geq 0, \\ &p_{m,k}(\alpha\varphi) = |\alpha| p_{m,k}(\varphi), \\ &p_{m,k}(\varphi_1 + \varphi_2) \leq p_{m,k}(\varphi_1) + p_{m,k}(\varphi_2). \end{aligned} \tag{4.2.39}$$

It is *not* true that $p_{m,k}(\varphi) = 0 \Leftrightarrow \varphi \equiv 0$. For example, if φ has the shape drawn in Figure 66, then $\varphi^{(j)}(x) = 0$ for $|x| \leq m$ and $j \geq 0$.

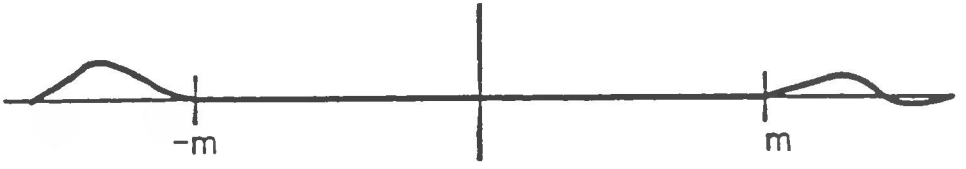


Figure 66. $p_{m,k}(\varphi) = 0$ *for all* k

For this φ,

$$p_{m,k}(\varphi) = 0 \quad (k \geq 1),$$

yet φ is not the zero function. But for the fact that

$$p_{m,k}(\varphi) \text{ can be } 0 \text{ even if } \varphi \neq 0, \tag{4.2.40}$$

$p_{m,k}$ would be a norm.

4.2.41 Definition. *A real-valued function p on a real or complex linear space V is called a* ***seminorm*** *if*

$$\begin{aligned} &i)\ p(v) \geq 0, \ \textit{all}\ v \in V \\ &ii)\ p(\alpha v) = |\alpha| p(v), \ v \in V, \ \alpha \in \mathbf{R}(\textit{resp. } \mathbf{C}) \\ &iii)\ p(v + w) \leq p(v) + p(w). \end{aligned}$$

(It is not required that $p(v) = 0 \Leftrightarrow v = 0$). ∎

4.2.42 Definition. *A subbase for the topology on $C^\infty(\mathbf{R})$ is given by the sets*

$$B_{m,k,\epsilon}(\varphi) \stackrel{\text{def}}{=} \{\psi \mid p_{m,k}(\varphi - \psi) < \epsilon\}$$

where m, k range over $\{0, 1, 2, \ldots\}$, $\epsilon > 0$ and φ ranges over C^∞. The topology is defined by the subbase in the usual way (see (3.4.10)). ∎

So, when are two smooth functions on $\mathbf{R}$ "close" to each other in the $C^\infty(\mathbf{R})$ topology? When they agree, up to their k^{th} derivatives, on a large interval $[-m, m]$. They are "closer" when more derivatives agree on the same interval, or on a larger one.

• **4.2.43 Exercise.** Show that $p_{m,k}(\varphi) \geq p_{m',k'}(\varphi)$ if $m \geq m'$ and $k \geq k'$. ∎

• **4.2.44 Exercise.** Draw two functions $\varphi_1, \varphi_2 \in C^\infty(\mathbf{R})$ for which $p_{1,k}(\varphi_1) = p_{1,k}(\varphi_2)$, but $p_{2,k}(\varphi_1) > p_{2,k}(\varphi_2)$. ∎

4.2.45 Proposition. *$C^\infty(\mathbf{R})$ is a metric space, with the metric*

$$d(\varphi_1, \varphi_2) = \sum_{m=1}^{\infty} \frac{p_{m,m}(\varphi_1 - \varphi_2)}{1 + p_{m,m}(\varphi_1 - \varphi_2)} \times \frac{1}{2^m}.$$

This metric defines the same topology as (4.2.42). ∎

The proof will not be given, because this metric is hardly ever used. For any application, the definition of the topology in (4.2.42) is much more convenient. It is, however, worth noting that *all* derivatives of two $C^\infty(\mathbf{R})$ functions can be compared over *all* of $\mathbf{R}$ by the single measure of distance (4.2.45). Moreover, observe that the metric d is fixed up so that it yields a finite value even if the individual seminorms $p_{m,m}$ tend to ∞.

4.2.46 Remark. The topology on $C^\infty(\mathbf{R}^n)$ is got in a similar way. Introduce the following seminorms:

$$p_{m,k}(\varphi) = \max_{j_1+\cdots+j_n \leq k} \sup_{\|x\|_2 \leq m} \left| \frac{\partial^{j_1+\cdots+j_n}}{\partial x_1^{j_1} \ldots \partial x_n^{j_n}} \varphi(x) \right|.$$

(Rewrite this using Notation 4.2.1). Define the basis for the topology as in (4.2.42). ∎

In the theory of distributions, the topology on test function spaces is used, first of all, to define *convergence of a sequence of test functions*. Definitions 4.2.42 and 4.2.46 have the following consequence.

4.2.47 Corollary. *A sequence $\{\varphi_m\}$ of functions in $C^\infty(\mathbf{R}^n)$ converges to φ if, and only if, for every $\epsilon > 0$, $K \in \mathbf{N}$, and $k \in \mathbf{N}$, there is an $M_{\epsilon,K,k}$ such that*

$$\sup_{|x| \leq K, |\alpha| \leq k} |D^\alpha(\varphi_m(x) - \varphi(x))| < \epsilon$$

whenever $m > M_{\epsilon,K,k}$. ∎

• **4.2.48 Exercise.** Prove this Corollary. ∎

In words (and this should make perfect sense) the sequence $\{\varphi_m\}$ converges to φ if, on any closed ball $\{x \mid \|x\|_2 \leq K\}$, all partial derivatives of $\varphi_m - \varphi$, up to any specified order, can be made as close to zero as desired.

Let us now think about a distribution on $C^\infty(\mathbf{R}^n)$, i.e., a linear function T from $C^\infty(\mathbf{R}^n)$ to the complex numbers. This T will be continuous if

$$T(\varphi_m) \to T(\varphi) \quad \text{whenever} \quad \varphi_m \to \varphi.$$

Recall from Section 3.2 that continuity cannot in general be defined in terms of convergent sequences. By Proposition 4.2.45, however, $C^\infty(\mathbf{R}^n)$ is a metric space, and in this case Proposition 3.2.11 guarantees that the usual definition of continuity, (3.2.1), will be satisfied.

Because the functions in $C^\infty(\mathbf{R}^n)$ are allowed to grow very rapidly at infinity, not all the distributions defined in Subsection 4.1 are well defined.

4.2.49 Examples

A. Define δ on $C^\infty(\mathbf{R})$ by the standard formula $\delta(\varphi) = \varphi(0)$. We have

$$|\delta(\varphi_m) - \delta(\varphi)| = |\varphi_m(0) - \varphi(0)|. \tag{4.2.50}$$

Assume that $\varphi_m \to \varphi$ in the sense of Corollary 4.2.47. Then, for a given ϵ, the right-hand side in

$$|\varphi_m(0) - \varphi(0)| \leq \sup_{|x|\leq 1,\, \alpha \leq 0} |D^\alpha(\varphi_m(x) - \varphi(x))|$$

can be made less than ϵ. This shows that the Dirac delta is continuous on $C^\infty(\mathbf{R})$: it is really a distribution, according to the Pseudo-Definition 4.1.28.

B. The Cauchy Principal Value $\mathrm{PV}1/x$ does not even make sense for the "typical" $C^\infty(\mathbf{R})$ function. Take

$$\varphi(x) = e^{x^2} - 1.$$

There is no problem with convergence at $x = 0$ in the integral

$$\int_{-\infty}^{\infty} \frac{\varphi(x)}{x}\, dx,$$

but the integral won't converge due to the growth of the integrand at $\pm\infty$. ∎

The topology on $C_0^\infty(\mathbf{R}^n)$ is more complicated. This is only to be expected: it has to keep track of all the derivatives, just like the $C^\infty(\mathbf{R}^n)$ topology, but in addition it must worry about the compact support restriction. For most applications to distribution theory, it suffices to define what is meant by "convergence of a sequence of functions in $C_0^\infty(\mathbf{R}^n)$." It is (fortunately) not necessary to describe open sets, neighborhoods, bases, and so on.

4.2.51 Definition. *A sequence* $\{\varphi_j\} \subset C_0^\infty(\mathbf{R}^n)$ *converges to* 0 *(the zero function) if*

a) *there is an* R *such that* $\varphi_j(x) = 0$ *when* $|x| > R$, *for all* j,

and

b) *for each* k, *the seminorms* $p_{R,k}(\varphi_j)$ *appproach* 0 *as* $j \to \infty$.

One says that $\{\varphi_j\}$ *converges to* φ *if* $\varphi_j - \varphi$ *converges to* 0. ∎

The new ingredient is a). It requires that *all* functions φ_j in a convergent sequence have support within a single (possibly large) set, $|x| < R$. So, the sequence shown in Figure 67 could not converge to zero in $C_0^\infty(\mathbf{R})$, even though it does converge to zero in $C^\infty(\mathbf{R})$. The topology on $C_0^\infty(\mathbf{R}^n)$ is therefore not just the "restriction" of the topology on $C^\infty(\mathbf{R})$.

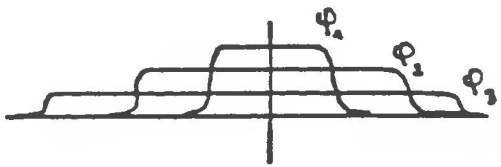


Figure 67. *How not to converge to zero in* $C_0^\infty(\mathbf{R})$

4.2.52 Examples

A. A simple example will illustrate the need for condition a) in Definition 4.2.51. Choose $\varphi_j \in C_0^\infty(\mathbf{R})$, satisfying (see Exercise 4.2.30)

$$\varphi_j(x) = \begin{cases} 1/2j, & \text{if } |x| \le j, \\ \in [0, 1/2j], & \text{if } -j - 1/j \le x \le -j, \\ \in [0, 1/2j], & \text{if } j \le x \le j + 1/j, \\ 0, & \text{otherwise.} \end{cases}$$

(Sketch the graph!) You might think that the definitions of convergence in $C_0^\infty(\mathbf{R})$ would only confirm what seems obvious from the graph, that this sequence tends to 0. To see why Definition 4.2.51 does *not* agree that $\varphi_j \to 0$, consider a (potential) distribution T defined on test functions $\varphi \in C_0^\infty(\mathbf{R})$ by

$$T(\varphi) = \int_{-\infty}^{\infty} \varphi(x)\, dx.$$

We have

$$T(\varphi_j) \geq \int_{-j}^{j} \frac{1}{2j}\, dx = 1.$$

On the other hand, $T(0) = 0$. A definition of convergence that says $\varphi_j \to 0$ would then force the conclusion that

$$\lim_{j\to\infty} T(\varphi_j) \neq T(\lim_{j\to\infty} \varphi_j).$$

In other words, this T would not be a continuous function on $C_0^\infty(\mathrm{R})$. There are many other distributions that would suddenly turn out to be discontinuous, and the usefulness of the theory would be diminished.

B. Let $\varphi \in C_0^\infty(\mathrm{R})$, and define

$$\varphi_j(x) = \frac{1}{j}\varphi(\frac{x}{j}).$$

This sequence cannot converge in $C_0^\infty(\mathrm{R})$. For, suppose that $\varphi(a) \neq 0$ for some $a \neq 0$. Then $\varphi_j(ja) \neq 0$, and condition a) of (4.2.51) is violated: the φ_j do not all vanish for $|x| >$ some R (independent of j).

C. Condition a) in 4.2.51 is the more unexpected requirement of the two, but one should not get the impression that condition b) is irrelevant. Consider a sequence of bump functions $\varphi_j \in C_0^\infty(\mathrm{R})$, $j \geq 1$, such that

$$\varphi_j(x) = \begin{cases} 1, & \text{if } |x| \leq 1, \\ \in [0,1], & \text{if } -1 - 1/j \leq x \leq -1, \\ \in [0,1], & \text{if } 1 \leq x \leq 1 + 1/j, \\ 0, & \text{otherwise.} \end{cases}$$

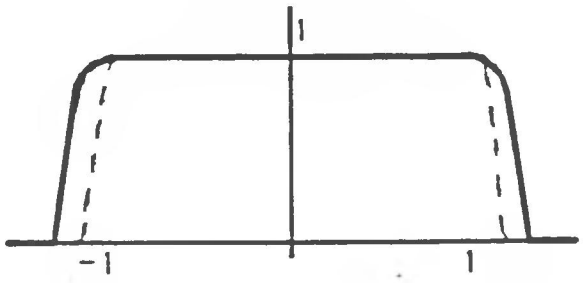


Figure 68. ***Bumps that don't converge in*** $C_0^\infty(\mathrm{R})$

Then $\varphi_j(x) = 0$ for $|x| \geq 2$, independent of j, so that condition a) in 4.2.51 is satisfied. What about condition b)? Suppose that $\varphi_j \to \varphi$ in the sense of Definition 4.2.51. Among other things, it must be true that

$$\sup_{|x|\leq 2} |\varphi_j(x) - \varphi(x)| = p_{2,0}(\varphi_j - \varphi) \to 0,$$

so that

$$\lim_{j\to\infty} |\varphi_j(x) - \varphi(x)| = 0$$

for every $x \in [-2, 2]$. The purported limit function φ is therefore discontinuous: it must be 1 on $[-1, 1]$ and 0 on $[-2, -1) \cup (1, 2]$. Such a function cannot belong to $C_0^\infty(\mathbf{R})$. Hence, the sequence φ_j does not converge in $C_0^\infty(\mathbf{R})$. ∎

The topology on $C_0^\infty(\mathbf{R}^n)$ can be defined by a collection of seminorms; see (4.2.42) for the corresponding definition in $C^\infty(\mathbf{R}^n)$. The following is reproduced from EDM II, Article 125 G.

4.2.53 Definition. *For every nondecreasing sequence* $\mathbf{a} = \{a_0, a_1, \ldots\}$ *of positive numbers, and every nondecreasing sequence* $\mathbf{k} = \{k_0, k_1, \ldots\}$ *of nonnegative integers, set*

$$p_{\mathbf{a},\mathbf{k}}(\varphi) = \sup_{j\geq 0} \sup_{|\alpha|\leq k_j} \sup_{|x|>j} a_j |D^\alpha \varphi(x)|. \tag{4.2.54}$$

A subbase for the topology on $C_0^\infty(\mathbf{R}^n)$ *is then given by the sets*

$$\{\psi \mid p_{\mathbf{a},\mathbf{k}}(\varphi - \psi) < \epsilon\},$$

as φ *ranges over* $C_0^\infty(\mathbf{R}^n)$ *and* $\mathbf{a}$, $\mathbf{k}$ *range over the collection of sequences described.* ∎

This definition grows on you (like a wart) when you think about it long enough, but you can certainly understand why people prefer to concentrate on convergence, (4.2.51). Maybe the one point to pay attention to is how cleverly this definition incorporates compact support. When a function is *not* zero for $|x| >$ some j, then you can pick a sequence $\mathbf{a}$ in which a_j is huge; the corresponding seminorm $p_{\mathbf{a},\mathbf{k}}$ will then be large because of $\sup_{|x|>j} a_j|\varphi(x)|$. This, in the end, is what keeps sequences $\{\varphi_m\}$ from converging to zero unless all functions φ_m vanish outside a fixed bounded region.

We will not need to understand the topology on $C_0^\infty(\mathbf{R}^n)$ except through the definition of convergence. Contrary to popular opinion, mathematicians are not masochists, and (unless they are really excited about topology for its own sake) they use only those portions of theories that apply to their own research.

4.2.55 Example. It was pointed out in Example 4.2.3.B that $\mathrm{PV}(1/x)$ is not a distribution on $C^\infty(\mathbf{R})$ because the integral $\int_{|x|>\epsilon} \varphi(x)/x\,dx$ does not usually converge at infinity. The test functions in $C_0^\infty(\mathbf{R})$, however, are zero at infinity, so this problem has been hypothesized away. We have a right to expect that $\mathrm{PV}\,\frac{1}{x}$ is, indeed, a continuous function on $C_0^\infty(\mathbf{R})$. Let us prove this claim.

One can simplify notation by making use of linearity. If T is *any* linear function of φ, then

$$T(\varphi_m) \to T(\varphi) \quad \text{if} \quad \varphi_m \to \varphi$$

is the same as

$$T(\varphi_m - \varphi) \to 0 \quad \text{if} \quad (\varphi_m - \varphi) \to 0.$$

Hence, it is always sufficient to show that (replace $\varphi_m - \varphi$ by φ_m)

$$T(\varphi_m) \to 0 \quad \text{if} \quad \varphi_m \to 0.$$

Suppose now that $\varphi_m \to 0$ in $C_0^\infty(\mathbf{R})$, according to Definition 4.2.51. We want to show that the *numbers* $\mathrm{PV}\,\frac{1}{x}(\varphi_m)$ approach zero.

Use formula (4.1.27) for the principal value. The definition of convergence in $C_0^\infty(\mathbf{R})$ requires all φ_m to vanish for $|x| >$ some (integer) R. Then

$$\left| \int_1^\infty \frac{\varphi_m(x)}{x}\, dx \right| \le \int_1^R \left| \frac{\varphi_m(x)}{x} \right| dx \le \int_1^R |\varphi_m(x)|\, dx \le R \sup_{x \in \mathbf{R}} |\varphi_m(x)|. \tag{4.2.56}$$

A similar estimate holds for the $\int_{-\infty}^{-1}$ integral.

The integral

$$\int_{-1}^{1} \frac{\varphi_m(x) - \varphi_m(0)}{x}\, dx \tag{4.2.57}$$

is estimated by the Mean Value Theorem. Since

$$\frac{\varphi_m(x) - \varphi_m(0)}{x} = \varphi_m'(c_x),$$

the integral (4.2.57) is bounded by

$$\int_{-1}^{1} \sup_{y \in \mathbf{R}} |\varphi_m'(y)|\, dx = 2 \sup_{y \in \mathbf{R}} |\varphi_m'(y)|. \tag{4.2.58}$$

Combining (4.2.56) and (4.2.58), we find that

$$\left| \mathrm{PV}\, \frac{1}{x}(\varphi_m) \right| \le 2R \sup_{x \in \mathbf{R}} |\varphi_m(x)| + 2 \sup_{y \in \mathbf{R}} |\varphi_m'(y)|. \tag{4.2.59}$$

The expression (4.2.59) is $\le$

$$2R\, p_{R,0}(\varphi_m) + 2\, p_{R,1}(\varphi_m),$$

which tends to zero (by the definition of $C_0^\infty(\mathbf{R})$ convergence,(4.2.51)) when $\varphi_m \to 0$ in $C_0^\infty(\mathbf{R})$. ∎

Finally, we give a very brief description of the topology on $\mathcal{S}(\mathbf{R}^n)$. That space will be important later in the course, because it is particularly well suited for the Fourier theory of distributions. From the purely topological point of view, there are no new ideas.

4.2.60 Definition. *For $\varphi \in \mathcal{S}(\mathbf{R}^n)$, set*

$$q_{m,k}(\varphi) \stackrel{\text{def}}{=} \max_{|\alpha| \le k} \sup_{x \in \mathbf{R}^n} |x|^m |D^\alpha \varphi(x)|. \blacksquare$$

Recall (from (4.2.4)) that a Schwartz function and all its derivatives decrease faster than any power of $|x|$. *A fortiori*, all $|x|^m |D^\alpha \varphi(x)|$ are bounded in x, and the quantities $q_{m,k}$ measure the size of these expressions.

4.2.61 Definition. *A sequence φ_j is said to converge to φ in $\mathcal{S}(\mathbf{R}^n)$, if*

$$\lim_{j \to \infty} q_{m,k}(\varphi_j - \varphi) = 0$$

for all $m, k = 0, 1, \ldots$. $\blacksquare$

• **4.2.62 Exercise.** Let $\varphi \in \mathcal{S}(\mathbf{R})$, and define

$$\varphi_j(x) = \frac{1}{j} \varphi(\frac{x}{j}).$$

Does $\varphi_j \to 0$ according to Definition 4.2.60? $\blacksquare$

• **4.2.63 Exercise.** Let $\varphi \in \mathcal{S}(\mathbf{R})$. Set $\varphi_j = \varphi/j$. Does this sequence converge in $\mathcal{S}(\mathbf{R})$? In $C_0^\infty(\mathbf{R})$? Set $\varphi_j(x) = \varphi(jx)/j$. Does this sequence converge in $\mathcal{S}(\mathbf{R})$? $\blacksquare$

• **4.2.64 Exercise.** Let p be a polynomial of one variable, and let $\varphi \in \mathcal{S}(\mathbf{R})$. Show that $p\varphi \in \mathcal{S}(\mathbf{R})$. If $\varphi_j \to \varphi$ in the sense of Definition 4.2.61, what can one say about convergence of the sequence $p\varphi_j$? $\blacksquare$

• **4.2.65 Exercise.** Does the sequence φ_j in Example 4.2.52.B converge to 0 in $C^\infty(\mathbf{R})$?. $\blacksquare$

• **4.2.66 Exercise.** Consider a sequence of functions φ_j of the shape shown in Figure 69. If one decides that $\varphi_j \to 0$, what will be wrong with the sequence PV $\frac{1}{x}(\varphi_j)$ as $j \to \infty$? $\blacksquare$

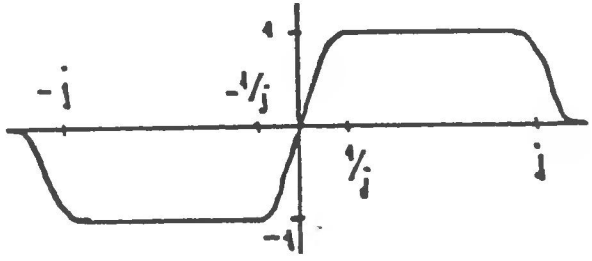


Figure 69.

• **4.2.67 Exercise.** Prove that PV $\frac{1}{x}$ is a continuous linear function on $\mathcal{S}(\mathbf{R})$. ∎

4.2.68 Exercise. Does the formula in Exercise 4.1.32 for PV $(1/x^2)$ define a continuous linear function on $C^\infty(\mathbf{R})$? On $C_0^\infty(\mathbf{R})$? On $\mathcal{S}(\mathbf{R})$? ∎

4.2.69 Exercise. For $\varphi \in \mathcal{S}(\mathbf{R})$, define

$$T(\varphi) = \sum_{n=-\infty}^{\infty} \varphi(n).$$

Is this a distribution (i.e., a continuous linear function)? (If it is a distribution, it would be written

$$\sum_{n=-\infty}^{\infty} \delta(x - n).)$$ ∎

4.2.70 Exercise*. Show that $(x+i0)^{-1}$, as defined in Supplement 4.1.36, is a continuous linear function on $C_0^\infty(\mathbf{R})$ and $\mathcal{S}(\mathbf{R})$. ∎

Subsection 4.3 Generalized functions

Before going further, I want to take stock of the abstract content of the discussion of distributions to this point. The *test function spaces* of Subsection 4.2 are linear spaces. They are equipped with rather complicated topologies; one knows that *none* of the topologies come from norms (that won't be proved), and I pointed out that at least the C^∞ topology comes from a metric. The very concrete distributions listed in Subsection 4.1 are linear functions on the test function spaces, and they are *continuous* with respect to the test function topologies. In short: I have illustrated Definition 3.2.9 yet again.

It is therefore all the more amazing that the abstract setup has led to concrete and practical calculations with singular functions and singular integrals. I now will explain the precise sense in which "distributions" generalize "functions". We must first recondition our response to the word *function*. The standard attitude is to think of a function $F : \mathbf{R} \to \mathbf{R}$ as a rule that associates a real number $F(x)$ to every real number x. True, but how does one generalize this idea and come up with delta functions and principal values?

For sake of simplicity, the discussion is confined to the test function space $C_0^\infty(\mathbf{R})$. Also for simplicity, only real-valued distributions are considered. The official definition of a distribution on $C_0^\infty(\mathbf{R})$ is no different from the Pseudo-Definition 4.1.28 — that was imprecise only because no topologies had been defined at that stage.

4.3.1 Definition. *A distribution on $C_0^\infty(\mathbf{R})$ is a continuous, linear, real-valued function on $C_0^\infty(\mathbf{R})$. The set of distributions on $C_0^\infty(\mathbf{R})$ is again a linear space, denoted by $C_0^\infty(\mathbf{R})'$.* ∎

Now let $F \in C^\infty(\mathbf{R})$ be a fixed function (yes, in $C^\infty(\mathbf{R})$, not necessarily in $C_0^\infty(\mathbf{R})$; the function F could grow very rapidly as $x \to \pm\infty$.) It defines a distribution T_F as follows:

$$T_F(\varphi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} F(x)\varphi(x)\,dx \tag{4.3.2}$$

for all $\varphi \in C_0^\infty(\mathbf{R})$. One must check: T_F is a linear function on $C_0^\infty(\mathbf{R})$, continuous in the $C_0^\infty(\mathbf{R})$ topology.

a) Linearity is trivial; it follows from linearity of the integral.

b) To establish continuity, we must show that $\varphi_j \to \varphi$ implies

$$\int_{-\infty}^{\infty} F(x)\varphi_j(x)\,dx \to \int_{-\infty}^{\infty} F(x)\varphi(x)\,dx. \tag{4.3.3}$$

Rewrite (4.3.3):

$$\int_{-\infty}^{\infty} F(x)\big(\varphi_j(x) - \varphi(x)\big)\,dx \to 0. \tag{4.3.4}$$

Recall, from Definition 4.2.51, that all the φ_j (and hence also φ) have support in some bounded set $|x| \le R$. Now, $\sup_{|x|\le R} |F(x)| \le K_R$ for some number K_R and

$$|\varphi_j(x) - \varphi(x)| \le p_{R,0}(\varphi_j - \varphi).$$

From these estimates, we deduce:

$$\left|\int_{-\infty}^{\infty} F(x)(\varphi_j(x) - \varphi(x))\,dx\right| \le \int_{-\infty}^{\infty} |F(x)|\,|\varphi_j(x) - \varphi(x)|\,dx$$

$$\le \int_{-R}^{R} K_R p_{R,0}(\varphi_j - \varphi)\,dx = 2RK_R\, p_{R,0}(\varphi_j - \varphi) \to 0.$$

So, every $C^\infty(\mathbf{R})$ function F defines a distribution T_F on $C_0^\infty(\mathbf{R})$. We now show — and this is the real point — that the numbers $T_F(\varphi)$, for every $\varphi \in C_0^\infty(\mathbf{R})$, completely determine the function F. In other words, a function $F \in C^\infty(\mathbf{R})$ can be specified in two ways:

1) Give the value $F(x)$ for every $x \in \mathbf{R}$, or
2) Give the value of $\int_{-\infty}^{\infty} F(x)\varphi(x)\,dx$ for every $\varphi \in C_0^\infty(\mathbf{R})$.

We are to determine the value $F(x_0)$, for any given $x_0 \in \mathbf{R}$, by using the known numbers

$$\int_{-\infty}^{\infty} F(x)\varphi(x)\,dx, \quad \varphi \in C_0^\infty(\mathbf{R}): \tag{4.3.5}$$

evidently, we should accomplish this by clever choices of φ. Take φ to be the heat equation solution

$$E(t, x - x_0) \stackrel{\text{def}}{=} \frac{1}{\sqrt{4\pi t}} e^{-(x-x_0)^2/4t}, \tag{4.3.6}$$

and pretend, for the moment, that it belongs to $C_0^\infty(\mathbf{R})$; this is of course false, but the idea becomes clearer. By hypothesis, we know the values of the integral

$$\int_{-\infty}^{\infty} F(x) E(t, x - x_0)\, dx \tag{4.3.7}$$

for every $t > 0$. Letting t decrease to 0, we now find the limit of (4.3.7): it is known to be $F(x_0)$ from Corollary 4.0.4. In this way, we recover the original function F.

But there is a gap in the proof: $E(t, x - x_0)$ is not $C_0^\infty(\mathbf{R})$, and F may not be bounded as required by Corollary 4.0.4. This is easy to fix, at the level of pictures. We just need to bend the function (4.3.6) down a little, smoothly (as we learned to do in Subsection 2), to make it have compact support:

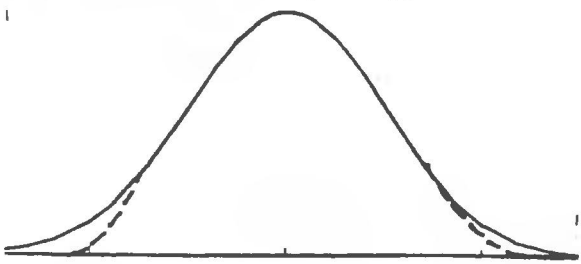


Figure 70. *Making $E(t, x - x_0)$ have compact support*

This takes care of both problems at once: the test function (4.3.6) is converted into a $C_0^\infty(\mathbf{R})$ function, which is good because we propose to work within the framework of the space $C_0^\infty(\mathbf{R})$, and the integral (4.3.7) converges no matter how fast F grows. To accomplish the feat, let ψ be a bump function (as in Exercise 4.2.30) with these properties:

$$\psi(x) = \begin{cases} 1, & \text{if } x_0 - 1 \le x \le x_0 + 1, \\ \in [0,1], & \text{if } x_0 - 2 \le x \le x_0 - 1, \\ \in [0,1], & \text{if } x_0 + 1 \le x \le x_0 + 2, \\ 0, & \text{otherwise.} \end{cases}$$

Now we show that the integrals

$$\int_{-\infty}^{\infty} F(x) \psi(x) E(t, x - x_0)\, dx \tag{4.3.8}$$

tend to $F(x_0)$ as $t \downarrow 0$. Let $G = \psi F$. This is a $C_0^\infty(\mathbf{R})$ function, and so is bounded and continuous on $\mathbf{R}$. Corollary 4.0.4 says that (4.3.8) tends to $G(x_0) = \psi(x_0)F(x_0)$ as $t \downarrow 0$, but since $\psi(x_0) = 1$, that limit is just $F(x_0)$, as was to be shown. Thus, the function F is completely determined by the collection of numbers (4.3.5): the distribution T_F defines the function F.

This is surely a completely new way to define functions on $\mathbf{R}$: we think of them as continuous functions on a much larger set, the test-function space $C_0^\infty(\mathbf{R})$. The resulting object is a distribution T_F. Knowing that T_F is derived from a function F according to (4.3.5), we can recover the usual description of F. There are, however, many distributions (= continuous linear functions on $C_0^\infty(\mathbf{R})$) that are *not* derived from ordinary functions according to the prescription (4.3.5). Those more general distributions are, naturally enough, called *generalized functions*.

4.3.9 Example. All the distributions we have met until now are, in a certain sense, *limits* of ordinary functions. The Dirac δ is a limit of $\mathcal{S}(\mathbf{R})$ functions — that is one way to phrase Proposition 4.0.3. The Cauchy PV $\frac{1}{x}$ is a limit of the discontinuous functions

$$f_\epsilon(x) = \begin{cases} 1/x, & \text{if } |x| > \epsilon, \\ 0, & \text{otherwise.} \end{cases}$$

The distribution T_{f_ϵ} determined by the function f_ϵ is just

$$T_{f_\epsilon}(\varphi) = \int_{-\infty}^{\infty} f_\epsilon(x)\varphi(x)\,dx = \int_{|x|>\epsilon} \frac{\varphi(x)}{x}\,dx,$$

and as $\epsilon \downarrow 0$, we get precisely PV $\frac{1}{x}$. These f_ϵ are not in $C^\infty(\mathbf{R})$, but — roughly speaking — the discontinuities can be rounded off in a smooth way to produce a sequence $f_\epsilon \in C^\infty(\mathbf{R})$ that still converges to PV $\frac{1}{x}$: if the f_ϵ have the shape in Figure 71, it should still be true that

(4.3.10). $$\lim_{\epsilon \downarrow 0} \int_{-\infty}^{\infty} f_\epsilon(x)\varphi(x)\,dx = \mathrm{PV}\,\frac{1}{x}(\varphi)$$

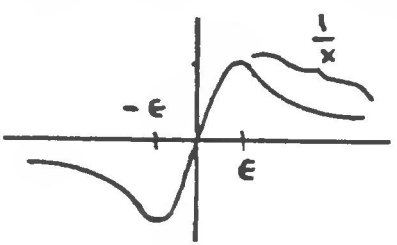


Figure 71. $C^\infty(\mathbf{R})$ *functions converging to* PV $\frac{1}{x}$

At this point, I hope you are thinking: but wait! the functions he just drew don't converge in $C^\infty(\mathbf{R})$! Quite true, and more power to you for seeing this important point. Your question leads to the next topic: convergence of *distributions*. You see, the limit (4.3.10) is not a limit of *functions*, it is a limit of *integrals of functions*. There is quite a difference between having a bunch of functions f_ϵ and all their derivatives converge pointwise (for each x), and having an integral (4.3.10) of those same functions tend to a limit. In a nutshell: the former is *strong* convergence, the latter is *weak* convergence. ∎

We will have just a few additional comments on convergence of distributions, but first it is time to set down the complete definitions of various classes of distributions and introduce standard vocabulary. The symbol $\mathcal{X}$ will stand for one of the spaces $C^\infty(\mathbf{R}^n)$, $C_0^\infty(\mathbf{R}^n)$, and $\mathcal{S}(\mathbf{R}^n)$, equipped with the topologies discussed in Subsection 4.2. The test functions are allowed to assume complex values.

4.3.11 Definition. *A distribution T on the test function space $\mathcal{X}$ is a function from $\mathcal{X}$ to $\mathbf{C}$ with these properties:*

a) T is complex linear, i.e.: $T(\alpha\varphi_1 + \beta\varphi_2) = \alpha T(\varphi_1) + \beta T(\varphi_2)$, for all $\alpha, \beta \in \mathbf{C}$ and $\varphi_1, \varphi_2 \in \mathcal{X}$.

ii) T is continuous with respect to the topology on $\mathcal{X}$. ∎

4.3.12 Example. The simplest illustration is, once again, the Dirac δ on the three test function spaces $C^\infty(\mathbf{R}), C_0^\infty(\mathbf{R})$, and $\mathcal{S}(\mathbf{R})$. In all three cases, one defines $\delta(\varphi) = \varphi(0)$. Linearity (property (4.3.11 a)) is trivial:

$$\delta(\alpha\varphi_1 + \beta\varphi_2) = \alpha\varphi_1(x) + \beta\varphi_2(x)\Big|_{x=0} = \alpha\varphi_1(0) + \beta\varphi_2(0) = \alpha\delta(\varphi_1) + \beta\delta(\varphi_2).$$

Continuity on $C^\infty(\mathbf{R})$ was checked in Example 4.2.49.A; the argument in all other cases is similar. ∎

4.3.13 Example. The derivative of the Dirac δ is another simple, but useful, distribution. It is defined by

$$\delta'(\varphi) = -\varphi'(0). \tag{4.3.14}$$

The minus sign might be unexpected. To understand why it arises, consider once again the solution (4.0.1) of the heat equation,

$$\frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}. \tag{4.0.1}$$

We have seen that (in a sense yet to be made completely precise), this function tends to δ as $t \downarrow 0$. Its derivative with respect to x,

$$-\frac{x}{4\sqrt{\pi t^3}} e^{-x^2/4t} \tag{4.3.15}$$

should therefore tend to δ'. Again, such a statement must be interpreted in terms of test functions: we will *define* the limit

$$\lim_{t \downarrow 0} \int_{-\infty}^{\infty} -\frac{x}{4\sqrt{\pi t^3}} e^{-x^2/4t} \varphi(x)\, dx \tag{4.3.16}$$

to be $\delta'(\varphi)$. To see what this limit might be, look at the graph of the function (4.3.15).

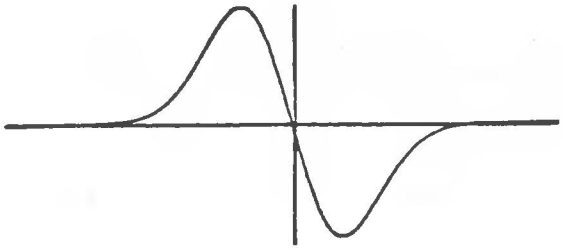


Figure 72. *The derivative of the Dirac delta*

This function has a large positive peak at $x = -\sqrt{2t}$ and a large negative peak at $x = \sqrt{2t}$. The integral (4.3.16) therefore picks out two values of the test function φ, with opposite signs:

$$\text{Integral (4.3.16)} \sim -A\Big[(\varphi(\sqrt{2t}) - \varphi(-\sqrt{2t}))\Big], \tag{4.3.17}$$

where A is an approximation to the area under each of the big peaks. It turns out that this area is comparable to $1/\sqrt{t}$, and (4.3.17) has the form (h replaces $\sqrt{t}$)

$$-\frac{\varphi(h) - \varphi(-h)}{2h} \to -\varphi'(0) \text{ as } h \to 0.$$

Such an argument is not easily made precise, so a more quantitative approach is outlined in Exercise 4.3.27.

We now want to show that δ' defines a distribution on $C_0^\infty(\mathbf{R})$. Complex linearity (as in (4.3.11 a)) is easy; let us prove that δ' is a continuous function.

Suppose that $\varphi_j \to \varphi$ in the sense of Definition 4.2.51. Then, in particular, $p_{R,1}(\varphi_j - \varphi) \to 0$, so that

$$|\varphi_j'(0) - \varphi'(0)| \le \max_{0 \le k \le 1} \sup_{|x| \le R} |\varphi_j^{(k)}(x) - \varphi^{(k)}(x)| = p_{R,1}(\varphi_j - \varphi) \to 0.$$

Hence

$$|\delta'(\varphi_j) - \delta'(\varphi)| = |\varphi_j'(0) - \varphi'(0)| \to 0. \blacksquare$$

Finally, we return, briefly, to convergence of distributions. This term covers not only limiting processes in which ordinary functions are interpreted as tending to a generalized function (as in Examples 4.3.9 and 4.3.13), but also convergence of sequences whose members are generalized functions. If there is to be a notion of "convergence of distributions", then the distributions themselves must be organized into a topological space.

4.3.18 Definition. *Let $\mathcal{X}$ be a test-function space (equipped with the appropriate topology). The space of distributions on $\mathcal{X}$, denoted by $\mathcal{X}'$, is the linear space of all complex - linear continuous functions from $\mathcal{X}$ to* C. ∎

4.3.19 Remark. The space of distributions on $C_0^\infty(\mathbf{R}^n)$ is often called $\mathcal{D}'(\mathbf{R}^n)$ (because $C_0^\infty(\mathbf{R}^n)$ is itself often denoted by $\mathcal{D}(\mathbf{R}^n)$). The distributions on $\mathcal{S}(\mathbf{R}^n)$ are called *tempered distributions.*[34] ∎

4.3.20 Definition. *Let T_j be a sequence of distributions in $\mathcal{X}'$. We say that $T_j \to T$ as $j \to \infty$ if $T_j(\varphi) \to T(\varphi)$ for all test functions $\varphi \in \mathcal{X}$.* ∎

The topology on $\mathcal{X}'$ is determined implicitly by the definition of convergence. It is possible, but somewhat difficult and usually not necessary, to provide explicit formulas for the sets of a subbase. Alternatively, one can (implicitly again) define the topology on $\mathcal{X}'$ to be the smallest topology in which the functions

$$\Phi_\varphi : \mathcal{X}' \to \mathbf{C}, \quad \Phi_\varphi(T) \stackrel{\text{def}}{=} T(\varphi),$$

as φ ranges over $\mathcal{X}$, are continuous. The topology on $\mathcal{X}'$ then fits the pattern of Proposition 3.4.19b).

When the distributions in Definition 4.3.20 are of the form $T_j = T_{F_j}$, defined by functions F_j, then the convergence of distributions gives rise to a new notion of convergence of functions. We state the definition in a special case.

4.3.21 Definition. *Let $\{F_j\}$ be a sequence of functions in $C^\infty(\mathbf{R})$, and suppose that the distributions T_{F_j} converge in $\mathcal{D}'(\mathbf{R})$ to a distribution T. Then the F_j are said to converge in the sense of distributions, or to converge weakly, to the distribution T.* ∎

4.3.22 Examples

---

[34] Why "tempered"? Recall, from Example 4.2.49, that distributions on $C^\infty(\mathbf{R}^n)$ must be very localized. Correspondingly, distributions on $C_0^\infty(\mathbf{R}^n)$ are quite unrestricted, because the test functions are so tame. In the middle sit the distributions on Schwartz space: the test functions don't grow, but they don't just whimper away to zero, either. Their distribution are neither localized, nor unrestrained — they are *tempered*.

A. We saw in Proposition 4.0.3 that the functions

$$T_j(x) \stackrel{\text{def}}{=} E(\frac{1}{j^2}, x) = \frac{j}{\sqrt{4\pi}} e^{-j^2 x/4}$$

have the property

$$\int_{-\infty}^{\infty} T_j(x)\varphi(x)\,dx \to \varphi(0) \text{ as } j \to \infty. \tag{4.3.23}$$

Think of the T_j not as functions, but as distributions, as in (4.3.5). The limit (4.3.23) then realizes Definition 4.3.21: $T_j(\varphi) \to \delta(\varphi)$.

There is no way to speak of convergence to δ of the T_j as *functions* on R. They do, for example, converge to the zero function almost everywhere in the sense of Lebesgue (see Chapter 2), but this is of little interest.

B. The δ-distribution on the sphere (see Supplement 4.1.57) is the limit, in the sense of Definition 4.3.20, of distributions δ_ϵ defined by (4.1.61). These distributions are in fact of the form T_F, for the functions d_ϵ in (4.1.59). The δ distribution on the sphere, therefore, is the *distribution limit* of a sequence (indexed by ϵ) of piecewise continuous functions. It is not necessary to confine distribution limits to smooth functions, but the technicalities would lead quickly into Lebesgue integration theory. ∎

We have now arrived at the beginning of distribution theory. We could learn to differentiate distributions, multiply generalized function by ordinary ones, and build distribution solutions of differential equations. These matters, however, go too far beyond the present, limited perspective of topology and continuity. They belong more properly in later discussions of linearity and duality.

• **4.3.24 Exercise.** Let δ_j $(j = 1, 2, \ldots)$ be the functions defined on R by

$$\delta_j(x) = \begin{cases} j/2, & \text{if } |x| \le 1/j, \\ 0, & \text{otherwise.} \end{cases}$$

Prove that each δ_j defines a distribution on $C_0^\infty(\mathbf{R})$, and that $\delta_j \to \delta$ as $j \to \infty$. This example shows that discontinuous functions can also be used in formula (4.4.14) to define distributions. ∎

• **4.3.25 Exercise.** Prove that δ is continuous on $C^\infty(\mathbf{R})$ and $S(\mathbf{R})$. ∎

• **4.3.26 Exercise.** Prove that δ' is continuous on $C^\infty(\mathbf{R})$ and $S(\mathbf{R})$. ∎

4.3.27 Exercise. Prove that the limit in (4.3.16) equals $-\varphi'(0)$ for every $\varphi \in C_0^\infty(\mathbf{R})$. This will justify the definition $\delta'(\varphi) = -\varphi(0)$. To this end, you must show that

$$\lim_{t \downarrow 0} \int_{-\infty}^{\infty} -\frac{x}{4\sqrt{\pi t^3}} e^{-x^2/4t} \varphi(x)\,dx = -\varphi'(0). \tag{4.3.28}$$

(Hint: integrate by parts in (4.3.28). ∎

• **4.3.29 Exercise.** Define $\delta''(\varphi) = \varphi''(0)$. Show that this is a distribution on each of the test-function spaces $\mathcal{X}$. (Verify the conditions in Definition 4.3.11). ∎

4.3.30 Exercise*. Show how to construct the functions f_ϵ sketched in Figure 71, and *prove* that (4.3.10) is true (for all $\varphi \in C_0^\infty(\mathbf{R})$.) ∎

• **4.3.31 Exercise.** Show that the rule

$$\varphi \mapsto \int_{-\infty}^{\infty} e^{ixt}\varphi(x)\,dx$$

defines a distribution on $C_0^\infty(\mathbf{R})$. ∎

• **4.3.32 Exercise.** Continue the preceding Exercise. Let T_j be the distribution defined by $\exp ijx$. Show that $T_j \to 0$ in the sense of Definition 4.3.20. (Hint: write

$$\int_{-\infty}^{\infty} e^{ijx}\varphi(x)\,dx = \frac{1}{ij}\int_{-\infty}^{\infty} \frac{d}{dx}\left(e^{ijx}\right)\varphi(x)\,dx$$

and integrate by parts. See Example 3.7.18. ∎

4.3.33 Exercise. Show that for each $t \in \mathbf{R}$ and integer $k \geq 0$,

$$\varphi \mapsto \int_{-\infty}^{\infty} x^k e^{itx}\varphi(x)\,dx$$

defines a distribution on $C_0^\infty(\mathbf{R})$; set $t = j$ and find the limit as $j \to \infty$ of the resulting sequence. ∎

4.3.34 Exercise. Show that the sum

$$\sum_{k=-\infty}^{\infty} a_k \delta(x-k)$$

converges in $\mathcal{D}'(\mathbf{R})$ (see (4.3.20)) for arbitrary $a_k \in \mathbf{C}$. [Hint: convergence of the series means: the partial sums $\sum_{-N}^{M}$ converge in the sense of Definition (4.3.20) as $N \to \infty, M \to \infty$.] ∎

The next five exercises form a set. The first problem is the hardest. If you take it for granted, the rest should be fairly straightforward.

4.3.35 Exercise*. Set $\varphi_j(x) = (\sin jx)/x$, where $j = 1, 2, \ldots$. Sketch the graph of φ_j for large j. If the integral

$$\int_{-\infty}^{\infty} \frac{\sin jx}{x}\,dx$$

exists, we can set $u = jx$ and convert that integral to

$$\int_{-\infty}^{\infty} \frac{\sin u}{u}\, du.$$

This is independent of j. Now notice how all the area under φ, concentrates in the main "bump" centered on $x = 0$. One can show, by contour integration, that the (improper) Riemann integral $\int_{-\infty}^{\infty} \sin x/x\, dx$ has the value π. This suggests that $\sin jx/x \to \pi\delta(x)$ as $j \uparrow \infty$. Prove this — i.e., prove that

$$\lim_{j\to\infty} \int_{-\infty}^{\infty} \frac{\sin jx}{x}\varphi(x)\, dx = \pi\varphi(0). \blacksquare$$

4.3.36 Exercise*. Define a distribution ($\in \mathcal{D}'$) as follows:

$$\mathrm{PV}\frac{\sin \infty x}{x}(\varphi) \stackrel{\text{def}}{=} \lim_{j\to\infty} \lim_{\epsilon\downarrow 0} \int_{|x|>\epsilon} \frac{\sin jx}{x}\varphi(x)\, dx.$$

In words: for each test function φ, compute $\mathrm{PV}\,\frac{1}{x}(\sin jx \cdot \varphi(x))$, and then take the limit $j \to \infty$. *Prove* that this limit is $\pi\delta(x)$. (Hint: use formula (4.1.27). The integral $\int_{|x|\geq 1}$ tends to zero, as in Exercise 4.3.32, and the limit of the integral $\int_{|x|\leq 1}$ is found from Exercise 4.3.35.) $\blacksquare$

4.3.37 Exercise*. Define a distribution ($\in \mathcal{D}'$) as follows:

$$\mathrm{PV}\frac{\cos \infty x}{x}(\varphi) \stackrel{\text{def}}{=} \lim_{j\to\infty} \lim_{\epsilon\downarrow 0} \int_{|x|>\epsilon} \frac{\cos jx}{x}\varphi(x)\, dx.$$

Show that this distribution is identically *zero* on all test functions $\varphi \in C_0^\infty(\mathbf{R})$. (Hint: use (4.1.27), (4.3.32), and some common sense.) $\blacksquare$

4.3.38 Exercise*. Define a distribution on $\mathcal{D}'$ as follows:

$$\frac{e^{i\infty x}}{x + i0}(\varphi) \stackrel{\text{def}}{=} \lim_{j\to\infty} \frac{1}{x + i0}(e^{ijx}\varphi(x)).$$

Show that this distribution is identically zero (i.e., the limit in the last equation is 0 for all test functions $\varphi \in C_0^\infty(\mathbf{R})$). $\blacksquare$

4.3.39 Exercise*. Find the following limits in $C_0^\infty(\mathbf{R})'$:

$$\lim_{j\to\infty} \frac{e^{ijx}}{x - i0},$$

$$\lim_{j\to\infty} \frac{e^{-ijx}}{x - i0},$$

$$\lim_{j\to\infty} \frac{e^{-ijx}}{x + i0}. \blacksquare$$

Section 5 The Size of Sets

Until now, we have been concerned exclusively with the size of elements of a set and with the distance between two elements. Now we want to measure the size of the sets themselves: the volume of a sphere, the surface area of an ellipsoid, the size of the set of differentiable functions as subset of the set of all functions.

There are *quantitative* measures of size ("the area of a circle is πr^2") and *qualitative* measures ("this set is puny"). We begin, in Subsection 1, with some *qualitative* measures of size; these can give a precise meaning to statements like "typically," "generically," or "almost always."

It is rather more difficult to assign numerical values of size to sets; that is the concern of *measure theory*, and an outline is given in Subsection 2. There are two aspects to measure theory. In what is sometimes called *abstract measure theory*, one assumes given a method for assigning size to subsets of a set X; of course, some axioms must be satisfied. (This way of assigning size to subsets will be called a *measure on* X.) One then draws further conclusions from these axioms. This is a bit like the beginning of general topology, where $(X, \mathcal{T})$ appeared out of the blue, and we simply manipulated elements of the abstract set X. As we saw in Subsections 3.6 and 3.7, it can be quite difficult to take a specific set X and equip it with an interesting and useful topology. This is even more true in measure theory, whose second component is the *construction* of measures on sets that arise in "concrete" problems, for example in integration theory or probability theory.

The survey in Subsection 2 concentrates on abstract measure theory. Very little is ever said about the subtle and often quite technical problem of constructing measures; I merely try to indicate where, exactly, the problem lies. In Subsections 3 and 4, there are two typical applications of measure theory: to Hausdorff measure, a method for measuring the size of "fractured" sets in Euclidean space, and to Brownian motion, a probabilistic theory of diffusion. In both cases I will present formulas arising from certain measures, but the construction of the measures will not be explained.

Subsection 5.1 Small sets, large sets

Topology, we have seen, gives a meaning to "close" without answering the question "what's the distance". In this subsection, I will explain how topology gauges the size of sets, without answering the question "how big is it?" We begin with a definition from Section 3.

3.5.12 Definition. *Let A be a subset of X. A point $x \in A$ is called an* ***interior point*** *of A if there is an open U so that $x \in U \subset A$. Let*

$$A^\circ = \{x \mid x \text{ is an interior point of } A\}.$$

The set A° is called the ***interior*** *of A.* ∎

There are degrees of smallness, and degrees of largeness. When a set is small in certain sense, its complement is, in a natural sense, large (and vice versa). Several of these notions have accepted technical names, which are now introduced. Sometimes the corresponding property of the complement does not have an official name; those situations are dealt with in the exercises. Be sure to read them. In all statements, an underlying topological space $(X, \mathcal{T})$ is assumed given.

5.1.1 Definition. *A set $A \subset X$ is said to be* ***nowhere dense*** *if its closure $\overline{A}$ has empty interior, i.e. $(\overline{A})^\circ = \emptyset$.* ∎

5.1.2 Definition. *If A is a countable union of nowhere dense sets, it is called* ***meager*** *or a* ***set of first category***[35]. ∎

5.1.3 Definition. *The complement of a meager set is called a* ***residual set.*** ∎

5.1.4 Definition. *A set $A \subset X$ is said to be* ***dense in*** *X if $\overline{A} = X$. If $A \subset B$ then A is said to be* ***dense in*** *B if $B \subset \overline{A}$.* ∎

- **5.1.5 Exercise.** A set A is nowhere dense if, and only if, its complement has dense interior, i.e. $\overline{(A^c)^\circ} = X$. ∎

- **5.1.6 Exercise.** A set A is dense if, and only if, its complement has empty interior, i.e. $(A^c)^\circ = \emptyset$. ∎

---

[35] Also a set of *first Baire category*, after René Baire (1874-1932), who encountered such sets in the study of discontinuous functions. Incidentally, any set that is not of the first category is, cleverly, said to be of *second category*.

• **5.1.7 Exercise.** The complement of a meager set is residual. ∎

These notions of "small" and "large" can now be ordered and paired as follows. Examples will show that the nontechnical descriptions could be misleading—use them as a rough guide only.

Very small :	nowhere dense
Very large :	dense interior
Small :	meager
Large :	residual
Medium small :	empty interior
Medium large :	dense

5.1.8 Examples

A. Let $A = \{(x, y) \in \mathbf{R}^2 \mid x = y\}$. This is a closed subset of $\mathbf{R}^2$, but it contains no points that are interior points of $\mathbf{R}^2$: no point on the line A is enclosed by an open disk that is again contained in the line A. The set A is therefore nowhere dense in $\mathbf{R}^2$.

B. The closure of the set $\mathbf{Q}$ of rational numbers in $\mathbf{R}$ is all of $\mathbf{R}$. Hence, $\mathbf{Q}$ is dense in $\mathbf{R}$.

C. The rationals are a countable subset $\{r_1, r_2, r_3, \ldots\}$ of $\mathbf{R}$. Each singleton set $\{r_n\}$ is nowhere dense, and therefore

$$\{r_1, r_2, \ldots\} = \bigcup_{n=1}^{\infty} \{r_n\}$$

is a set of the first category (a meager set).

Even though the rational numbers are dense ("medium large"), they are still "small" in the technical sense of being meager. They are also "medium small" in that they have empty interior. This reveals shortcomings in the list above!

D. The irrational numbers in $\mathbf{R}$ form a dense subset of $\mathbf{R}$. Because the complement is meager, the irrationals are a residual set. It may seem obvious that they are not also a meager set, but it is in fact possible for a set to be both meager and residual (Exercise 5.1.14). That this does not happen with the irrationals is a consequence of the *Baire category theorem* II.1.1.36. ∎

E. The set of pairs (x_1, x_2) with both coordinates rational is dense in $\mathbf{R}^2$.

F. Let $A \subset l^1(\mathrm{R}, \mathrm{N})$ be the set of sequences $(y_0, y_1, \ldots, y_m, 0, 0, \ldots)$ in which all but a finite number of entries are zero. The number m of (possibly) nonzero entries is not fixed. The set A is dense in $l^1(\mathrm{R}, \mathrm{N})$.

To prove this, one must show that the closure of A is all of $l^1(\mathrm{R}, \mathrm{N})$, i.e., that every $\mathbf{x} \in l^1(\mathrm{R}, \mathrm{N})$ is a point of closure of A. Choose an arbitrary $\epsilon > 0$, and show that the l^1 ball $B(\mathbf{x}, \epsilon)$ contains an element of A.

Here is the argument. Since

$$\sum_0^\infty |x_j| < \infty,$$

there is an m such that

$$\sum_{m+1}^\infty |x_j| < \epsilon.$$

Now define y by

$$y_j = \begin{cases} x_j, & \text{if } 0 \le j \le m, \\ 0, & \text{if } j \ge m+1. \end{cases}$$

Clearly,

$$\|\mathbf{x} - \mathbf{y}\|_1 = \sum_0^m |x_j - x_j| + \sum_{m+1}^\infty |x_j - 0| < \epsilon,$$

and since $\mathbf{y} \in A$, the claim is established. ∎

- **5.1.9 Exercise.** Let (X, T) be a topological space. Prove that a subset A of X is nowhere dense if and only if for each nonempty open $U \subset X$, $A \cap U$ is not dense in U. This is the origin of the term "nowhere dense." ∎

- **5.1.10 Exercise.** Let (X, T) be a topological space. Prove or find a counterexample: a set A is dense in X if, and only if, its complement A^c is nowhere dense. ∎

- **5.1.11 Exercise.** Prove that the union of a finite number of nowhere dense sets is nowhere dense. ∎

- **5.1.12 Exercise.** Give examples to show that a countable union of nowhere dense subsets of a topological space (X, T) can be a) dense, b) nowhere dense, c) neither dense nor nowhere dense. ∎

5.1.13 Exercise. Show that the Cantor set (Example 3.1.6) is nowhere dense in $[0, 1]$. ∎

5.1.14 Exercise. The rationals Q with the usual metric, $|\cdot - \cdot|$, form a metric space. Show: a) If $r \in \mathrm{Q}$, then $\mathrm{Q} - \{r\}$ is open. b) There exists a countable family of dense open sets whose intersection is empty.

Hence, a residual set, which one wants to think of as "large", may in fact be exceedingly small. A topological space in which every residual set is dense is called a **Baire space**. ∎

5.1.15 Exercise. Let $1 \leq p < \infty$. Prove that the set

$$A = \{(x_0, x_1, \ldots) \in l^p(\mathbf{R}, \mathbf{N}) \mid \text{all but finitely many } x_i \text{ are } = 0\}$$

is dense in (the usual topology on) $l^p(\mathbf{R}, \mathbf{N})$. What happens for $p = \infty$? ∎

5.1.16 Exercise. Let f be a continuous real-valued function on R. Prove that the set $\{x \mid f(x) = 0\}$ is nowhere dense if, and only if, there is no interval (a, b) on which f is identically zero. ∎

We have seen that sets of the first category, which are considered to be "small", can be dense (e.g., the rationals in R). It would be reassuring to have an independent confirmation of "smallness", and in particular, to answer the question: how small? We therefore turn now to the concept *Lebesgue measure zero*. This is only the first attempt to quantify "size". In the next subsection we will look at "bigger" sets whose size is a positive number.

Sets of measure zero are "small", in a certain well - defined sense. The concept agrees with Baire's first category some of the time. Nowhere dense sets or first category sets, however, need not have measure zero. The notion "measure zero" will be useful in some situations and irrelevant in others.

5.1.17 Definition. *A subset A of $\mathbf{R}^n$ is said to have measure zero (more precisely, Lebesgue measure zero) if for every $\epsilon > 0$, there is a countable collection $\{B_n\}$ of open balls*

$$B_n = B(x_n, \delta_n) = \{x \mid ||x - x_n||_2 < \delta_n\}$$

such that

$$A \subset \bigcup_{n=1}^{\infty} B_n \tag{5.1.18}$$

and

$$\sum_{n=1}^{\infty} \text{volume of } B_n \leq \epsilon. \tag{5.1.19}$$ ∎

Note: the open balls $B(x_n, \delta_n)$ need not be disjoint (they usually aren't); the sum in (5.1.19) refers to the sum of the indiviual volumes of these balls, which is usually larger than the volume of the union.

5.1.20 Examples

A. The rationals $\mathbf{Q} \subset \mathbf{R}$ have measure zero. They are a countable set, $\{r_1, r_2, \ldots\}$, and the collection of balls

$$B_n = \{x \mid |x - r_n| < \frac{\epsilon}{2^n}\}$$

covers Q (requirement (5.1.18)) and the total volume, or length, is

$$\sum_{n=1}^{\infty} \text{length of } B_n < \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = \epsilon.$$

This result should be thought about. The set Q is dense in R, but it can be covered by a union of open intervals of total length less than your favorite ϵ. This is difficult to picture! Since the length of R is infinite, there must a be huge chunk of R that is *not* covered by those intervals. Yet, the rationals are dense, so how is it possible to cover them with open intervals and still avoid a large portion of R? My imagination fails me, and I, at least, simply believe the estimate above.

B. Countable sets *always* have Lebesgue measure zero: just enumerate the elements as $\{x_1, x_2, \ldots\}$ and enclose x_j in a ball of volume $\epsilon/2^j$.

C. There are *uncountable* sets of measure zero as well. The Cantor middle thirds set C has this property.

Recall the construction of C ((3.1.6)). It proceeds in steps. At step 1, the open middle third is removed form $[0,1]$, and C is contained in the union $[0,1/3] \cup [2/3,1]$, whose total length is $2 \times (1/3)$. At step 2, the middle third is removed from the intervals $[0,1/3]$ and $[2/3,1]$, and C is contained in the union $[0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$, of total length $4 \times (1/9)$. At step n, C is contained in a union of 2^n closed intervals, each of length $(1/3)^n$; the intervals in the union have total length $2^n(1/3)^n = (2/3)^n$. (The official proof of the statement about step n uses mathematical induction). Now, if $\epsilon > 0$ is given, we choose n so large that $(2/3)^n < \epsilon$, and cover C by the 2^n closed intervals constructed in step n. Their total length is $(2/3)^n < \epsilon$.

Definition 5.1.17 required the set A to be covered by *open* intervals. The stepwise construction of the Cantor set makes it natural to cover it by *closed* intervals. It turns out that one could equally well phrase Definition 5.1.17 using "closed ball", see Exercise 5.1.21. Assuming the result of that exercise, we have then proved that the Cantor set has measure zero. ∎

- **5.1.21 Exercise.** The requirement in Definition 5.1.17 that the set A be covered by countably many *open* balls of total volume $< \epsilon$ can be changed: show that if (for each $\epsilon > 0$) a set A can be covered by countably many *closed* balls of total volume $< \epsilon$, then it has measure zero in the sense of Definition 5.1.17. ∎

- **5.1.22 Exercise.** Let $A \subset \mathbf{R}$ have measure zero. Show that the complement A^c is dense in R. ∎

- **5.1.23 Exercise.** Prove: a countable union of sets of measure zero in $\mathbf{R}^n$ also has measure zero. ∎

• **5.1.24 Exercise.** Let $f : \mathbf{R} \to \mathbf{R}$ be continuously differentiable. Prove: if $A \subset \mathbf{R}$ has measure zero, then $f(A)$ also has measure zero. (Hint: It suffices to prove that $f(A \cap [-n, n])$ has measure zero for each n. Take for granted the following fact from calculus (Proposition II.1.3.8): the continuous function f' is bounded on the interval $[-n, n]$, say $\sup_{-n \le x \le n} |f'(x)| \le M$. Now cover A by intervals $\{B_n\}$ of total length $< \epsilon$. Use the Mean Value Theorem to estimate the length of an appropriate cover of $f(A)$ by open intervals.)

Interesting fact: if f is merely continuous, then $f(A)$ need *not* have measure zero. ∎

5.1.25 Exercise. Let $A = \{(x, y) \mid y = x\} \subset \mathbf{R}^2$. That is, A is the graph of the function $f(x) = x$. Show that A has Lebesgue measure zero in $\mathbf{R}^2$. ∎

5.1.26 Exercise*. Refer to Example 5.1.20A. Cover the rationals by intervals $B_{n,N} = (r_n - 1/N2^n, r_n + 1/N2^n)$, and set $A_N = \bigcup_{n=1}^{\infty} B_{n,N}$. Is

$$\mathbf{Q} = \bigcap_{N=1}^{\infty} A_N \, ? \, ∎$$

Our qualitative measures of the size of sets are used to describe "typical" properties of the elements of various interesting sets. There are several different notions of "typical". We define one of them, and indicate others in examples and a supplement.

5.1.27 Definition. *Let (X, T) be a topological space. A property P that elements of X may or may not have is called* ***generic (in the sense of Baire category)*** *if the set $\{x \in X \mid x$ has property $P\}$ is a dense residual set.* ∎

5.1.28 Examples

A. For real numbers, the property of being irrational is generic, since by Example 5.1.8D the irrational numbers are a dense residual set.

B. There is another sense in which the property of being irrational is generic: the complement of the set of irrationals has Lebesgue measure zero.

C. In the linear space of continuous functions on $[0, 1]$, the set of differentiable functions is meager. In fact, a continuous function generically (in the sense of Baire category) has no derivative anywhere; see Example II.1.4.28E. There is also a way to say that differentiable functions are a set of measure zero in the space of continuous functions; this is described in Section 5.4.

D. A quadratic equation $x^2 + bx + c = 0$, with b, c complex, has distinct roots exactly when $b^2 - 4c \neq 0$. In the 2-dimensional space of quadratic equations coordinatized by (b, c), this condition defines a dense open set. A dense open set is a particular, "very large" version of a dense residual set. Hence the

property of having distinct roots is generic for quadratic equations, and in fact it is "typical" in a stronger sense than required by Definition 5.1.27. Further details are given in Supplement 5.1.31.

E. There are nowhere dense sets that do not have measure zero, see Exercise 5.2.30.

F. Most of the really important statements about generic properties are highly technical and difficult to prove. There is, for example, a result along the following lines: it is impossible to find an explicit formula for the solution of the generic conservative mechanical system[36]. Obviously, a lot of vague words have to be made precise here! ∎

5.1.29 Example. Consider the function $f(x,y,z) = z^2 - x^2 - y^2$ from R^3 to R. The set defined by $f(x,y,z) = c$ is either empty, or a smooth surface (a hyperboloid), or, when $c = 0$, a circular cone. The vertex of the cone is an example of a *singular point*, a point where the surface is pinched in some way. One learns in multivariable calculus that this singular point occurs because all partial derivatives of f vanish at $x = y = z = 0$.

Now let $f : \mathrm{R}^n \to \mathrm{R}^m$ be a smooth function. For what set $\{c = (c_1, \ldots, c_m) \in \mathrm{R}^m\}$ does $f^{-1}(c)$ have singular points? For what c is it smooth? This is an important question, and a general answer is provided by *Sard's Theorem*: the set of $c \in \mathrm{R}^m$ for which $f^{-1}(c) \subset \mathrm{R}^n$ has singular points is of Lebesgue measure zero. The proof is difficult except when $n = m = 1$. In this elementary case, the geometric interpretation is not very interesting. The analytical problem is:

Proposition. *Let $f : \mathrm{R} \to \mathrm{R}$ be twice continuously differentiable. Then the set*

$$\{y \mid y = f(x) \Rightarrow f'(x) \neq 0\}$$

has measure zero. ∎

Proof. Let $C = \{x \mid f'(x) = 0\}$. We need to show that $f(C)$ has measure zero (even though C itself need not, as when $f \in C_0^\infty(\mathrm{R})$). By Exercise 5.1.23, it is enough to prove that $C \cap [-n, n]$ has measure zero for every n. On $[-n, n]$, the continuous function f'' is bounded, say by M. Now recall Taylor's formula with remainder:

$$f(y) = f(x) + f'(x)(y - x) + (1/2)f''(\xi)(y - x)^2$$

for some ξ between x and y. When $x \in C \cap [-n, n]$, we have

$$f(y) = f(x) + (1/2)f''(\xi)(y - x)^2,$$

(36) L. Markus and K. Meyer, "Generic Hamiltonian systems are neither integrable nor ergodic", *Memoirs of the American Mathematical Society*, vol. 144 (1974)

and hence

$$|f(y) - f(x)| \leq \frac{M}{2}|y - x|^2. \tag{5.1.30}$$

Divide $[-n, n]$ into K subintervals of equal length $2n/K$. If I is one of them and $x \in C \cap [-n, n]$ belongs to I, then by (5.1.30), $f(I)$ is contained in an interval of length $M \cdot (2n/K)^2$, centered at $f(x)$. There are at most K such intervals. Hence $f(C \cap [-n, n])$ is covered by closed intervals of total length no greater than $M \cdot (2n/K)^2 \cdot K = 4n^2M/K$. This can be made arbitrarily small by taking K lage enough. ∎

5.1.31 SUPPLEMENT: Zariski generic sets

The *Zariski topology* was introduced in Supplement 3.3.13. Recall that a set is Zariski open if it is the complement of the solution set of a system of polynomial equations. By Exercise 3.5.24, the closure of a Zariski open subset of R is all of R so that a Zariski open subset of R is *dense* (in the Zariski topology on R). This is also true for the Zariski topologies on $\mathbf{R}^n$, $\mathbf{C}^n$, or on more complicated spaces.

The *complement* of a Zariski open set is (besides being Zariski closed) a very small set, by the standards of Zariski topology: it is *nowhere dense* (because its open complement is dense). A countable union of Zariski closed sets should therefore be thought of as "small": in the Zariski topology, such sets are of first Baire category.

Sets that are closed in the Zariski topology on $\mathbf{R}^n$ also have Lebesgue measure zero on $\mathbf{R}^n$. The proof of this statement will not be given; see Exercise 5.1.25 for an example. Since Zariski-closed subsets of $\mathbf{R}^n$ are nowhere dense *and* have measure zero, being Zariski-closed is a more precise description of "smallness" than either or both of the others.

5.1.32 Definition. *Let X be the set $\mathbf{R}^n$ or $\mathbf{C}^n$, equipped with the Zariski topology. A set A is called (Zariski) generic, if it contains a set that is open in the Zariski topology. A property P is called generic if the set of $x \in X$ having property P is a Zariski generic set. (Equivalently: a property is called (Zariski) generic if the set of points not having that property is contained in the solution set of finitely many algebraic equations.)* ∎

In this new language, Exercise 3.3.19 and Example 5.1.28D say that the property P: *a quadratic equation $x^2 + bx + c = 0$ has distinct roots* is Zariski - generic in $\mathbf{C}^2 = \{(b, c)\}$. The Zariski version of "typical" (or "large set" or "generic property") is usually the best concept to use in situations involving polynomial constraints.

We look at a few illustrations.

5.1.33 Example. The generic monic [37] polynomial of degree n,

$$p(x) = x^n + a_1x^{n-1} + \ldots + a_{n-1}x + a_n,$$

with complex coefficients, has n distinct roots.

[37] "monic" means that the leading coefficient is 1

The proof is an exercise in algebra. Even analysts should try to understand the argument (because the kind of "critical behavior" involved here is important in problems involving stability and bifurcation, and because, at a more fundamental level, mathematicians of the 1990's should know at least 10% of what Gauss knew).

To introduce the ideas, consider the polynomial $p(x) = x^2 + bx + c$. It has x_0 as a repeated root if, and only if, both $p(x_0)$ and $p'(x_0)$ vanish. We want conditions on the coefficients b, c that guarantee the existence of a simultaneous solution to $x_0^2 + bx_0 + c = 0$ and $2x_0 + b = 0$. The first equation implies that $x_0^3 + bx_0^2 + cx_0 = 0$. The second equation implies that $2x_0^3 + bx_0^2 = 0$ and that $2x_0^2 + bx_0 = 0$. We now have *three* equations for the three unknowns x_0^3, x_0^2, and x_0:

$$\begin{aligned} x_0^3 + bx_0^2 + cx_0 &= 0 \\ 2x_0^3 + bx_0^2 + 0x_0 &= 0 \\ 0x_0^3 + 2x_0^2 + bx_0 &= 0. \end{aligned}$$

For a nontrivial solution to exist, the coefficient determinant must vanish. This leads to the condition $-b^2 + 4c = 0$, which is of course expected from the quadratic equations formula

$$x_0 = \frac{-b \pm \sqrt{b^2 - 4c}}{2}.$$

This is a somewhat heuristic argument. The problem is that solutions of the homogeneous system are not obviously related by being the square and the cube of a particular one of them. In fact, that is only true for *one* of the infinitely many solution sets of that system. The heuristic method leads to the correct answer, however, and a more convincing argument is sketched below.

Next, we ask when the *cubic* polynomial $p(x) = x^3 + bx + c$ has a multiple root. Again, the condition is that $p(x_0) = p'(x_0) = 0$ for some x_0. The calculations are left as exercise; the result is the following.

The conditions

$$\begin{aligned} x_0^2 p(x_0) &= 0 \\ x_0 p(x_0) &= 0 \\ x_0^3 p'(x_0) &= 0 \\ x_0^2 p'(x_0) &= 0 \\ x_0 p'(x_0) &= 0 \end{aligned}$$

translate into 5 homogeneous linear equations for $x_0^5, \ldots x_0$. The 5×5 coefficient determinant must vanish, and that condition works out to

$$4b^3 + 27c^2 = 0. \tag{5.1.34}$$

Now consider the general monic polynomial

$$p(x) = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0. \tag{5.1.35}$$

The calculations must be done intelligently if they are not to become overwhelming, but one finds, again, that the vanishing of a certain determinant involving the coefficients

a_j is necessary for the existence of a multiple root, and sufficient for the existence of a double root. This equation, determinant = 0, is a polynomial equation in the space ($\mathbf{R}^n$ or $\mathbf{C}^n$) of coefficients $a_0, \ldots, a_{n-1}$. The complement is Zariski open; it is the set of coefficients for which the polynomial (5.1.25) has only simple roots. The property of having simple roots is therefore *generic*.

Here is a rigorous way to get the determinant condition. Suppose p is a polynomial of exact degree n, with complex coefficients. If p and p' have a nonconstant common factor q, then $p = qf$ and $p' = qg$ for polynomials f, g; moreover, f has degree at most $n-2$ and g has degree at most $n-2$. Note that evidently

$$(*) \qquad g(x)p(x) = f(x)p'(x).$$

If, conversely, (*) holds with some f, g of the indicated maximum degrees, then all n factors $(x - a_j)$ of p must divide fp'. But since f has at most $n-1$ such factors, at least one root of p must also be a root of p'.

If now (*) is written out, one gets a system of homogeneous equations for the coefficients of the polynomials f and g. This system has nontrivial solutions precisely when the determinant we arrived at in another way vanishes.

Another interpretation might help. Suppose p has multiple roots. Its coefficients then lie on the small set "determinant = 0". If you change the coefficients by a small amount, then typically (or "generically") they no longer lie on this small set. The perturbed polynomial will have simple roots only.

The polynomial in the coefficients of p whose vanishing implies multiple roots is called the *discriminant of the n^{th}-degree polynomials*. The set where the discriminant vanishes is quite complicated and has interesting and useful geometric features. It is called a *discriminant locus*; discriminant loci appear in bifurcation problems, in catastrophe theory, and elsewhere in problems involving singular behavior. ∎

5.1.36 Example. The generic complex $n \times n$ matrix $\mathbf{A}$ has distinct eigenvalues.

To prove this, we just look at the characteristic polynomial, $\chi(x) = \det(\mathbf{A} - x\mathbf{I})$. *Generically*, according to Example 5.2.3, it has distinct roots. The assertion can be made more explicit: the set of $n \times n$ (complex) matrices is identified with $\mathbf{C}^{n^2}$: the n^2 entries of these matrices serve as coordinates $x_1, \ldots, x_{n^2}$ for $\mathbf{C}^{n^2}$. The requirement that there be "repeated roots" imposes polynomial equations on those x_j. ∎

• **5.1.37 Exercise.** Carry out the details of the calculation of the discriminant (5.1.34) of the cubic polynomial. Test the condition on the polynomial $x^3 - 3x + 2$, and find the double root. ∎

5.1.38 Exercise. For a fourth-degree polynomial $p(x) = x^4 + bx^2 + cx + d$, the condition that there be a multiple root comes from the vanishing of a 7×7 determinant. Write it down and evaluate it. (There are lots of zeros, so preliminary simplification by row operations is useful). ∎

5.1.39 Exercise. Let $p(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$. What is the determinant whose vanishing gives the condition (on the a_i) that forces p to have a multiple root? ∎

• **5.1.40 Exercise.** In what sense is an $n \times n$ (complex) matrix "generically" invertible? ∎

• **5.1.41 Exercise.** Prove: the generic quadric curve in $\mathbf{R}^2$,

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

is *not* a parabola. ∎

• **5.1.42 Exercise.** Does the graph of the generic *real* cubic, $y = ax^3 + bx^2 + cx + d$ have an inflection point? (Your answer should give a precise meaning to the term "generic"). Show that *every* real cubic has an inflection point. The problem is more reasonable if "cubic" is replaced by "quartic". ∎

• **5.1.43 Exercise.** Write polynomials that define the following set: all $(a, b, c, d) \in \mathbf{C}^4$ for which the eigenvalues of

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

are not distinct. ∎

5.1.44 Exercise*. Think of the set of 2×2 complex matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ as $\mathbf{C}^4$. Use polynomial equations and inequalities to describe the set A of matrices that are similar to a Jordan block

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

for some $\lambda \in \mathbf{C}$. ∎

5.1.45 Exercise*. Let X be the set of nonzero 3×3 (complex) matrices whose eigenvalues are all zero. Thus, the Jordan canonical form of such a matrix is one of the following:

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Show that, generically, a matrix from the set X has a single Jordan block. ∎

• **5.1.46 Exercise.** Let $n \leq m$. Think of the set X of $n \times m$ real matrices as the Euclidean space $\mathbf{R}^{nm}$. Prove: The matrices of rank n are a Zariski - open subset of X. (One says: matrices of "full rank" are generic). ∎

• **5.1.47 Exercise.** Prove: Generically, two 2-dimensional planes in $\mathbf{R}^4$ intersect in one point. (A 2-dimensional plane is a set

$$\{\vec{v} \in \mathbf{R}^4 \mid \vec{v} = \vec{a} + r\vec{b} + s\vec{c},\ r, s \in \mathbf{R}\}.$$

Again, define "generic" before you try to prove anything). ∎

5.1.48 Exercise*. Let $n \leq k + l \leq 2n$. In $\mathbf{R}^n$, a k-dimensional plane and an l-dimensional plane generically intersect in a $k + l - n$-dimensional plane. (If you can't prove this, try the special case $n = 3, k = l = 2$.) ∎

Subsection 5.2 Measure Theory

A *measure* assigns a number ("volume," "probability," or, neutrally, "size") to certain subsets of a set X. The original impetus for the development of measure theory came from problems involving Fourier series and integration theory. While most applications of measures are tied up with integrals, there is still a significant component of the theory that is concerned only with the "size of sets." In this section, we will – among other things – measure the size of fractal sets, give meaning to the statement: "most functions are not differentiable anywhere," and lay the foundation for later developments of integration theory.

Why does one need a theory?

The object, as I just said, is to assign a measure of "size" to subsets of a set X. It is usually not very difficult to do this for special classes of simple sets; only when infinite unions or intersections or limits are brought in do latent contradictions reveal themselves.

5.2.1 Examples

A. We try to assign a "size" to subsets of the set X of rational numbers in $[0,1]$. If $E \subset X$ is the intersection of X with an interval of the form (a,b), $[a,b]$, $(a,b]$, or $[a,b)$, define

$$s(E) = b - a.$$

("s" for "size.") Now consider some consequences of this definition.

a) Since $\{r\} = X \cap [r,r]$ when r is rational, $s(\{r\}) = r - r = 0$. (reasonable)

b) Since $X = X \cap [0,1]$, $s(X) = 1 - 0 = 1$. (reasonable)

c) On the other hand, $X = \bigcup_{n=1}^{\infty}\{r_n\}$, where $\{r_1, r_2, \ldots\}$ is an enumeration of the countable set X. We therefore add up countably many sets of size 0, and produce a set of size 1.

Is this desirable? The ordinary concept of length does not behave like that. For example,

$$(0,1] = \bigcup_{n=1}^{\infty}(\frac{1}{n+1}, \frac{1}{n}].$$

The interval $(1/(n+1), 1/n]$ has length $1/n - 1/(n+1)$, and if we add up all the lengths, we get

$$\sum_{n=1}^{\infty}(\frac{1}{n} - \frac{1}{n+1}) = 1 = \text{ length of } (0,1].$$

What, then, is wrong with our definition of the "size" of subsets of X?

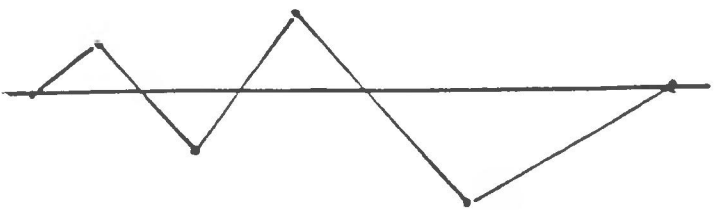
B. If curves $y = f_n(x), 0 \le x \le 1$, have finite length, and if

$$\max_{0 \le x \le 1} |f_n(x) - f(x)| \to 0 \text{ as } n \to \infty$$

(the f_n and f are continuous), does the curve $y = f(x)$ have finite length? *Not necessarily*: Connect the points

$$(\frac{2}{3}, -\frac{2}{3}), (\frac{2}{5}, \frac{2}{5}), \ldots, (\frac{2}{2n+1}, (-1)^n \cdot \frac{2}{2n+1}), (0,0)$$

by line segments.



These line segments are longer than $2/3, 2/5, \ldots, 2/(2n+1)$. The curve is the graph of a continuous function f_n on $[0,1]$. Let $n \to \infty$. The functions f_n tend to a continuous function f, which has a zig-zaggy graph that wiggles down to 0 at $x = 0$. The arclength is greater than the infinite series

$$\sum_{n=1}^{\infty} \frac{2}{2n+1},$$

which diverges to ∞ by comparison with the harmonic series. Hence, the limiting curve has infinite length.

C. If you approximate a circular cylinder by an inscribed polygon, do you necessarily approximate its surface area $2\pi rh$? *No, you can get the polygons to approximate any value* $\ge 2\pi rh$. (See EDM II, Article 246B, for an example due to H. A. Schwarz).

D. Is it possible to assign a "length" to every subset of R, so that an interval $[a, b]$ is assigned the usual length $b - a$, and so that two sets that differ only by rigid translation have the same length? *No.* (This is discussed in every text on Lebesgue measure theory). ∎

These examples illustrate that intuition may be of limited value when one starts to talk about "arbitrary subsets," "limits," and "infinity."

The goals of measure theory are (roughly speaking):

1) to start with a reasonable concept of "size" for simple subsets of the set of interest (we called it X), and to extend the measure of "size," without contradictions, to more complicated subsets;
2) to use the resulting structure in the analysis of a variety of problems.

The axioms of measure theory are easy to state, but rather harder to use, much harder to appreciate, and much much harder to verify in examples. We start with the axioms, and an attempt to interpret them.

5.2.2 Definition. *A measurable space is a set X together with a collection $\mathcal{B}$ of subsets of X satisfying*

$$X \in \mathcal{B},\ \emptyset \in \mathcal{B}, \tag{5.2.3}$$

$$\text{if } A \in \mathcal{B}, \text{ then } A^c \in \mathcal{B}, \tag{5.2.4}$$

$$\text{if } A_1, A_2, \ldots \in \mathcal{B}, \text{ then } \bigcup_{n=1}^{\infty} A_n \in \mathcal{B}. \tag{5.2.5}$$

A collection of subsets $\mathcal{B}$ of a set X is called a σ-algebra (read: "sigma algebra") if it satisfies (5.2.3), (5.2.4), (5.2.5). The sets in the collection $\mathcal{B}$ are called measurable sets. ∎

The Greek letter "σ" stands for "s," which is the first letter of the German "Summe," or "sum," which is the old term for "union." I.e., "σ" is a mnemonic for property (5.2.5).

"Measurable" means "capable of supporting a measure (of size of sets)." The significance of the concept should become clearer in a little while; for now just notice that the collection $\mathcal{B}$ is *not* the same as topology. In a topology, the complement of an open set is not usually open (e.g.: if $A = (a, b) \subset \mathbb{R}$, then $A^c = (-\infty, a] \cup [b, \infty)$, which is *closed* but not open). On the other hand, (5.2.5) asks only that $\mathcal{B}$ be closed under formation of *countable* unions, whereas the axiom (3.0.5) for a topology requires that *uncountable* unions of sets in $\mathcal{T}$ also belong to $\mathcal{T}$.

5.2.6 Remark. Repeated application of rules (5.2.4) and (5.2.5) shows that a σ-algebra is also closed under the formation of *countable intersections*. Indeed, suppose that $A_n, n = 1, 2, \ldots$, belong to $\mathcal{B}$. Then

$$\begin{aligned} A_n \in \mathcal{B} &\overset{(5.2.4)}{\Longrightarrow} A_n^c \in \mathcal{B} \\ &\overset{(5.2.5)}{\Longrightarrow} \bigcup_{n=1}^{\infty} A_n^c \in \mathcal{B} \\ &\overset{\text{de Morgan}}{\Longrightarrow} \Big(\bigcap_{n=1}^{\infty} A_n\Big)^c \in \mathcal{B} \\ &\overset{(5.2.4)}{\Longrightarrow} \bigcap_{n=1}^{\infty} A_n \in \mathcal{B}. \end{aligned}$$

A good first picture of a σ-algebra, therefore, is the following. Take a bunch of sets already in $\mathcal{B}$, and form countable unions and intersections and take complements until you get tired. All the sets you have created will again belong to $\mathcal{B}$. ∎

5.2.7 Example. You should be warned at the outset that the sets belonging to $\mathcal{B}$ cannot be characterized in any explicit way, except in the most trivial of examples. Usually, one starts with elementary sorts of sets that one desires to be "measurable," and one builds $\mathcal{B}$ by infinitely many repetitions of rules (5.2.4) and (5.2.5) (rather as one builds general open sets from the basis of a topology). Therefore, none of our examples will tell you exactly what the typical set in $\mathcal{B}$ looks like.

For example, imagine that a σ-algebra $\mathcal{B}$ of subsets of $\mathbf{R}$ is known to contain all open intervals. It then contains all open sets, because those are countable unions of open intervals (see 3.1.4). It contains all closed sets, because those are complements of open sets. It contains all F_σ and G_δ sets, and so on. This example contains the germ of an important construction that will appear later in this subsection. ∎

When $(X, \mathcal{B})$ is a measurable space, one usually wants to measure the size of the sets $A \subset X$ that belong to $\mathcal{B}$.

5.2.8 Definition. *Let $(X, \mathcal{B})$ be a measurable space. A* ***measure*** *on $(X, \mathcal{B})$ (or, more simply, on "X", if $\mathcal{B}$ is given) is a function*

$$\mu : \mathcal{B} \longrightarrow [0, \infty]$$

such that

$$\mu(\emptyset) = 0 \tag{5.2.9}$$

and

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n) \tag{5.2.10}$$

whenever $A_1, A_2, \ldots$ are disjoint sets belonging to $\mathcal{B}$, i.e., $A_n \cap A_m = \emptyset$ if $n \neq m$. A triple $(X, \mathcal{B}, \mu)$ consisting of a set X, a σ-algebra $\mathcal{B}$, and a measure μ is called a ***measure space.*** ∎

The conditions to be satisfied by a measure μ are fairly easy to understand. Note that one is permitted to have $\mu(A) = \infty$: there may well be sets whose "size" is infinite. (It should seem reasonable that the set $\mathbf{R}$ of real numbers has infinite length). The subtlest condition is (5.2.10); it tries to capture the relation

$$\text{length of } ((0,1]) = \sum_{n=1}^{\infty} \text{length of } \Big(\frac{1}{n+1}, \frac{1}{n}\Big]$$

mentioned in Example 5.2.1.A. The fact that countably *infinite*, rather than *finite*, sequences of sets enter that condition has surprisingly deep consequences. One cannot appreciate those without more detailed study of the foundations of measure theory.

There are two general results that will be useful in the examples. Both are intuitively obvious. It is understood that there is a fixed measurable space $(X, \mathcal{B})$ (5.2.2) with given measure μ. The proofs illustrate a technique that is applied in many situations. Since condition (5.2.10) deals with *disjoint* sets belonging to $\mathcal{B}$, statements about *overlapping* sets must always be converted to statements about disjoint sets. The key steps are isolated in a Lemma.

5.2.11 Lemma. *Let $A \subset B$. Then*

a) $B = (B - A) \cup A$ *is a disjoint union,*

b) $\mu(B) = \mu(B - A) + \mu(A)$. ∎

Proof. a) Left to the reader. b) In (5.2.10), take $A_1 = B - A$, $A_2 = A$, and $A_n = \emptyset$ for $n \geq 3$. The infinite sum in (5.2.10) reduces to a finite sum, since $\mu(\emptyset) = 0$, and the result follows. ∎

5.2.12 Proposition. *If $A, B \in \mathcal{B}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$.* ∎

Proof. By Lemma (5.2.11),

$$\mu(B) = \mu((B - A)) + \mu(A) \geq \mu(A). \blacksquare$$

5.2.13 Proposition. *If $A_n \in \mathcal{B}$ for $n \geq 1$, if $A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$, and if $\mu(A_1) < \infty$, then*

$$\mu(\bigcap_{n=1}^{\infty} A_n) = \lim_{m \to \infty} \mu(A_m). \blacksquare$$

Proof. We have $A_1 = (A_1 - A_2) \cup A_2$, $A_2 = (A_2 - A_3) \cup A_3$, and so forth. Write $A = \bigcap_{n=1}^{\infty} A_n$. Then

$$A_1 = A \cup \bigcup_{n=1}^{\infty} (A_n - A_{n+1}).$$

This is a *disjoint* union, so that by (5.2.10)

$$\mu(A_1) = \mu(A) + \sum_{n=1}^{\infty} \mu(A_n - A_{n+1}).$$

By Lemma (5.2.11), part b), $\mu(A_n - A_{n+1}) = \mu(A_n) - \mu(A_{n+1})$, and the last equation becomes

$$\mu(A_1) = \mu(A) + \sum_{n=1}^{\infty} \mu(A_n) - \mu(A_{n+1}). \tag{5.2.14}$$

This sum telescopes: by definition,

$$\begin{aligned} \sum_{n=1}^{\infty} \mu(A_n) - \mu(A_{n+1}) &= \lim_{m\to\infty} \sum_{n=1}^{m-1} \mu(A_n) - \mu(A_{n+1}) \\ &= \mu(A_1) - \lim_{m\to\infty} \mu(A_m). \end{aligned} \tag{5.2.15}$$

Insert (5.2.15) into (5.2.14) to get

$$\mu(A_1) = \mu(A) + \mu(A_1) - \lim_{m\to\infty} \mu(A_m). \tag{5.2.16}$$

The hypothesis $\mu(A_1) < \infty$ is now required: we can cancel $\mu(A_1)$ from both sides, to obtain

$$\mu(A) = \lim_{m\to\infty} \mu(A_m).$$

(With $\mu(A_1) = \infty$, (5.2.16) could stand for $\infty = 3 + \infty - 2$, from which we cannot deduce $3 = 2$).

This concludes the proof. ∎

One more general result will be needed; its proof is left as an exercise:

• **5.2.17 Exercise.** Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $B_n \in \mathcal{B}$ for $n = 1, 2, \ldots$. Show that

$$\mu\left(\bigcup_{k=1}^{\infty} B_k\right) \leq \sum_{k=1}^{\infty} \mu(B_k). \quad \blacksquare$$

5.2.18 Example. It is time to put these general results to some use. Let $X = \mathbf{R}$. We want to define "length" for subsets of R. The measure whose elementary features are described in this example is, up to a technical point explained in Chapter II, the most famous of all measures: *Lebesgue measure on* R.

The simplest sets to be assigned a length are the intervals, and we set

$$\mu((a, b)) = b - a. \tag{5.2.19}$$

If $a = -\infty$ and/or $b = \infty$, the interval will have length ∞. The open intervals are to be the basic sets from which the rest of the collection $\mathcal{B}$ is built by use of rules (5.2.4), (5.2.5), see Example 5.2.7.

Take $-\infty < a < b < \infty$. What is the length $\mu((a,b])$ of the half-open interval? Intuitively, the extra point has zero length, so one expects that $\mu((a.b]) = b - a$ again. Since we want the axioms of measure theory to apply, we require that

$$\mu((a,b]) = \mu((a,b)) + \mu(\{b\}). \tag{5.2.20}$$

Why should $\mu(\{b\}) = 0$? Let $A_n = (b - 1/n, b + 1/n)$. The intersection (over n) of these intervals is $\{b\}$. By Proposition 5.2.13, therefore,

$$\mu(\{b\}) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \frac{1}{2n} = 0.$$

Thus, a single point must, for sake of consistency, be assigned measure 0, and then (5.2.20) shows that $\mu((a,b]) = b - a$.

Evidently, similar arguments will give $\mu([a,b]) = b - a$, for finite closed intervals $[a,b]$.

The logic of this last argument is important. We want to measure the size of messy subsets of R (for whatever reason). Definition 5.2.8 is to govern "size". Once we admit open intervals, and declare their lengths to be given by $\mu((a,b)) = b - a$, *we have no choice about the measure to be assigned to* $\{b\}$ or $(a,b]$: those sets *must* belong to $\mathcal{B}$, and their measure *must* be, respecticely, 0 and $b - a$. This inevitability may also give rise to a fatal contradiction; see the Caution 5.2.21 below.

Finally, we compute the measure of the set of rationals and of the set of irrationals.

The rationals can be written as a countable, disjoint union of singleton sets,

$$\mathbf{Q} = \bigcup_{n=1}^{\infty} \{r_n\}.$$

Since singletons have measure 0, it *is inevitable* from (5.2.10) that $\mu(\mathbf{Q}) = 0$.

In Example 5.1.20, it was shown that the rationals have "Lebesgue measure zero" in the sense of Definition 5.1.17. Has anything new happened? Yes, the whole perspective has changed. In Subsection 1, only the term "measure zero" had meaning; there was no machinery for finding the "measure", or "size", of any set that did not have measure zero as defined in (5.1.17). We now have the beginning of a more comprehensive theory, measure theory, whose axioms and techniques, if cleverly enough applied, will allow us to compute a measure of the size of many more sets.

A simple illustration is afforded by the set $\mathbf{I}$ of irrational numbers. Since $\mathbf{R} = \mathbf{I} \cup \mathbf{Q}$ is a disjoint union,

$$\mu(\mathbf{I}) = \mu(\mathbf{R}) - \mu(\mathbf{Q}) = \mu(\mathbf{R}).$$

I claim that $\mu(\mathbf{R}) = +\infty$ (which seems obvious enough). To prove it, note that by Lemma 5.2.11

$$\mu(\mathbf{R}) \geq \mu((-n.n)) = 2n$$

holds for every n. It should also be noted that $\mathrm{R} \in \mathcal{B}$ (see (5.2.3)). ∎

5.2.21 Caution. Careful study of the last example will reveal the main difficulty inherent in the construction of a measure on a specific set. I have only shown how one can use the axioms and their corollaries to compute the measure of the rationals and the irrationals, once the measure of intervals is declared to be their usual length. I have *not* established, in a mathematical way, *consistency* of my starting definitions (5.2.19). Perhaps one can write the set Q of rationals in a quite different way (maybe a countable intersection of countable unions of countable intersections of open intervals), use the axioms correctly, and come up with the value $1/2$. It seems unlikely, but the possibility remains open.

The problem of *constructing* a measure, then, is invariably the following. One starts with a bunch of sets one wants to measure (i.e., they will belong to $\mathcal{B}$), and one declares what their measures are. To prove: every other set in $\mathcal{B}$ has an unambiguously defined measure, that value being independent of how the set happens to be built from the basic sets.

This matter is studied in detail in texts on Real Analysis or Probability Theory. A few sketchy hints are given in Subsection 4. ∎

When an example revolves around the familiar open and closed intervals, as does Example 5.2.18, it is not always clear what, if any, features have wider applicability. One of the possible abstract settings involves topology. The construction of the σ-algebra in 5.2.18 is generalized in the concept of *Borel σ-algebra*, which is explained next. Brownian motion in Subsection 4 affords a concrete example.

Let $(X, \mathcal{T})$ be a topological space. The set X usually has a large number of σ-algebras, any one of which could be used to build a measurable space $(X, \mathcal{B})$. Only one of these σ-algebras, however, interacts with the topology $\mathcal{T}$ in a nice way; it is often important that topology and measure theory be compatible with each other.

5.2.22 Proposition. *Let $(X, \mathcal{T})$ be a topological space. There is a unique smallest σ-algebra containing all open sets of the topology $\mathcal{T}$; it is called the* **Borel σ-algebra** *on X, written $\mathcal{B}(\mathcal{T})$ (or $\mathcal{B}$ for short, once the topology is fixed). The sets in that σ-algebra are called* **Borel sets**. *A measure on the measurable space $(X, \mathcal{B}(\mathcal{T}))$ is called a* **Borel measure**. ∎

Proof. The proof is outlined in Exercises 5.2.27,28, and 29. ∎

• **5.2.23 Exercise.** Consider the purported measure s in Example 5.2.1.A. On the one hand, by Proposition 5.2.12, one finds

$$s(X) \geq s(X \cap [\frac{1}{n}, 1 - \frac{1}{n}]) = 1 - \frac{1}{2n},$$

and on the other hand, axiom (5.2.10) requires

$$(X) = \sum_{n=1}^{\infty} s(\{r_n\}) = 0.$$

Think about this in light of Example 5.2.18 and Caution 5.2.21. ∎

• **5.2.24 Exercise.** Mimic the arguments in Example 5.2.18 to show that the Cantor set (5.1.19.C) must satisfy $\mu(C) = 0$. ∎

• **5.2.25 Exercise.** Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $A_n \in \mathcal{B}$ for $n = 1, 2, \ldots$. Show that

$$\mu\Big(\bigcup_{j=1}^{\infty} A_j\Big) = \lim_{n\to\infty} \mu\Big(\bigcup_{j=1}^{n} A_j\Big). \text{ ∎}$$

• **5.2.26 Exercise.** Let X be a set, and let $\mathcal{B}$ be the collection of all subsets of X. Show that $\mathcal{B}$ is a σ-algebra. ∎

5.2.27 Exercise. Let $\mathcal{B}_n$, $n = 1, 2, \ldots$, be a countable set of σ-algebras. Show that

$$\bigcap_{n=1}^{\infty} \mathcal{B}_n$$

is again a σ-algebra. Show that the same result holds if the countable set $\{\mathcal{B}_n\}$ is replaced by an uncountable set $\{\mathcal{B}_\lambda\}$. ∎

5.2.28 Exercise. Let X be a set, let $\mathcal{C}$ be a collection of subsets of X (not necessarily a σ-algebra), and let $\mathbf{B}$ be the collection of all σ-algebras that contain $\mathcal{C}$. There is at least one such σ-algebra: the set of all subsets of X. Show that the intersection $\mathcal{B}(\mathcal{C})$ of all the σ-algebras in $\mathbf{B}$ is again a σ-algebra, and that it contains $\mathcal{C}$. We say that $\mathcal{B}(\mathcal{C})$ is the *σ-algebra generated by $\mathcal{C}$*. ∎

5.2.29 Exercise. Use the two preceding exercises to prove Proposition 5.2.22. ∎

5.2.30 Exercise*. Adapt the construction in Example 3.1.6 of the Cantor set as follows. Fix $0 < c < 1$. At stage n of the process, instead of removing the middle third (having length 3^{-n}) of each closed interval, remove an interval of length $c3^{-n}$. Show that set that is left over is closed, nowhere dense, and that its measure is $1 - c$. This is a *generalized Cantor set*. ∎

Probability Theory

Measure theory was developed, around 1900, as foundation for the rigorous study of integration and trigonometric series. These problems being rather advanced, a beginner may find it difficult to appreciate the motivation for the abstract axioms of measure theory. Probability theory, a comparatively young subject, is an application of measure theory that invests many general ideas with simple intuitive content. To end this subsection, I want to explain Definitions 5.2.2 and 5.2.8 of σ-algebra and measure

in the language of probability. The material is set apart in small type, because most readers will find much of it a review.

In probability theory, the basic set X is called the *sample space*, and it is often written Ω rather than X. A typical element is denoted by the Greek letter ω. Each ω represents the possible "outcome" of an "experiment." The points of Ω are called *sample points*.

Example. If the experiment consists of tossing a coin three times, Ω would be the set $\{(H,H,H),(H,H,T),(H,T,H)\cdots\}$. Here H stands for *Heads*, and T stands for *Tails*. Each $\omega \in \Omega$ represents one outcome (like (H,H,T)) of the three-toss sequence. ∎

One is not usually interested in individual outcomes, but rather in collections of outcomes. Loosely speaking, a collection of outcomes (= a subset of Ω) is called an *event*.

Example. In the preceding example, the subset $\{(H,T,T),(T,H,T),(T,T,H)\}$ contains all outcomes for which exactly one Head occurs. It is the event "one Head in three tosses". ∎

In simple experiments, such as the coin-toss just described, one always can determine, by performing the experiment and recording the outcome, whether or not a particular event occurred. In more complicated experiments one can imagine subsets of Ω whose occurrence or non-occurence *cannot* be decided by observation of the outcome (unless, perhaps, the scope of the experiment is broadened). This leads to the intuitive definition:

5.2.31 Pseudo - Definition. *An event is an observable subset of the sample space Ω.* ∎

This is, of course, not terribly useful, because "observable" depends on the experiment one has in mind. Events are, instead, characterized by the properties one imagines "observability" to have:

5.2.32 Definition. *The events in a sample space Ω are the sets in a given σ-algebra $\mathcal{B}$.* ∎

Suppose given a sample space Ω of "outcomes" and a σ-algebra $\mathcal{B}$ of "events". We focus on one or more events, $A \subset \Omega, B \subset \Omega, \ldots$. The experiment is performed, and the result is a certain outcome ω. If $\omega \in A$, we say "A happened". If $\omega \notin A$, we say "A did not happen".

Let us summarize, using this language, the intuitive meaning of the simple set operations, leaving the interpretation of more complicated operations as exercises.

(5.2.33)

A^c:	A does not happen
$A \cup B$:	A or B happens, (or both happen)
$A \cap B$:	both A and B happen
$(A - B) \cup (B - A)$:	A happens, or B happens, but not both.

Now we interpret properties (5.2.3), (5.2.4), (5.2.5) in the definition of σ-algebra in this concrete setting. $\Omega, \emptyset \in \mathcal{B}$ means: we can tell whether or not *something* happened. That is plausible enough. If $A \in \mathcal{B}$, we can tell whether or not A happened. Obviously,

we then also know whether or not "not A", which is A^c, happened. So, (5.2.4) is reasonable as well. The last condition, (5.2.5), is the most delicate, because it involves infinite processes. Still, it sounds quite reasonable, too: if we can tell, for each n individually, whether or not event A_n happened, then we should also be able to tell, from our observation of the outcome of the experiment, whether or not *at least one* of the A_n happened. That latter set is $\bigcup_n A_n$, which ought therefore also belong to $\mathcal{B}$.

• **5.2.34 Exercise.** It is useful to think of a sequence $A_1, A_2, \ldots$ in $\mathcal{B}$ in the following way. An experiment is done repeatedly, at times $1, 2, 3, \ldots$. At time 1, we check whether or not A_1 has occurred, at time 2 we check whether or not A_2 occurred, *etc.*. Explain why the following interpretations are reasonable:

If $\omega \in \bigcap_{n=1}^{\infty} A_n$, then ω happens every time.

If $\omega \in \bigcup_{n=1}^{\infty} A_n$, then ω happens at least once.

If $\omega \in \liminf_{n\to\infty} A_n \stackrel{\text{def}}{=} \bigcup_{m=1}^{\infty} \left(\bigcap_{n=m}^{\infty} A_n \right)$, then ω "eventually occurs all the time."

If $\omega \in \limsup_{n\to\infty} A_n \stackrel{\text{def}}{=} \bigcap_{m=1}^{\infty} \left(\bigcup_{n=m}^{\infty} A_n \right)$, then ω "happens infinitely often." ∎

• **5.2.35 Exercise.** Prove that *all* set operations described in the preceding problem preserve membership in the σ-algebra $\mathcal{B}$. ∎

5.2.36 Example. ***Brownian motion.*** (Recall Subsection 3.6.) Let Ω be the set of all continuous functions $\omega : [0, \infty) \to \mathbf{R}$ satisfying $\omega(0) = 0$. In principle, any such function ω represents a possible motion of the diffusing particle, but we will see in this section that some types of motion are much more likely than others.

For each $T > 0$, one defines a σ-algebra $\mathcal{B}_T$ of subsets of Ω; the details are rather complicated, and won't be given, but the idea is simple. We want each set belonging to $\mathcal{B}_T$, i.e. each event in $\mathcal{B}_T$, to represent a feature of the motion whose occurrence or non-occurence can be determined from observations made from time 0 to time T. For example:

1) The set $\{\omega \mid \omega(t) = 0 \text{ for some } 0 < t < T\}$ belongs to $\mathcal{B}_T$. We can tell, by watching the particle up to time T, whether it returns to the starting point (the origin) before time T.

2) The set $\{\omega \mid \omega(t) \neq 0 \text{ for } t > \text{ some } t_0\}$ does not belong to $\mathcal{B}_T$. By observing the particle up to time T, we cannot decide whether it eventually stops returning to the origin.

3) The set $\{\omega \mid \omega(t) \neq a \text{ for } t < t_0\}$ does belong to $\mathcal{B}_T$, for any $T \geq t_0$. We can tell whether the particle got to distance a by time t_0, if we watch it for at least that long. (This is an example of a "hitting time").

4) If $T = \infty$, all these sets should belong to $\mathcal{B}_\infty$.

Of course, these statements are completely heuristic – nothing has been defined. They only show how, in certain settings, some subsets of Ω could be observable, others not. There is really nothing very exotic about the non - observability of the set in 2) above, and it is obvious that (under the condition that the particle be watched only up to time T) there is no point in assigning a "size" to a set that cannot be observed. Examples of

sets that cannot be measured are much more difficult to construct and understand in non - probabilistic settings. ∎

Continuing with probability theory, we have the sample space Ω and the (observable) events $\mathcal{B}$, but most important, each event is to be assigned a *probability*. This is a number in $[0, 1]$, interpreted as follows: if $A \in \mathcal{B}$, the probability $p(A)$ is the limit

$$\lim_{n\to\infty} \frac{\text{number of occurrences of } A \text{ in } n \text{ repetitions of the experiment}}{n}.$$

(This is again heuristic, and is intended only to relate the mathematical setting to everyday experience).

The assignment $\mathcal{B} \ni A \longmapsto p(A)$ = probability of A is supposed to be a *measure*, i.e., to satisfy requirements (5.2.8). A *probability measure* differs from other measures in that $p(\Omega) = 1$: the probability (or fraction of times) that *something* happens should be 1.

Properties (5.2.8) of "measure" are probably familiar from elementary probability theory:

a) $p(\emptyset) = 0$ means: if the experiment is performed, the probability of no outcome is zero.

b) If A_1 and A_2 are mutually exclusive events ($A_1 \cap A_2 = \emptyset$, or A_1 and A_2 can't happen simultaneously) then

$$p(A_1 \text{ or } A_2) = p(A_1 \cup A_2) = p(A_1) + p(A_2).$$

Standard probability theory also depends on the much stronger requirement

$$p\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} p(A_n) \text{ whenever } A_n \cap A_m = \emptyset,\ n \neq m. \tag{5.2.37}$$

This *does not follow* from the finite version

$$p\Big(\bigcup_{n=1}^{N} A_n\Big) = \sum_{n=1}^{N} p(A_n), \text{ if } A_n \cap A_m = \emptyset,\ n \neq m. \tag{5.2.38}$$

Property (5.2.36) (the same as (5.2.10)) is referred to as the *countable additivity* of the measure p. (5.2.37) is called *finite additivity*. (5.2.36) is by far the most subtle of the three axioms.

It should seem reasonable to require that if events $A, B, A_1, A_2, \ldots$ are observable and are assigned a probability of occurrence, then all events constructed from them in (5.2.32) and in Exercise 5.2.33 should also be observable, and have a definite probability of occurrence.

That is the basic idea behind "measurable space," "σ-algebra," and "measure."

5.2.39 Examples

A. The simplest probability that can be defined on a measurable space $(\Omega, \mathcal{B})$ is the so-called δ-measure. Let ω_0 be a fixed element of Ω, and define the measure δ_{ω_0} by:

$$\delta_{\omega_0}(A) = \begin{cases} 1, & \text{if } \omega_0 \in A, \\ 0, & \text{if } \omega_0 \notin A. \end{cases}$$

In words: the outcome ω_0 is certain to happen, and there is really nothing "random" about the experiment. An event A has probability 1 of occurring if it includes the outcome ω_0, and it has probability 0 of occurring if it does not contain ω_0.

As a special case, take $\Omega = \mathbf{R}$, and $\omega_0 = 0$. The corresponding measure, δ_0, is said to be *concentrated at* 0. It is related to the Dirac delta distribution.

B. To generalize the preceding example a little, choose n distinct sample points $\omega_1, \ldots, \omega_n$, and n positive numbers p_j such that $p_1 + \ldots + p_n = 1$. Define a probability measure by

$$p(A) = \sum_{j=1}^{n} p_j \delta_{\omega_j}.$$

Now there are n different outcomes that can happen with nonzero probability.

C. Let $\mathcal{B}$ be the Borel σ-algebra (see (5.2.22)) on $\mathbf{R}$ (with its usual topology). Let f be a continuous non-negative function on $\mathbf{R}$ for which

$$\int_{-\infty}^{\infty} f(x)\,dx = 1. \tag{5.2.40}$$

For an interval $[a, b]$, define

$$p_f([a, b]) = \int_a^b f(x)\,dx. \tag{5.2.41}$$

Then, using properties (5.2.9) and (5.2.10) as in Example 5.2.18, one can extend (5.2.40) to get a measure on the Borel σ-algebra. Because of (5.2.39), the measure p_f satisfies

$$p_f(\mathbf{R}) = \int_{-\infty}^{\infty} f(x)\,dx = 1.$$

One can therefore think of $\mathbf{R}$ as a sample space Ω, and p_f is a *(Borel) probability measure* on $\mathbf{R}$. The function f is called the *probability density* for the probability measure p_f. (An "outcome" of the "experiment" is a real number, presumably obtained by performing some kind of measurement.)

For example, when

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right),$$

the measure p_f is the so-called *normal probability measure with mean μ and standard deviation σ*. It plays a fundamental role in probability and statistics. ∎

Subsection 5.3 Application: Hausdorff Measure

Example (5.2.18) took the first few steps towards extending the usual concept of "length" to fairly general subsets of $\mathbf{R}$. Some sets, it turned out, have length zero, even though they contain lots of points: the rationals or the Cantor middle-thirds set, for instance. Sets of Lebesgue measure zero are not all equally small. The Cantor set is, in one sense at least, larger than the set of rationals: it is uncountable. Sets of measure zero on the line $\mathbf{R}$, or in higher-dimensional space $\mathbf{R}^n$, appear in numerous problems.

1) Strange attractors of many systems of differential equations have measure zero.
2) Results of complex function theory assert that certain measure zero subsets of $\mathbf{C}$ (thought of as $\mathbf{R}^2$) don't have much influence on the properties of (certain types of) analytic functions.
3) The path traced out in the plane $\mathbf{R}^2$ by a diffusing Brownian motion particle (Exercise 3.6.18) is a wiggly curve of planar Lebesgue measure zero.

It would take forever to define all the terms. I only want to point out that sets of Lebesgue measure zero arise in very different subjects, and that, furthermore the measure zero sets just listed are "wild": very wiggly curves, like Brownian motion, or fractured point sets, like the Cantor set.

The theory of Hausdorff[38] measures gives one method for studying the size of sets that are small by the standards of length in $\mathbf{R}$ (or area in $\mathbf{R}^2$, volume in $\mathbf{R}^3$). Indeed, the theory does more: it points out that — since a set of Lebesgue measure zero in $\mathbf{R}$ has length zero — Lebesgue measure may not be the correct yardstick to use on small sets. It then introduces the idea of *fractional dimension*, and an adaptation of the notion of length (or area, etc.) to non-integer dimensions allows one to measure more precisely the size of small sets.

This subsection starts with motivation, which leads into the definition of Hausdorff measure in dimension α. Some examples are done (incompletely), and then the fractional (Hausdorff) dimension of a set is shown to exist. The exercises give more examples. A calculus supplement shows how to compute the volume of a ball in α dimensions; this may be of some interest.

5.3.1 Example. In preparation for the concept of Hausdorff measure, let us use Definition 5.1.13 to prove that *the line segment* $[0, 1]$ *has Lebesgue measure zero in the plane.*

Cover the segment $[0, 1]$ by *closed* disks B_j of radius $1/N$, centered at the points j/N, $j = 0 \ldots, N$. There are $N + 1$ such disks, each having area π/N^2. The total area of the $N + 1$ disks is then $(N + 1)\pi/N^2$, which tends to zero as $N \to \infty$. According to Definition 5.1.13 (as modified in Exercise 5.1.17), the (planar) measure of $[0, 1]$ is therefore zero. ∎

---

[38] The same Hausdorff as in Hausdorff space.

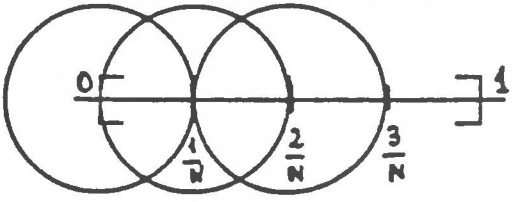


Figure 73. *A covering of* $[0, 1]$

One should understand exactly *why* the planar measure turns out to be zero. The line segment $[0, 1]$ is 1-dimensional. It can be covered by a number of disks ($N + 1$ of them, in the argument above). The number of disks necessary to cover $[0, 1]$ grows towards infinity much more slowly than the *area* π/N^2 of each disk tends to zero. Hence,

Number of disks $\times$ area per disk $\sim$ total area $\longrightarrow 0$.

If we try the same argument on the square $0 \le x, y \le 1$, we might decide to use disks centered at points $(i/N, j/N)$; there would be (at least) $(N + 1)^2$ of them, and the total area would be estimated as $(N + 1)^2\pi/N^2$, which does not tend to zero.

Hausdorff measure, roughly speaking, does the following. It recognizes that the number of disks covering $[0, 1]$ is not in balance with the area of the typical disk, and it takes various powers α of the area, until it finds an α for which

$$\sum (\text{disk area})^\alpha$$

remains finite. For example:

$$\sum_{j=0}^{N} \left(\frac{\pi}{N^2}\right)^1 \to 0;$$

$$\sum_{j=0}^{N} \left(\frac{\pi}{N^2}\right)^{1/3} \to \infty,$$

$$\sum_{j=0}^{N} \left(\frac{\pi}{N^2}\right)^{1/2} \to \text{a finite limit.}$$

This procedure, as we shall see, identifies a *dimension* of the set in question, and computes the size or *measure* of this set in the correct dimension.

The precise definition of Hausdorff measure works with the radii of the disks rather than with their area. By raising the radii to various powers, it captures the change of dimension. In dimension 1, length is proportional to radius to power 1. In dimension 2, area is proportional to radius to the power 2. And so forth. By analogy, in dimension 2/3, "volume" should be proportional to radius to the power 2/3.

5.3.2 Definition. *Let $\alpha > 0$ be given, let $A \subset \mathbf{R}^n$. Cover A by a finite or countable union of closed balls of radius $\leq \epsilon$ and let $r_1, r_2, \ldots$ be the radii of those balls. Call this an ϵ-covering. The Hausdorff measure of A in dimension α is defined to be*

$$H_\alpha(A) = \lim_{\epsilon \downarrow 0} \inf_{(\text{over all } \epsilon\text{-coverings of } A)} \sum r_n^\alpha;$$

it could be 0, ∞, or a finite number. ∎

5.3.3 Remark. This is a lie. 5.3.2 only defines *Hausdorff outer measure*. Remember that a measure requires a σ-algebra of measurable sets. The expression H_α does not assign sizes in a consistent way to arbitrary subsets of $\mathbf{R}^n$. If, however, one restricts A in the definition to be a *Borel subset* (see (5.2.22)), then one does obtain a Borel measure. Since the tools to prove this have not been developed, I plan to ignore the point altogether.

If you want to understand the intuition behind fractional dimension and Hausdorff measure, you should look at two famous books by Benoit Mandelbrot: *Fractals: Form, Chance, and Dimension* and *The Fractal Geometry of Nature*. A thorough theoretical treatment is given by K. J. Falconer, *The geometry of fractal sets*. The term *fractal* was coined by Mandelbrot; it refers not to the possibly fractional dimension, but to the "fractured" appearance of sets of fractional dimension. ∎

Definition 5.3.2 becomes more palatable when it is broken up into steps. First, let $\Gamma(\epsilon)$ be the collection of all coverings of A by closed balls B_n, $n \geq 1$ with radii $r_n \leq \epsilon$. For $\gamma \in \Gamma(\epsilon)$, set

$$H_\alpha(A, \epsilon, \gamma) = \sum_{n=1}^{\infty} r_n^\alpha.$$

Next, let

$$H_\alpha(A, \epsilon) = \inf_{\gamma \in \Gamma(\epsilon)} H_\alpha(A, \epsilon, \gamma). \tag{5.3.4}$$

5.3.5 Lemma. *The function $H_\alpha(A, \epsilon)$ is non-decreasing as $\epsilon \downarrow 0$ while α remains fixed.* ∎

Proof. Let $\epsilon_1 < \epsilon_2$. In the definition (5.3.4) of $H_\alpha(A, \epsilon_2)$, one takes an infimum over all covers by balls with radii $\leq \epsilon_2$. Since $\epsilon_1 < \epsilon_2$, those covers include *all* covers by balls of the smaller radii $\leq \epsilon_1$. When you take an infimum over a larger collection, you have more chance to make the minimum value smaller. Hence, $H_\alpha(A, \epsilon_2) \leq H_\alpha(A, \epsilon_1)$. ∎

Now, if the function $H_\alpha(A, \epsilon)$ increases (or stays constant) as ϵ decreases, it will have a limit. This limit may be ∞, it may be 0 (if all the $H_\alpha(A, \epsilon)$ are zero), or

it may be a nonzero finite value. In any case, it exists. It is denoted by $H_\alpha(A)$, the Hausdorff measure of A in dimension α.

• **5.3.6 Exercise.** Prove: If f is a function on $(0,\infty)$ that is nondecreasing as $x \downarrow 0$ (this means: $x_1 \le x_2 \Rightarrow f(x_1) \le f(x_2)$), then $\lim_{x\downarrow 0} f(x)$ exists (but it may be ∞). ∎

5.3.7 Example. Let us try to compute the Hausdorff measure in dimension α of the line segment

$$A = \{(x,y) \,|\, 0 \le x \le 1, y = 0\} \subset \mathbf{R}^2.$$

In Example 5.3.1, A was covered by $N+1$ balls of radius $1/N$. Call this cover $\gamma(1/N)$ (the ϵ in (5.3.4) is $1/N$). Evidently,

$$H_\alpha(A, 1/N, \gamma(1/N)) = \sum_{j=0}^{N} \left(\frac{1}{N}\right)^\alpha = \frac{N+1}{N^\alpha}.$$

The infimum over *all* covers (radii $\le 1/N$) can only be smaller:

$$H_\alpha(A, 1/N) \le \frac{N+1}{N^\alpha}.$$

Now let $N \to \infty$ (same as $\epsilon \downarrow 0$ in (5.3.4):

$$H_\alpha(A) \le \lim_{N\to\infty} \frac{N+1}{N^\alpha}.$$

We can conclude from this that $H_\alpha(A) = 0$ when $\alpha > 1$, that $H_1(A) \le 1$, and that $H_\alpha(A) \le \infty$ when $\alpha < 1$.

It should be obvious why $H_\alpha(A) = 0$ for $\alpha > 1$. Intuitively, we are certain that the set A has dimension 1. It stands to reason that A will look thin and small in any larger dimension, and sure enough, its Hausdorff measure is zero in dimension $\alpha > 1$, just as its Lebesgue measure was zero in dimension 2, Example 5.3.1.

It turns out that the Hausdorff measure of A in dimension $\alpha < 1$ is ∞. So far, we only know that it is $\le \infty$, but the eventual answer can be anticipated. To see what is going on, imagine a simpler picture. What would be the *one*-dimensional measure of the two-dimensional square $\{0 \le x, y \le 1\} \subset \mathbf{R}^2$? The line segment $[0,1]$ lives in dimension 1, and it imagines itself to have length 1. It realizes that infinitely many copies of itself are needed to cover the square, and the total length of all these line segments is ∞. By the standards of 1-dimensional length, the square has infinite measure. In the same way, our set $A = [0,1]$ has infinite measure by the standards of α-dimensional "length" when $\alpha < 1$.

That is not a proof that $H_\alpha(A) = \infty$ when $\alpha < 1$, of course, and actually the proof is not completely trivial. Look at what needs to be done.

One must prove that $H_\alpha(A,\epsilon) \uparrow \infty$ as $\epsilon \downarrow 0$, see Proposition 5.3.5. Thus, given $M > 0$, one must find $\eta_M > 0$ such that

$$H_\alpha(A,\epsilon) > M \text{ whenever } 0 < \epsilon < \eta_M.$$

This condition is now translated by use of (5.3.4). $H_\alpha(A.\epsilon) > M$ is the same as

$$\inf_{\gamma\in\Gamma(\epsilon)} H_\alpha(A,\epsilon,\gamma) > M,$$

and by the meaning of inf,

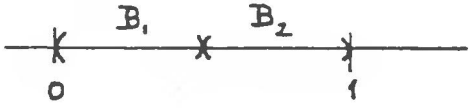
$$H_\alpha(A,\epsilon,\gamma) > M \text{ for all } \gamma \in \Gamma(\epsilon), \text{ whenever } \epsilon < \eta_M.$$

This makes sense: when ϵ is small ($< \eta_M$), it takes many balls of radius $< \epsilon$ to cover the set A. As a rough estimate, it takes $1/\epsilon$ balls. When $\alpha < 1$, then

$$\sum_{n=1}^{\infty} r_n^\alpha \sim \sum_{n=1}^{1/\epsilon} \epsilon^\alpha \sim \frac{\epsilon^\alpha}{\epsilon} \to \infty.$$

With the translation accomplished, and the statement to be proved interpreted, the rest of the argument is fairly straightforward. See Exercise 5.3.18.

There remains the problem of computing $H_1(A)$, the 1-dimensional measure of $A = [0,1]$. We know that it is ≤ 1; you probably expect it to be 1. That need not be the case — it could be some other finite number. Definition 5.3.2 does not really adhere to any "natural" standard of length or area. In fact, $H_1(A)$ turns out to be $1/2$; this only means that 5.3.2 is rescaling the usual length; if you don't like it, just put a 2 in front of the sum in the definition. [39] To see why the value $1/2$ might arise, look at this picture:



The two balls of radius $1/4$ cover A. The sum $\sum$ (radius)$^\alpha$ is $1/4 + 1/4 = 1/2$. In dimension 1, the *radius* of an interval is only $1/2$ of its *length*.

The *proof* that $H_1(A) = 1/2$ is left as exercise. ∎

---

[39] This will not fix things so that you get the "correct" area for a 2-dimensional set. To make Hausdorff measure agree with the usual (Lebesgue) measure in integer dimensions, one must put in a rather ugly correction term, which will be explained later.

5.3.8 Example. Let C be the Cantor middle thirds set, Example 5.1.16.C. In this example, we saw that for each n, the set C could be covered by 2^n *disjoint* closed balls, each of radius $(1/2)3^{-n}$. Call this cover γ_n, and let ϵ_n be the common value of the radius. Then, as in Example 5.3.7,

$$H_\alpha(C, \epsilon_n) \leq H_\alpha(C, \epsilon_n, \gamma_n) = \sum_{k=1}^{2^n} \left(\frac{1}{2 \cdot 3^n}\right)^\alpha = \frac{2^n}{2^\alpha 3^{\alpha n}}. \tag{5.3.9}$$

Now let $n \to \infty$, which corresponds to $\epsilon_n \to 0$. The expression on the right of (5.3.9) can be written

$$\frac{1}{2^\alpha}\left(\frac{2^{1/\alpha}}{3}\right)^{\alpha n} \tag{5.3.10}$$

The limit of (5.3.10) is 0 if $2^{1/\alpha} < 3$. For such α, the Hausdorff measure is $H_\alpha(C) = 0$. When $2^{1/\alpha} > 3$, the limit is ∞, and we cannot conclude anything about $H_\alpha(C)$ without further argument, just as in Example 5.3.7; it again turns out, however, that $H_\alpha(C) = \infty$ for those α. Finally, the limit is finite when $2^{1/\alpha} = 3$. Solve for α to get $\alpha = \ln 2 / \ln 3 = .6309\ldots$. It can be proved that $H_\alpha(C)$ is finite in this case. One then concludes that *the Cantor middle thirds set has dimension* $\ln 2 / \ln 3$. The value of $H_{.6309\ldots}(C)$ will not be computed (it is not hard to guess). ∎

5.3.11 Example. The two preceding examples dealt with nice sets that were easy to cover with closed intervals. This example goes to the opposite extreme: take any $A \subset \mathbf{R}^n$ whatsoever. When $\alpha > n$, then A will "look small", because it lies in a space of dimension lower than α. More precisely,

$$H_\alpha(A) = 0 \text{ for all } \alpha > n.$$

Nothing can be said in this generality about $\alpha \leq n$, since A could equally well be a single point or all of $\mathbf{R}^n$.

Not all details of the proof can be given; the missing steps require some (plausible) applications of results from the theory of outer measures (cf. Remark 5.3.3).

First, let C be the unit cube, $\{(x_1, \ldots, x_n) \mid 0 \leq x_i \leq 1\}$. Choose $\epsilon > 0$, and pick N so that $\sqrt{n}/N < \epsilon$. Cover C by $(N+1)^n$ balls of radius $\sqrt{n}/N$ centered at the points $(m_1/N, \ldots, m_n/N)$, with integers m_i between 0 and N. Call this cover γ_N.

We have

$$H_\alpha(C, \epsilon, \gamma_N) = \sum_{n=1}^{(N+1)^n} r_n^\alpha = (N+1)^n \times \left(\frac{\sqrt{n}}{N}\right)^\alpha;$$

as long as $\alpha > n$,

$$= (\sqrt{n})^\alpha \frac{(N+1)^n}{N^\alpha} \to 0$$

as $N \to \infty$. Because $H_\alpha(C, \epsilon, \gamma_N) \to 0$ as $N \to \infty$, we have $H_\alpha(C) = 0$. This much of the argument just reproduces a portion of Example 5.3.7, with $[0, 1]$ generalized to C.

One of the virtues of a good theory (like measure theory) is that it lets you deduce things without excessive calculation. Observe: if A is any subset of the unit cube C, then by (5.2.12)

$$H_\alpha(A) \le H_\alpha(C) = 0. \tag{5.3.12}$$

(Strictly speaking, A should be a Borel set, but see below.) Next,, if $C_{\mathbf{m}}$ is the unit cube with lowest left corner at $\mathbf{m} = (m_1, \dots, m_n)$, then of course $H_\alpha(C_{\mathbf{m}}) = 0$ by the argument given above, and if $A \subset \mathbf{R}^n$ is arbitrary, we may write

$$A = \bigcup_{\mathbf{m}} (A \cap C_{\mathbf{m}}),$$

and then

$$H_\alpha(A) = H_\alpha\left(\bigcup_{\mathbf{m}} A \cap C_{\mathbf{m}}\right) \le \sum_{\mathbf{m}} H_\alpha(A \cap C_{\mathbf{m}}) = \sum_{\mathbf{m}} 0 = 0. \tag{5.3.13}$$

The results (5.2.12) and (5.2.17) needed to establish (5.3.12) and (5.3.13) were proved earlier, but only for *measures*; they turn out to be true for *outer measures* as well — but you will have to take my word for that. ∎

We return to the general properties of Hausdorff measure and fractional dimension. Examples 5.3.7 and 5.3.8 suggest that $H_\alpha(A)$ is either zero or infinity, except perhaps for one particular value of the dimension α. This is indeed the case, and a complete proof can be given. Note that no properties of measures are used in the next proposition: the argument establishes the existence of a unique dimension for every subset of $\mathbf{R}^n$.

5.3.14 Proposition. *Let $A \subset \mathbf{R}^n$. There is a unique $\alpha_0 \in [0, \infty)$ such that*

$$H_\alpha(A) = \begin{cases} \infty & \text{if } 0 \le \alpha < \alpha_0, \\ 0 & \text{if } \alpha > \alpha_0. \end{cases}$$

*This α_0 is called the **Hausdorff dimension** of A.* ∎

Proof. I claim that if $\alpha < \beta$, then

$$H_\alpha(A, \epsilon) \ge \epsilon^{\alpha - \beta} H_\beta(A, \epsilon). \tag{5.3.15}$$

Accept this for the moment, and analyze the consequences.

Suppose first that $\lim_{\epsilon \downarrow 0} H_\beta(A,\epsilon) > 0$, i.e. that $H_\beta(A) > 0$. Then, since the exponent $\alpha - \beta$ in (5.3.15) is negative, the right side tends to ∞. It follows that $\lim_{\epsilon \downarrow 0} H_\alpha(A,\epsilon) = \infty$, whence $H_\alpha(A) = \infty$.

Obviously, if $H_\beta(A) = \infty$, then $H_\alpha(A) = \infty$ as well.

Suppose next that $\lim_{\epsilon \downarrow 0} H_\alpha(A,\epsilon) = H_\alpha(A) < \infty$. Then (5.3.15) shows that $H_\beta(A)$ must be zero (or else the right side would tend to ∞ as $\epsilon \downarrow 0$, leading to the nonsensical conclusion $H_\alpha(A) \geq \infty$).

We can now define α_0. Let $\mathcal{A}_\infty$ be the set of all α for which $H_\alpha(A) = \infty$, and let $\mathcal{A}_0$ be the set of α for which $H_\alpha(A) = 0$. We have shown that $\mathcal{A}_\infty < \mathcal{A}_0$, in the sense that every element of the first set is $<$ every element of the second set. The special number α_0 is the "boundary" between these two sets, i.e.,

$$\alpha_0 = \inf \mathcal{A}_0, \quad \text{or } \alpha_0 = \sup \mathcal{A}_\infty. \tag{5.3.16}$$

As is explained in Remark 5.3.17 following this proof, the set $\mathcal{A}_\infty$ may be empty; in that case, $\alpha_0 = 0$. On the other hand (Example 5.3.11), $\mathcal{A}_0$ is not empty, and so α_0 cannot be ∞.

Let us now prove (5.3.15). For a cover $\gamma \in \Gamma(\epsilon)$, we have defined

$$H_\alpha(A,\epsilon,\gamma) = \sum_{n=1}^{\infty} r_n^\alpha = \sum_{n=1}^{\infty} r_n^\beta (r_n^{\alpha-\beta});$$

since $r_n \leq \epsilon$ and $\alpha - \beta < 0$, the estimate

$$r_n^{\alpha-\beta} \geq \epsilon^{\alpha-\beta}$$

holds, whence

$$H_\alpha(A.\epsilon.\gamma) \geq \epsilon^{\alpha-\beta} \sum_{n=1}^{\infty} r_n^\beta = \epsilon^{\alpha-\beta} H_\beta(A,\epsilon,\gamma).$$

Now take the infimum over $\gamma \in \Gamma(\epsilon)$ (see (5.3.4)) to get (5.3.15). ∎

5.3.17 Remark. Let A be a set containing just one point. Then A can be covered by a single ball of radius $r \leq \epsilon$, and the sum in Definition 5.3.2 reduces to

$$\sum_{n=1}^{1} r_n^\alpha = r^\alpha \leq \epsilon^\alpha.$$

It follows that $H_\alpha(A.\epsilon) = 0$ if $\alpha > 0$ and, when $\alpha = 0$, $H_0(A,\epsilon) = 1$. Thus, it is possible to have *no* α for which $H_\alpha = \infty$. ∎

• **5.3.18 Exercise.** Finish the proof that $H_\alpha([0,1]) = \infty$ for $\alpha < 1$, following the outline in Example 5.3.7. You will need to use properties of ordinary length (Lebesgue measure) on $[0,1]$. ∎

• **5.3.19 Exercise.** Let C be the Cantor middle thirds set. What is a reasonable guess for the value of $H_\alpha(C)$ when $\alpha = \ln 2/\ln 3$? ∎

• **5.3.20 Exercise.** Define the Cantor middle fifth set C_5 in the obvious way: at stage 1, remove the middle fifth from $[0,1]$, at stage 2, remove the middle fifth from each of the remaining two intervals, and so on. Mimic Example 5.3.8 to find a guess for the dimension of that set. ∎

5.3.21 Exercise. Define a set $A \subset \mathbf{R}^2$ as follows. Start with the unit square S. At step 1, remove everything but the four closed corner squares of side 1/4. At step 2, remove everything from each of those squares but its four closed corner squares of side 1/16. Continue in this way. Make a guess at the Hausdorff dimension of the resulting set. ∎

5.3.22 Exercise*. Let f be a continuously differentiable function defined on $[0,1]$. Prove: the curve

$$C = \{(x, f(x)) \,\big|\, x \in [0,1]\}$$

has Hausdorff dimension 1.

Here is an outline of the argument; you should fill in the details. Cover C with closed balls B_n whose radii are $r_n \le \epsilon$. Let π be the projection from C to $[0,1]$, $\pi : (x, f(x)) \mapsto x$. The intervals $\pi(B_n)$ cover $[0,1]$, so $\sum 2r_n \ge 1$. Hence $H_1(C) \ge 1$.

Conversely, one must show that for every $\eta > 0$, $H_{1+\eta}(C) = 0$. Divide $[0,1]$ into N intervals of length $\le 1/N$. On such an interval, $|f(x) - f(x')| \le K/N$, where K is a constant determined from the Mean Value Theorem. Hence,

$$\|(x, f(x)) - (x', f(x'))\|_\infty \le \frac{1}{N} + \frac{K}{N} = \frac{K'}{N}. \tag{5.3.23}$$

The square (5.3.23) can be enclosed in a disk of radius $\sqrt{2}K'/N$. For this cover, one finds

$$\sum_{n=1}^{N} \left(\frac{\sqrt{2}K'}{N}\right)^{1+\eta} = N \times \frac{(\sqrt{2}K')^{1+\eta}}{N^{1+\eta}},$$

which tends to zero as $N \to \infty$. ∎

5.3.24 Exercise*. I was told that the following example was described on the TV series Nova; I don't know who invented it.

Draw a triangle with vertices A, B, and C. Pick a point P in the triangle. Roll a (fair) die. If the number is 1 or 2, move halfway along the line from P to A. If the number is 3 or 4, move halfway towards B, and if the number is 5 or 6, move halfway towards C. Call the new point P', and start all over again. For most choices of starting points P, this procedure will fill in triangle ABC with a very complicated, "fractured" pattern.

Program a computer to run this experiment. Can you estimate the Hausdorff dimension of the resulting set? (An interesting question: is this dimension fixed, or does it vary randomly with choice of the starting point and roll of the die?) ∎

5.3.25 SUPPLEMENT: Volume of the ball in fractional dimension

We conclude this section with a calculation that is really only of secondary importance in the theory of Hausdorff measure, but is a standard example in analysis with which you should be familiar. Namely, we compute the volume ω_α of the unit sphere in dimension α. One can then (as do some authors) redefine Hausdorff measure by setting

$$H_\alpha(A, \epsilon, \gamma) = \sum_{n=1}^{\infty} \omega_\alpha r_n^\alpha.$$

This modification has some small advantage: for "ordinary" sets, such as $[0,1] \subset \mathbf{R}^2$, the Hausdorff measure then agrees with the usual value of length, area, volume, etc. On the other hand, the measure of "exotic" sets like the Cantor set becomes significantly more ugly.

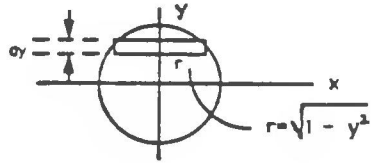
We first obtain a formula for the volume ω_n of the unit ball in n dimensions. (The derivation is not the slickest one possible, but it is elementary.) A clever construction then provides a formula that applies even when the dimension is not an integer.

The "unit sphere" in $\mathbf{R}$ is the interval $[-1,1]$, and its "volume" (length, in this case) is $\omega_1 = 2$. The volume of a ball of radius r is then $\omega_1 r^1$.

The "volume" (=area) of the unit ball (=disk) in two dimensions is given by the integral

$$\int_{-1}^{1} \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} 1 \, dx \, dy = 2 \int_{-1}^{1} \sqrt{1-y^2} \, dy.$$

The last integral has the following interpretation.



The unit disk is cut into slices of thickness dy and radii $\omega_1(\sqrt{1-y^2})^1$; the 1-dimensional volumes of these slices are added up from $y = -1$ to $y = +1$.

More generally, to find the volume of the unit ball in $\mathbf{R}^{n+1}$, we introduce coordinates $(x_1, \ldots, x_n, y)$, and cut the $n+1$-ball into slices of thickness dy, so that the cross-sections of the slices are n-dimensional balls of radii $\sqrt{1-y^2}$. Their n-dimensional volumes are $\omega_n(\sqrt{1-y^2})^n$, since the volume of an n-ball of radius r is $\omega_n r^n$. The $n+1$-dimensional volumes of the slices, $\omega_n(\sqrt{1-y^2})^n \, dy$, are then added up from $y = -1$ to $y = +1$:

$$\omega_{n+1} = \omega_n \int_{-1}^{1} (\sqrt{1-y^2})^n \, dy.$$

Introduce $y = \sin\theta$ in this integral, and write

$$\omega_{n+1} = 2\omega_n \int_0^{\pi/2} \cos^{n+1}\theta \, d\theta.$$

Let $c_n = \int_0^{\pi/2} \cos^n \theta \, d\theta$. We obtain a *recursive formula* for the volume ω_j, namely

(5.3.26) $$\omega_{n+1} = 2c_{n+1}\omega_n.$$

We still need to find a closed formula, giving the value of ω_{n+1} directly.

First, an integration by parts relates c_{n+1} and c_{n-1}:

$$\int_0^{\pi/2} \cos^{n+1}\theta \, d\theta = \int_0^{\pi/2} \cos^n\theta \left(\frac{d}{d\theta}\sin\theta\right) d\theta$$

$$= n \int_0^{\pi/2} \cos^{n-1}\theta \, \sin^2\theta \, d\theta = nc_{n-1} - nc_{n+1},$$

whence

$$c_{n+1} = \frac{n}{n+1} c_{n-1}.$$

From this, we see that

(5.3.27) $$c_n = \begin{cases} \frac{\pi}{2} \cdot \frac{1}{2} \cdot \frac{3}{4} \cdots \frac{n-1}{n}, & \text{if } n \text{ is even,} \\ \frac{2}{3} \cdot \frac{4}{5} \cdot \frac{6}{7} \cdots \frac{n-1}{n}, & \text{if } n \text{ is odd.} \end{cases}$$

• **5.3.28 Exercise.** Use (5.3.26) and (5.3.27) to find the volumes of the unit balls in dimensions 2, 3, 4, starting with $\omega_1 = 2$. ∎

It would be possible to produce a formula for ω_{n+1} after some experimentation with (5.3.26) and (5.3.27), but such a formula would not generalize to fractional dimensions. Therefore, we proceed in a different and much less straightforward way.

The basic ingredient will be the *Gamma function*, which is defined for $x > 0$ by

(5.3.29) $$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} \, dt.$$

(See (4.1.9).) Integration by parts shows that

(5.3.30) $$\Gamma(x+1) = x\Gamma(x).$$

• **5.3.31 Exercise.** Prove 5.3.30. ∎

By evaluation of the integral in (5.3.29) one finds that $\Gamma(1) = 1$. Then (5.3.30) gives

$$\Gamma(2) = \Gamma(1+1) = 1 \cdot \Gamma(1) = 1,$$
$$\Gamma(3) = \Gamma(2+1) = 2 \cdot \Gamma(2) = 2.$$
$$\Gamma(4) = \Gamma(3+1) = 3 \cdot \Gamma(3) = 2 \cdot 3,$$

and so on:

$$\Gamma(n+1) = n!.$$

Thus, the Gamma function at integer arguments $n+1$ coincides with the factorial function, and one may consider the Gamma function at other, noninteger, arguments to be an interpolation of the factorial function.

For example, one has

$$\Gamma(\tfrac{1}{2}) = \int_0^\infty e^{-t} \frac{1}{\sqrt{t}}\, dt = 2\int_0^\infty e^{-u^2}\, du = \sqrt{\pi}.$$

and then

$$\Gamma(\tfrac{3}{2}) = \frac{\sqrt{\pi}}{2}, \quad \Gamma(\tfrac{5}{2}) = \frac{3\sqrt{\pi}}{4},$$

and so forth:

$$\Gamma(\frac{2n+1}{2}) = \frac{2n-1}{2} \cdots \frac{3}{2} \cdot \frac{1}{2}\sqrt{\pi}.$$

I now claim that the volume of the n-dimensional unit ball is given by the formula

$$\omega_n = \frac{(\sqrt{\pi})^n}{\Gamma(\frac{n+2}{2})}. \tag{5.3.32}$$

In view of (5.3.26), it suffices to prove that the quantities ω_n defined in (5.3.32) satisfy

$$c_{n+1} = \frac{1}{2}\frac{\omega_{n+1}}{\omega_n}, \tag{5.3.33}$$

where c_{n+1} is given by (5.3.27). This is a useful exercise in manipulation with the Gamma function, and is left to the reader.

• **5.3.34 Exercise.** Verify (5.3.33), and hence establish (5.3.32). ∎

We have reached the goal. Formula (5.3.32) for the volume of the unit ball makes sense even when the "dimension" n is taken to be an arbitrary non-negative number. The volume of a ball of radius r in dimension α is *defined to be*

$$\frac{\pi^{\alpha/2}}{\Gamma(\frac{\alpha+2}{2})} r^\alpha,$$

and the definition of Hausdorff measure in dimension α is changed to

$$\lim_{\epsilon \downarrow 0} \inf_{\Gamma(\epsilon)} \sum_{n=1}^\infty \frac{\pi^{\alpha/2}}{\Gamma(\frac{\alpha+2}{2})} r_n^\alpha. \tag{5.3.35}$$

With this revised definition, the N-dimensional unit ball S will now have n-dimensional Hausdorff measure $H_N(S) = \omega_N$, whereas with the original definition (5.3.2) we would have had $H_N(S) = 1$. This is perhaps not a big deal, but some authors adopt the normalization in (5.3.35), and one might as well know where it comes from. Most important, all mathematicians should see the computation of the volumes of higher-dimensional balls just once.

• **5.3.36 Exercise.** Find the volume of the n-dimensional ellipsoid

$$\frac{x_1^2}{a_1^2} + \ldots + \frac{x_n^2}{a_n^2} = 1. \blacksquare$$

5.3.37 Exercise. Find a formula for the *surface area* σ_n of the unit sphere in n dimensions. E.g., $\sigma_2 = 2\pi, \sigma_3 = 4\pi$ etc. ∎

Subsection 5.4 Application: Brownian motion

Subsection 3.6 developed topological properties of the space Ω of Brownian motion sample paths. The sets of the form

$$U(t_0, a, b) \stackrel{\text{def}}{=} \{\omega \,|\, \omega(t_0) \in (a, b)\} \tag{5.4.1}$$

were declared to be open, and (in accordance with the concept of *subbase for a topology*) arbitrary unions of finite intersections of these basic sets form the general open subsets of Ω. The goal of the theory of Brownian motion is to make predictions about the typical behavior of a diffusing particle, and to this end one must also introduce a *probability measure* which will tell one how "likely" are the motions described by a set of paths such as $U(t_0, a, b)$.

In this subsection, we first define the probability measure on the basic sets (5.4.1) and their finite intersections. Theorems from measure theory guarantee that this definition will extend to the σ-algebra of Borel sets of Ω (see (5.2.22)). These theorems will not be proved. We then give a few examples of the kinds of conclusions one draws from the probability measure. Calculations and proofs involving Brownian motion probability tend to be rather intricate, drawing not only on general measure theoretic results and techniques but also on subtle estimates peculiar to Brownian motion. It will not be possible to develop such a large mathematical arsenal in the limited space available.

Probability theory is, in a sense, a branch of measure theory. After all, a measure P on a measurable space $(X, \mathcal{B})$ becomes a probability measure as soon as $P(X) = 1$. Yet, probability theory has a very distinctive flavor, which is due to ideas and constructions that make sense in a general measure theoretic context but are really only used by probabilists. The most important special probabilistic idea is the concept of *independence* of events. It is fundamental to the theory of Brownian motion, and we must explain it.

5.4.2 Definition. *Let $(X, \mathcal{B}, P)$ be a probability measure space. Two events $A, B \in \mathcal{B}$ are said to be independent if $P(A \cap B) = P(A)P(B)$. More generally, events $A_1, A_2, \ldots$ are said to be independent if $P(A_{i_1} \cap \ldots \cap A_{i_m}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_m})$ for every finite subcollection of the set of events $\{A_n\}$.* ∎

If one thinks of $P(A)$ (resp. $P(B)$) as "the fraction of times A happens" (resp. B happens), then the definition of independence, rewritten as

$$\frac{P(A \cap B)}{P(A)} = P(B),$$

says: of the times when A happens, the fraction of occurrences of B is the same as the overall frequency of B. Or: B does not become more or less likely once A has occurred. Or: occurrence of A does not affect occurrence (or non-occurrence) of B. Or: A and B are independent.

• **5.4.3 Exercise.** Show that if A and B are independent and also mutually exclusive ($A \cap B = \emptyset$), then one of $P(A)$ or $P(B)$ must be 0. Intuitively: if A and B are mutually exclusive, and if A occurs, then it is almost certain that B cannot occur. ∎

• **5.4.4 Exercise.** Let $X = \{a, b, c, d\}$, and assign each point the probability $1/4$. Let $A = \{a, d\}$, $B = \{b, d\}$, and $C = \{c, d\}$. Show that any two of the sets A, B, C are independent, but that the three sets A, B, C are *not* independent. Hence, events may be independent in pairs without being independent. ∎

5.4.5 Exercise*. Suppose that $A_1, \ldots, A_n$ are independent. Show that the complements $A_1^c, \ldots, A_n^c$ are also independent. ∎

5.4.6 Example. In elementary probability theory, one studies *Bernoulli* trials (also called *success - failure trials*). Here is a simple example. Toss a fair six-sided die infinitely often. Each toss is called a "trial." The outcome of one trial is called a "success" if the die shows "6", and "failure" if the die shows any other number. The probability p of success on any given trial is $1/6$, and the probability q of failure is $5/6$. It is reasonable to assume that the outcome of one trial does not influence the outcome of any other trial. For instance, the events A="success on first trial" and B="success on second trial" are considered to be independent. Hence, by (5.4.2), their intersection ("success on the first two trials") should have probability $(1/6) \times (1/6)$.

The general setting for Bernoulli trials is described next. Let X be the set of sequences $(\omega_0, \omega_1, \ldots)$ of 0's and 1's. An $\omega \in X$ is thought of as a sequence of successive trials, each of which can end either in "success" ($\omega_i = 1$) or "failure" ($\omega_i = 0$). This model was introduced[40] in Example 1.1.45, and its topology was described in Exercise 3.4.16. The topology will be important, so it is reviewed briefly.

---

[40] Earlier, the index set was Z, and now it is N; this is a minor point.

A base for the topology on X is given by the collection of sets

$$U_{\epsilon_1,\ldots,\epsilon_r} \stackrel{\text{def}}{=} \{\omega \,|\, \omega_{i_1} = \epsilon_1, \ldots, \omega_{i_r} = \epsilon_r\}, \tag{5.4.7}$$

indexed by all the finite sequences $(\epsilon_1, \ldots, \epsilon_r)$ of 0's and 1's and $i_1 < \ldots < i_r$ of nonnegative integers. These sets are analogs of the Brownian motion window sets (3.6.3), as is illustrated in Figure 74. This topology is the smallest one in which all projection functions π_i, defined by $\pi_i(\omega) = \omega_i$, are continuous when the range $Y = \{0, 1\}$ of the π_i is given the discrete topology (in which all subsets are open). See Proposition 3.4.19b. In less forbidding words: we certainly want to be able to pick an arbitrary trial — # i, say — and, for this trial, distinguish outcome 1 from outcome 0. This obvious requirement, when expressed in the language of topology, forces X and Y to carry the topologies just introduced.

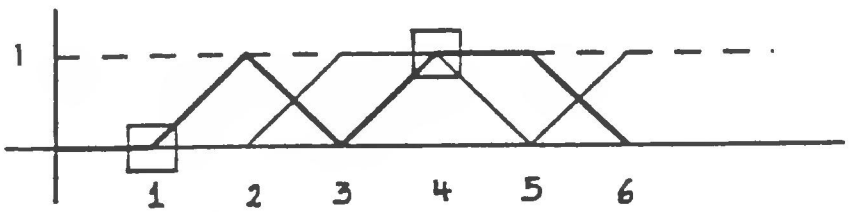


Figure 74. *ω's passing through the windows $\omega_1 = 0, \omega_4 = 1$*

We wish to define a probability measure that incorporates three features:

a) $P($ success $) = p$; $P($ failure $) = q$ (with $p + q = 1$);
b) the trials are independent;
c) the window sets (5.4.7) should be events.

Write $\{\omega_i = 1\}$ for the event $\{\omega \in X \,|\, \omega_i = 1\}$ (the i^{th} trial is a success), and likewise write $\{\omega_i = 0\}$ for the event "the i^{th} trial is a failure". Require that

$$P(\omega_i = 1) = p, \quad P(\omega_i = 0) = q. \tag{5.4.8}$$

This is feature a). Independence is expressed by defining, for each pair of sequences

$$(\epsilon_1, \ldots, \epsilon_r), \quad (i_1, \ldots, i_r),$$

the probability of the window set (5.4.7) to be

$$P(\omega_{i_1} = \epsilon_1, \ldots, \omega_{i_r} = \epsilon_r) = \prod_{k=1}^{r} P(\omega_{i_k} = \epsilon_k) = p^{\#\text{ of }1\text{'s}} \cdot q^{\#\text{ of }0\text{'s}}. \tag{5.4.9}$$

Probability and topology interact in an interesting way. We already know that the window sets are a base for the only reasonable topology on X. It turns

out that once probabilities are defined consistently on these window sets, they extend (consistently) to the σ-algebra generated by the window sets. (According to Proposition 5.2.22, this will be the Borel σ-algebra corresponding to the natural topology on X.) There is now a small-print digression on the technical aspects of this extension. The reader may just pick up the discussion at the other end. ∎

Window sets were defined as in (5.4.7) in order to make the notation in (5.4.9) simple. For the purpose of checking consistency of probabilitites, however, (5.4.7) is a little too limited.

Example. Here is a typical consistency check. Prescription (5.4.9) gives

$$P(\omega_1 = 1, \omega_2 = 0) = p^1 q^1 \quad \text{and} \quad P(\omega_1 = 1, \omega_2 = 1) = p^2 q^0. \tag{5.4.10}$$

Prescription (5.4.8) gives

$$P(\omega_1 = 1) = p. \tag{5.4.11}$$

But

$$\{\omega_1 = 1\} = \{\omega_1 = 1, \omega_2 = 0\} \cup \{\omega_1 = 1, \omega_2 = 1\}, \tag{5.4.12}$$

and the sets on the right side are disjoint, so one should have

$$P(\omega_1 = 1) = P(\omega_1 = 1, \omega_2 = 0) + P(\omega_1 = 1, \omega_2 = 1). \tag{5.4.13}$$

It is easy to check directly that (5.4.13) is satisfied:

$$pq + p^2 = p(p + q) = p \cdot 1 = p,$$

but there are many more complicated relations of this type that must all be checked. They impose severe restraints on one's creativity: the probabilities of window sets cannot be postulated at will. ∎

The key feature of the consistency check is this: as illustrated by (5.4.12), one must consider unions of window sets, and in the process, some windows may be "lost". (On the right side of (5.4.12), there are constraints on ω_2; on the left side, there are none.) One therefore generalizes *window set* to *cylinder set*, by allowing windows that include *both* 0 and 1 or *neither* 0 nor 1.

Definition. *Let $A_1, \ldots, A_r$ be subsets of $\{0, 1\}$, and let $\mathbf{S} = \{i_1, \ldots, i_r\}$ be a finite set of integers. A set of the form*

$$\{\omega \,|\, \omega_{i_1} \in A_1, \ldots, \omega_{i_r} \in A_r\}$$

is called a **cylinder set**. *The collection of all cylinder sets with fixed $\mathbf{S}$ is denoted by $\mathcal{Z}_{\mathbf{S}}$.* ∎

A cylinder set belongs to infinitely many $\mathcal{Z}_{\mathbf{S}}$. For instance, $\{\omega_1 = 1\}$ belongs to $\mathcal{Z}_{\{1\}}$, but since

$$\{\omega_1 = 1\} = \{\omega \mid \omega_1 \in \{1\}, \omega_2 \in \{0, 1\}\},$$

it also belongs to $\mathcal{Z}_{\{1,2\}}$. It also belongs to

$$\mathcal{Z}_{\{1,2,3\}}, \ldots, \mathcal{Z}_{\{1,2,\ldots,10\}}, \ldots$$

This definition provides a convenient formalism for dealing with relations like (5.4.12).[41] The consistency argument for Bernoulli trials proceeds along the following lines. If $A \in \mathcal{Z}_{S_1}$ and $A \in \mathcal{Z}_{S_2}$, one shows that also $A \in \mathcal{Z}_{S_1 \cup S_2}$, and that the probabilities assigned by (5.4.9) are the same for all three index sets S. The details are omitted.

Once it is established that P is consistent on cylinder sets, one appeals to a theorem from general measure theory. A collection $\mathcal{S}$ of subsets of a set X is called a *semialgebra* if it has the following properties:

i) $\emptyset \in \mathcal{S}$;

ii) if $A, B \in \mathcal{S}$ then $A \cap B \in \mathcal{S}$;

iii) if $A, B \in \mathcal{S}$ and $A \subset B$, there are disjoint sets $C_1, \ldots, C_r \in \mathcal{S}$ such that $B - A = C_1 \cup \ldots \cup C_r$.

The extension theorem [42] says that a finitely additive measure (see (5.2.38)) on a semialgebra $\mathcal{S}$ extends in a unique way to a countably additive measure (see (5.2.37)) on the σ-algebra generated by $\mathcal{S}$.

To apply this result in the Bernoulli setting, one checks that the collection of all cylinder sets forms a semi-algebra and that the probability measure assigned to cylinder sets by (5.4.8) and (5.4.9) satisfies (5.2.38). One then knows that it extends to a unique Borel measure on the sequence space X.

The same sort of argument leads, in much more general settings, to the *Kolmogorov extension theorem* which proves existence of many important stochastic processes.[43]

Since independence is such an important concept, we present another example. Recall (Exercise 5.2.34) that if $\{A_n\}$ is a sequence of events, then

$$\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$$

is the event "A_n happens infinitely often," abbreviated below as "A_n i.o." A somewhat technical result is proved first, and its implications (which are quite striking) are discussed afterwards.

5.4.14 Proposition (Borel-Cantelli Lemma). *Let $(X, \mathcal{B}, P)$ be a probability measure space, and let $\{A_n\}$ be a sequence of events.*

---

[41] The term *cylinder set* arises as follows. If you have a function $f : X \to Y$, and picture Y along the horizontal axis and X along the vertical axis, then, for $U \subset Y$, the set $f^{-1}(U)$ looks like a cylinder with base U, extending into the X direction.

[42] See A. N. Kolmogorov and S. V. Fomin, *Elements of the Theory of Functions and Functional Analysis*.

[43] K. Krickeberg, *Probability Theory*.

a) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, *then* $P(A_n \text{ i.o.}) = 0$.

b) If the A_n *are independent and* $\sum_{n=1}^{\infty} P(A_n) = \infty$, *then* $P(A_n \text{ i.o.}) = 1$. ∎

Proof. a) $P(A_n \text{ i.o.}) = P(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n)$; by Proposition 5.2.13, this is equal to the limit

$$\lim_{m\to\infty} P\left(\bigcup_{n=m}^{\infty} A_n\right).$$

which, by Exercise 5.2.17, is

$$\leq \lim_{m\to\infty} \sum_{n=m}^{\infty} P(A_n).$$

and that limit is zero because the series $\sum P(A_n)$ converges.

b) From the proof of part a), we know that

$$P(A_n \text{ i.o.}) = \lim_{m\to\infty} P\left(\bigcup_{n=m}^{\infty} A_n\right).$$

By de Morgan's laws and Lemma 5.2.11b),

$$P\left(\bigcup_{n=m}^{\infty} A_n\right) = 1 - P\left(\bigcap_{n=m}^{\infty} A_n^c\right).$$

It suffices to prove that

$$\lim_{m\to\infty} P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 0. \tag{5.4.15}$$

In fact, more is true: I claim that for every m,

$$P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 0. \tag{5.4.16}$$

If every term of a sequence is zero, its limit is of course also zero, so that (5.4.15) follows from (5.4.16).

To establish (5.4.16), I will first show that (for fixed m)

$$\lim_{N\to\infty} P\left(\bigcap_{n=m}^{N} A_n^c\right) = 0. \tag{5.4.17}$$

Then (5.4.15) follows by an application of Proposition (5.2.13). By independence of the A_n^c (Exercise 5.4.5), we have

$$P\Big(\bigcap_{n=m}^{N} A_n^c\Big) = \prod_{n=m}^{N} P(A_n^c) = \prod_{n=m}^{N} (1 - P(A_n)).$$

Now comes a clever step. The Taylor expansion

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$$

shows that $\log(1-x) \le -x$ when $0 < x < 1$. Hence,

$$\log\Big(\prod_{n=m}^{N} (1 - P(A_n))\Big) = \sum_{n=m}^{N} \log(1 - P(A_n)) \le -\sum_{n=m}^{N} P(A_n).$$

Since the series $\sum P(A_n)$ diverges,

$$\lim_{N\to\infty} -\sum_{n=m}^{N} P(A_n) = -\infty,$$

whence

$$\lim_{N\to\infty} \log\Big(\prod_{n=m}^{N} (1 - P(A_n))\Big) = -\infty,$$

and so

$$\lim_{N\to\infty} \prod_{n=m}^{N} (1 - P(A_n)) = 0.$$

This establishes (5.4.17) and with it, the proposition. ∎

The two halves of the Borel-Cantelli lemma, taken together, have a quite striking interpretation.

5.4.18 Corollary (Borel zero-one law). *Let $\{A_n\}$ be a sequence of independent events. Then the event A_n i.o. can only have probability zero or one.* ∎

Proof. The sum $\sum P(A_n)$ either converges or diverges. ∎

A *zero-one law* is a theorem that says: a certain event has either probability zero (it happens almost never) or probability one (it happens almost surely.)

5.4.19 Examples

A. The curious phrases "almost never" and "almost surely" in the last sentence deserve explanation. In elementary probability theory, when $P(A) = 0$, then

A never happens. This is peculiar to the sample spaces with finitely many outcomes that are the staple of beginning probability; confusion may be further encouraged by imprecise use of mathematical concepts.

Let A_n be the event: the n^{th} trial results in success, i.e. $A_n = \{\omega_n = 1\}$. The events A_n are independent by the very construction of the probability measure. Since $P(A_n) = p$, the series $\sum P(A_n)$ diverges, and by (5.4.18) the event A_n i.o. has probability 1. One says: "Almost surely, every sequence ω of Bernoulli trials has infinitely many successes." It is necessary to insert the qualification "almost", because there are sequences that do *not* have infinitely many successes, such as the sequence consisting entirely of 0's. Those sequences, however, form a subset of X of (probability) measure 0. Roughly speaking: if you ask N people to perform one sequence of Bernoulli trials each, then these exceptional sequences may occur, but the proportion of such occurrences is vanishingly small as $N \to \infty$.

B. Let Y be a fixed finite sequence of 0's and 1's, $Y = (\epsilon_1, \dots, \epsilon_k)$. Let A_n be the event

$$\omega_{nk+1} = \epsilon_1, \omega_{nk+2} = \epsilon_2, \dots, \omega_{nk+k} = \epsilon_k.$$

Thus, an $\omega \in A_n$ contains the sequence Y at indices $nk+1, \dots, (n+1)k$. The A_n are independent, because the outcomes ω_j are prescribed at distinct trials. The probability $P(A_n)$ is a positive number, and is independent of n (it is p to the number of 1's in Y times q to the number of 0's in Y). Therefore, $\sum P(A_n) = \infty$, and $P(A_n \text{ i.o.}) = 1$:

> Almost surely, the pattern Y will be repeated infinitely often starting at trials with the special index $nk+1$. ∎

• **5.4.20 Exercise.** Let P be the ordinary measure of length on $[0,1]$ (Lebesgue measure). It is also a probability measure (for the experiment of picking a point at random from the unit interval). Show that $A = [0, 1/2]$ and $B = [0, 1/4] \cup [1/2, 3/4]$ are independent. ∎

5.4.21 Exercise. Refer to Exercise 3.1.7. Define a map F from the Bernoulli sample space X to $[0,1]$ by

$$F : \omega \mapsto \sum_{n=1}^{\infty} \frac{\omega_n}{2^n}. \tag{5.4.22}$$

Suppose that $p = q = 1/2$. Find the image under F of the basic sets $\{\omega_i = 0\}$, $\{\omega_i = 1\}$, and compare the probability measure of the basic sets with the lengths of their images in $[0,1]$. Do the same for the sets in (5.4.9).

This exercise illustrates another way to prove that the probability measure defined on cylinder sets by (5.4.8) and (5.4.9) extends to the Borel sets of X: if one already knows that Lebesgue measure exists on $[0,1]$, one can use it to induce a measure on X. ∎

5.4.23 Exercise. Let B be the set of $x \in [0,1]$ that do not have a unique dyadic expansion (5.4.22) (see also Exercise 3.1.7). Show that B has measure zero. Show that the set of the corresponding ω's has probability zero: $P(F^{-1}(B)) = 0$. ∎

We now turn to the definition of the probability measure on the space of Brownian motion sample paths. There are two sides to the problem, neither of which can be described in as much detail as one would like. First, one must develop a model of the diffusion process. The arguments may be based on physics or on probabilistic analogies; they will result in a plausible formula for the probabilities of the window sets. Second, one must show that these probabilities extend to the Borel σ-algebra generated by the window sets. From the mathematical point of view, that is the difficult part; the existence of the measure was established by Norbert Wiener, and the Brownian motion measure is usually called *Wiener measure* in his honor.

There are by now many different constructions of Wiener measure, each illuminating the character of Brownian motion in a different way.

Method 1 *Limit of random walk.* The Bernoulli trials of Example 5.4.6 are interpreted as a random walk. Toss a fair coin ($p = q = 1/2$) infinitely often, and label the outcomes 1 for *Heads* and -1, rather than 0, for *Tails*. Place a particle at the origin of the t-axis, and move a unit step to the right (left) when the coin toss results in $+1$ (-1). Let Z_i be the result on the i^{th} toss, $0 \leq i < \infty$. The displacement from the origin at time n,

$$d(n) = Z_1 + \ldots + Z_n,$$

can be drawn as a step-function version of Brownian motion, as in Figure 75.

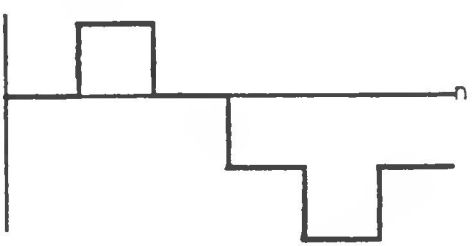


Figure 75. *Random walk*

One now takes a continuum limit of the random walk, by making shorter steps faster. Formally,

$$\lim_{N\to\infty} \frac{Z_1 + \ldots + Z_{[Nt]}}{\sqrt{N}} \stackrel{\text{def}}{=} \omega(t) \tag{5.4.24}$$

is the Brownian motion sample path; here $[x]$ denotes the integer part of x. By fixing a large finite N in (5.4.24) and using a random number generator to produce the ± 1's, one can produce computer simulations of Brownian motion.

Method 2 *Random Fourier series.* Wiener's original proof of the existence of a probability measure on the Brownian motion sample space was constructive, in the spirit of Exercise 5.4.21. There, roughly put, the typical ω was shown to be a number in $[0, 1]$ with random dyadic coefficients. Analogously, the typical Brownian motion ω can be shown to be a Fourier series with random coefficients: if $Y_0, Y_1, \ldots$ are independent normal random variables, mean 0 and variance 1, then

$$\frac{t}{\sqrt{\pi}} Y_0 + \sqrt{\frac{2}{\pi}} \sum_{n=1}^{\infty} \frac{\sin nt}{n} Y_n$$

is normalized Brownian motion for $0 \leq t \leq \pi$. (Don't worry if the words mean nothing to you). In both examples, one can use the (presumably) familiar measure theory on $\mathbf{R}$ to construct the desired probabilities on the abstract sample spaces.

Method 3 *Axiomatic development.* Examples 5.4.19 illustrated consequences of independence in Bernoulli trials. What happens up to trial # i does not influence the outcomes on subsequent trials. One can require this kind of behavior of a diffusing particle, and if the conditions are phrased cleverly, Brownian motion is the only possible result.

Assumption 1 $\omega(0) = 0$. This just requires the diffusing particle to start at a particular point.

Assumption 2 Fix $t > 0$ and $h > 0$. Then $\omega(t + h) - \omega(t)$ is independent of $\{\omega(\tau) \,|\, 0 < \tau \leq t\}$.

We have not really said enough about independence to explain all the concepts here. The physical sense, however, is clear enough. $\omega(t + h) - \omega(t)$ is the displacement of the particle between time t and time $t + h$. $\{\omega(\tau) \,|\, 0 < \tau \leq t\}$ is the complete history of the particle's motion up to the present time, t. The assumption requires the past (how the particle got to its current location) not to influence the subsequent motion *as seen from the particle's current location.*

Assumption 3 The probabilities

$$P(\omega(t + h) - \omega(t)) \in (a, b)$$

are independent of t.

This assumption complements the preceding one. Not only is the motion from time t onward independent of the history, all the probabilities describing how far the particle moves in time h from its location at time t are independent of t. The diffusion *starts over again at each instant of time.* I can look at a diffusing particle at any time $t_0 > 0$, call its location at that time the "origin," and I will see the same typical behavior as you would see were you to follow the particle's motion from time $t = 0$. See Figure 76. (This is a complicated requirement: the location of the particle at time t_0 is random, so the choice of origin changes with the sample path!)

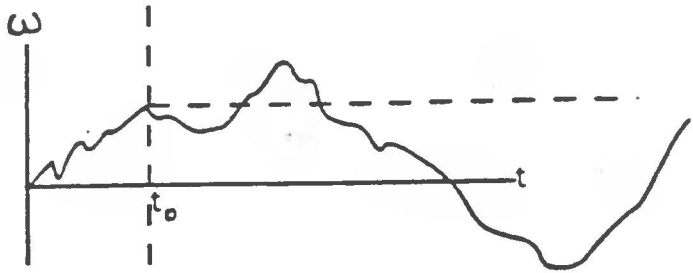


Figure 76. *Brownian motion referred to time* $t_0 > 0$

Assumption 4

$$\lim_{h \downarrow 0} \frac{P(\,|\omega(t+h) - \omega(t)| \geq \delta\,)}{h} = 0, \quad \text{for every } \delta > 0.$$

This is a subtle condition, important for the mathematics but not so easy to appreciate. It imposes a kind of averaged smoothness on the diffusion. Even though a diffusing particle *might* make a huge excursion in a short time interval, it is *unlikely* that it should do so.

Conclusion: One can then prove that there exist two constants μ and σ such that

$$P(\omega(t) \in (a,b)) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \int_a^b \exp -\frac{(x-\mu t)^2}{2\sigma^2 t}\, dx. \tag{5.4.25}$$

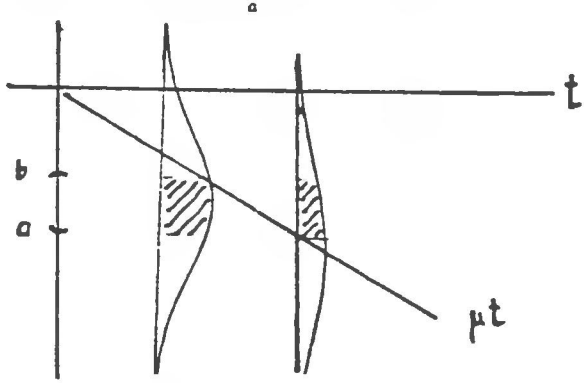


Figure 77. *Brownian motion probabilities*

The graph of the integrand in (5.4.25) is called a *normal curve*. One says (Example 5.2.38.B) that $\omega(t)$ is normal with mean μt and variance $\sigma^2 t$. A sketch of these functions at two values of t, say at $t_1 < t_2$, shows what happens. The

probability of finding the particle in a certain interval at time t_i is given by an area under the normal curve (shaded in Figure 77). If $\mu \neq 0$, then the most likely positions, under the central peak of the normal curve, move with constant velocity μ. This is called *drift*; you would see this effect if you were to impart an initial velocity μ to the particle before it begins to execute the random diffusing motion (assuming, somewhat unrealistically, that it is not subject to friction from the medium in which it moves). The normal curves spread out at a rate $\sigma\sqrt{t}$ (that is a measure of the width of such a curve). Put differently, about 95% of the area under one of our normal curves lies between $\mu t - 2\sigma\sqrt{t}$ and $\mu t + 2\sigma\sqrt{t}$; after the particle has been diffusing for a while, it is just as likely to be near the point μt as far from it. (It would be precisely at μt if there were only drift but no diffusion, i.e. no random behavior).

Remark. If you have a practical mind, you will wonder about practical matters that are evidently not incorporated in this model of diffusion. What about friction effects, as the diffusing particle moves through the fluid? Or, what about gravity and surface tension? It is of course unrealistic to pretend that the motion is unidirectional. It is also unrealistic to suppose (as in Assumptions 2 and 3) that the paricle starts its diffusion anew at each instant of time. When I look at the particle at time t, it has a certain velocity determined by its motion to that point, and its behavior from time t to time $t + h$ is surely affected by that velocity.

All this is true; whether or not the assumptions are modified to take account of such secondary effects depends on the use to which the model will be put. There is a modification of Brownian motion, the *Ornstein-Uhlenbeck* process, in which the particle remembers its velocity. One can include friction, or have the particle attached to a harmonic spring, and one can let it diffuse in two dimensions, or n dimensions, or on the surface of a sphere, or on a Riemannian manifold, or whatever. The *theory of diffusion processes* is a vast field, and we are only looking at the simplest example. ∎

Wiener measure is now described in more detail. The approach is not one of the systematic methods listed above; rather, we start with (5.4.25), where the basic set $U(t_0, a, b)$ is assigned a probability derived from a normal curve, we continue by defining suitable probabilities for window sets, and we do not worry about consistency or other technical matters. For simplicity, set $\mu = 0$ and $\sigma = 1$.

5.4.26 Notation. *Set*

$$p(\xi|\zeta;t) = \frac{1}{\sqrt{2\pi t}} \exp -\frac{(\zeta - \xi)^2}{2t}. \blacksquare$$

It is sometimes convenient, for notation's sake, to think about diffusions that do not necessarily start at the point 0. If one takes this point of view, one interprets $p(\xi|\zeta;t)\,d\zeta$ as the (infinitesimal) probability of finding the particle at position ζ at time t, given that it was at position ξ at time 0.

5.4.27 Definition. *The basic set $U(t_0, a, b)$ in (5.4.1) is assigned the probability*

$$P(U(t_0, a, b)) = \frac{1}{\sqrt{2\pi t_0}} \int_a^b \exp\left(-\frac{\zeta^2}{2t_0}\right) d\zeta = \int_a^b p(0|\zeta; t_0)\, d\zeta. \blacksquare$$

Be sure to understand the intuitive meaning of such a formula. There is an *infinitesimal* probability $p(0|\zeta; t_0)\, d\zeta$ that the path will be at ζ at time t_0. To find the (finite) probability that it will lie between a and b, just add up (integrate) the infinitesimal probabilities over $\zeta \in (a, b)$.

Assigning the probabilities (5.4.25) to the basic sets $U(t_0, a, b)$ is the counterpart of setting probabilities of success and failure at each Bernoulli trial equal to p and q. As in the Bernoulli case, we must next assign probabilities to *intersections* of basic sets, and we must do so in a consistent way. It will suffice to consider an intersection of just two sets, $U(t_1, a_1, b_1) \cap U(t_2, a_2, b_2)$. Figure 78 shows a path ω that starts at 0 when $t = 0$, hits at $\zeta_1 \in (a_1, b_1)$ at time t_1 ("event" A), and reaches $\zeta_2 \in (a_2, b_2)$ at time t_2. (I put quotes around "event" because I am using heuristic talk about infinitesimal probabilities.)

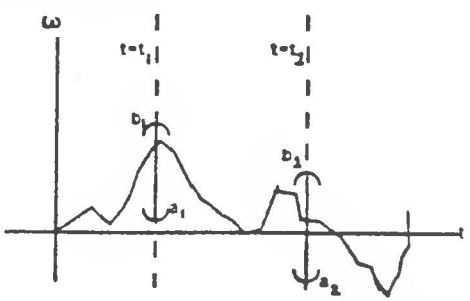


Figure 78. *Probability of passing through two windows*

The probability that ω hits the first window at ζ_1 is, according to the notation in (5.4.26), $p(0|\zeta_1; t_1)\, d\zeta_1$ (it is infinitesimal). At this stage, the diffusion starts anew, as described by Assumptions 2 and 3 above. It starts at ζ_1, not at zero, and we need to go to ζ_2 in the time $t_2 - t_1$. The (infinitesimal) probability for this "event" — call it B — is $p(\zeta_1|\zeta_2; t_2 - t_1)$. The probability that *both* A and B happen is, by independence,

$$p(0|\zeta_1; t_1)\, d\zeta_1 \times p(\zeta_1|\zeta_2; t_2 - t_1) d\zeta_2. \tag{5.4.28}$$

After these heuristic considerations, we now *define* the probability of

$$U(t_1, a_1, b_1) \cap U(t_2, a_2, b_2)$$

to be the result of summing the infinitesimal probabilities (5.4.28) over all $\zeta_1 \in (a_1, b_1)$ and $\zeta_2 \in (a_2, b_2)$:

$$P(U(t_1, a_1, b_1) \cap U(t_2, a_2, b_2)) \tag{5.4.29}$$
$$\stackrel{\text{def}}{=} \int_{a_2}^{b_2} \int_{a_1}^{b_1} p(0|\zeta_1; t_1)\, p(\zeta_1|\zeta_2; t_2 - t_1)\, d\zeta_1 d\zeta_2.$$

The next proposal is now natural.

5.4.30 Definition. *The probability of the intersection of a finite number of basic sets is defined to be*

$$P\left(\bigcap_{j=1}^{n} U(t_j, a_j, b_j)\right)$$
$$\stackrel{\text{def}}{=} \int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} p(0|\zeta_1; t_1) \cdots p(\zeta_{n-1}|\zeta_n; t_n - t_{n-1})\, d\zeta_1 \cdots d\zeta_n. \blacksquare$$

This last formula illustrates one virtue of the notation $p(\xi|\zeta; t)$: try to write this integrand explicitly! More than that, however, the notation reveals the structure of the integral. First, the path goes from 0 to $\zeta_1 \in [a_1, b_1]$ in time t_1, then it goes to $\zeta_2 \in [a_2, b_2]$ in time $t_2 - t_1$, etc., and in the end one sums over all the ways for ω to pass from window to window.

• **5.4.31 Exercise.** Verify that

$$p(\xi|\zeta; t) = \int_{-\infty}^{\infty} p(\xi|\eta; \tau) p(\eta|\zeta; t - \tau)\, d\eta \quad \text{for } 0 < \tau < t. \tag{5.4.32}$$

Give an intuitive interpretation to this formula. (To get from ξ to ζ in time t, the particle can go from ξ to η in time τ ...). This is the *Chapman - Kolmogorov - Smoluchowski* equation.

Remark: The CKS equation (5.4.32) is analogous to the relation

$$P(\omega_2 = 1) = P(\omega_1 = 0, \omega_2 = 1) + P(\omega_1 = 1, \omega_2 = 1)$$

in Bernoulli trials. It is used to prove *consistency* of (5.4.30) on cylinder sets. The intuitive interpretation of the CKS equation does not draw on all the specific properties of Brownian motion. Indeed, the functions

$$p(\xi|\zeta; t) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ix(\zeta - \xi)} e^{-t|x|^\alpha}\, dx$$

for $0 < \alpha \leq 2$ also satisfy (5.4.32). The case $\alpha = 2$ recovers (5.4.26); the others define new stochastic processes, different from Brownian motion, called *stable processes with exponent* α. See L. Breiman, *Probability*. ∎

5.4.33 Remark. The extension of (5.4.30) to a measure defined on all the Borel sets of Ω proceeds along the lines sketched in Example (5.4.6). Cylinder sets are sets determined by restrictions on ω at finitely many times t_j. One shows that (5.4.30) is consistent on the semialgebra of cylinder sets (this uses (5.4.32)) and then appeals to the extension theorem mentioned in (5.4.6). The procedure is codified in the *Kolmogorov extension theorem* for stochastic processes. ∎

To conclude, I pose just two of the questions one asks about Brownian motion once one has the probability measure under control, and I sketch a partial answer to one of them.

Question 1. Does the diffusing particle eventually leave a neighborhood of its starting point, never to return? **Answer:** It may, of course — just take the function $\omega(t) = t$. However, for a set $A \subset \Omega$ with $P(A) = 1$, the following is true: for any $\omega \in A$, and any $N > 0$, there is a $t > N$ such that $\omega(t) = 0$. In words: most diffusion paths return to the starting point at arbitrarily large times.

Question 2. Computer simulations show the typical ω to be very "jiggly." Describe the jiggles. **Answer:** This is a vast subject. For example, one knows that, except for a set of ω's of probability zero, the set $\{t \mid \omega(t) = 0\}$ is a *closed, perfect set of Lebesgue measure* 0. The term means, roughly speaking, that the set of zeros of ω is a Cantor-like set; moreover, every zero of ω is itself an accumulation point of a Cantor-like set of nearby zeros. The same is true if $\omega = 0$ is changed to $\omega = c$: near each value of the typical ω, the graph jiggles back and forth on almost unimaginably fine scales. Such behavior augurs ill for derivatives, and indeed one of the most famous properties of Brownian motion is:

Theorem. *The set*

$$\{\omega \mid \omega \text{ has no derivative at any } t \in [0, \infty)\}$$

has probability 1.[44] ∎

The proof of the theorem will not be given; instead, I present a fairly rigorous argument for another result of this type.

5.4.34 Proposition. *With probability* 1, ω *is not monotone on any interval.* ∎

Proof. Recall that a function f is *monotone* on an interval (a, b) if $f(s) \leq f(t)$ or $f(s) \geq f(t)$ for all $a < s < t < b$.

We prove first that with probability 0, ω is nondecreasing on $[0, 1]$. This is the interesting step. Let

$$A_n = \{\omega \mid \omega(\tfrac{i+1}{n}) - \omega(\tfrac{i}{n}) \geq 0 \text{ for } i = 0, \ldots, n-1\}.$$

[44] This means, in particular, that all the continuous functions you have studied until now will, most likely, form a set of measure zero.

Two useful lemmas are left as exercises.

• **5.4.35 Exercise.** Let $t_1 < t_2$. Using Definition 5.4.30, show that

$$P(\omega(t_2) - \omega(t_1)) \in (a, b)) = P(\omega(t_2 - t_1)) \in (a, b)).$$

(Hint: Use

$$\int_a^b \int_{-\infty}^{\infty} p(\zeta_1 | \zeta_1 + \eta; t_2 - t_1)\, p(0 | \zeta_1; t_1)\, d\zeta_1 d\eta.$$

First, interpret this integral: the path goes from 0 to ζ_1 in time t_1, and then ....) ∎

• **5.4.36 Exercise.** Let $t_1 < t_2 < t_3$. Show (by computing probabilities) that the events

$$\{\omega \,|\, \omega(t_2) - \omega(t_1) \in (a, b)\} \text{ and } \{\omega \,|\, \omega(t_3) - \omega(t_2) \in (c, d)\}$$

are independent. Generalize. ∎

Two things follow from these exercises. First,

$$P(\omega(\frac{i+1}{n}) - \omega(\frac{i}{n}) \geq 0) = \int_0^{\infty} p(0|\zeta; \frac{1}{n})\, d\zeta = \frac{1}{2}.$$

Second, the differences $\omega((i+1)/n) - \omega(i/n))$ in the definition of A_n are independent, so that

$$P(A_n) = \prod_{i=0}^{n-1} P\left(\omega(\frac{i+1}{n}) - \omega(\frac{i}{n}) \geq 0\right) = \left(\frac{1}{2}\right)^n.$$

Hence, $P(A_n) \to 0$, and by Proposition 5.2.13, $P(\bigcap_n A_n) = 0$.

It remains to identify $B \stackrel{\text{def}}{=} \bigcap_n A_n$ as the set of ω which are nondecreasing on $[0, 1]$.

• **5.4.37 Exercise.** Prove this. (Hint: if $\omega \in B$, then $\omega(i/n) \leq \omega(j/n)$ for all n and $0 \leq i < j \leq n-1$. If $s < t$ and $\omega(s) > \omega(t)$, find i/n near s and j/n near t to reach a contradiction. Conversely, if $\omega \notin B$, show that there must be $s < t$ with $\omega(s) > \omega(t)$.) ∎

Now we can use this special result to prove the proposition.

The interval $[0, 1]$ can be replaced by any interval $[a, b]$. Just look at

$$\omega(a + (b-a)\frac{i+1}{n}) - \omega(a + (b-a)\frac{i}{n}).$$

The notation is more complicated, but the argument remains the same, and we see:

> The set $B_{a,b}$ of ω that are nondecreasing on the interval $[a, b]$ has probability 0.

At this stage, one would like to say: throw out all the ω's that are nondecreasing on some interval. Those have measure (probability) zero. What's left are the ω that are nondecreasing on no interval, and they have probability 1. There is a flaw in this argument: there are uncountable many intervals $[a, b]$, and uncountably many sets $B_{a,b}$ of probability 0. Their union may have positive probability.

The remedy is (almost) always the same: use the fact that the rationals are a) countable and b) dense in the reals. There are only countably many intervals $[a, b]$ with *rational* endpoints. The countable union of the probability-zero sets $B_{a,b}$ has probability zero by Exercise 5.2.17. The *complement* of this set has probability 1; it consists of all the ω that are nondecreasing on no rational interval.

But now, if a continuous function is nondecreasing on an interval $[s, t]$ with irrational endpoints, then it is also nondecreasing on any rational interval contained in $[s, t]$. Hence, the set of ω that fail to be nondecreasing on any interval has probability 1.

In exactly the same way, one proves that the set of ω that fail to be *nonincreasing* on any interval has probability 1. The intersection consists of the ω that are not monotone on any interval, and it still has probability 1. ∎

It is impossible to draw the (typical) sample path ω, and probably impossible to imagine it. There is no interval, no matter how small, on which the graph of ω heads either down or up. Still, the graph can go quite far above or below the t-axis. It does this by frenetic jiggles that can only be understood as a limit of oscillations on successively smaller scales.

• **5.4.38** Exercise. Prove: if $P(A) = 1$ and $P(B) = 1$, then $P(A \cap B) = 1$. ∎

• **5.4.39** Exercise. Show that the set of ω that are eventually never zero is a Borel set. ∎

CHAPTER II

Convergence and Compactness

The last chapter dealt with measures of size and distance. These are usually only the first ingredients in the application of advanced techniques to any problem. If, for example, we need to solve an equation $f(x) = b$, with x and b in some function space, the mere definition of a norm in that space helps little. One must put the norm to some use. Solving equations is a very typical problem in applied mathematics, so let us relate the material in this chapter to that situation. Since the abstract equation $f(x) = b$ might be a standard linear system in $\mathbf{R}^n$, or a partial differential equation like $\Delta u = \phi$, or an integral equation like $\int_a^b K(r,s)u(s)\,ds = b(r)$, the general problem "$f(x) = b$" incorporates a great deal of mathematics.

Having fixed the right-hand side b, we define $g(x) = f(x) - b$, and seek to solve $g(x) = 0$. For instance, if g is the derivative of a function G, then solving $g(x) = 0$, or $G'(x) = 0$, is one step towards minimizing G, which is a ubiquitous and important problem in analysis.

To solve $g(x) = 0$, one often generates a sequence of approximate solutions, and one hopes that it converges to the true solution. (Example: x_n satisfies $g(x_n) = b_n$, where $b_n \to 0$. Does $\{x_n\}$ converge to a solution x of $g(x) = 0$?) This requires an understanding of *convergence*: what does it mean, in a complicated space, that "$x_n \longrightarrow x$"? How can one tell, without knowing the limit *a priori*, that a sequence converges? If the x_n satisfy the approximate equations $g(x_n) = b_n$, and if $x_n \longrightarrow x$ and $b_n \longrightarrow b$, will the limit equation $g(x) = b$ be satisfied? It will, if the function g, mapping one huge space to another, is continuous. So the concept of *continuity* also enters the problem.

When you propose to construct a solution of $g(x) = 0$ as limit of a sequence $\{x_n\}$ of approximations, you must be able to deduce properties of that solution x. For instance, if x is a function of a variable t, is it continuous if the $x_n(t)$ are continuous? Is it differentiable, if the $\dot{x}_n(t)$ exist? Can it be integrated? Such questions pose the problem of *interchange of limit and all the usual calculus operations*: is $\lim \int x_n(t)\,dt = \int x(t)\,dt$? is $\lim \dot{x}_n(t) = \dot{x}(t)$?

Mathematicians of the 19th century discovered that the then current theory of functions, derivatives, and integrals (which we still learn and teach in the standard undergraduate curriculum) was not flexible enough to handle the common limiting processes. Not only may $\lim \int x_n(t)\,dt$ fail to equal $\int x(t)\,dt$, the limit $x(\cdot)$ might not even have an integral in the calculus sense. The theory of the *Lebesgue integral* takes care of such problems, at the expense of admitting complicated and highly discontinuous functions. It is an indispensable tool in modern analysis.

Sometimes, it is more convenient to produce a whole slew of approximate solutions of which only a subsequence may converge to a true solution. The crucial ingredient is *compactness*: under appropriate conditions, the points of an infinite set $\{x_n\}$ must "pile up" somewhere, i.e., there must be an accumulation point. There will, in general, be several accumulation points, and this idea tends to prove existence, but not uniqueness, of solutions to $g(x) = 0$. To take a familiar example: if $g = G'$, we try to find minima (or maxima) of G by solving $g(x) = 0$. Obviously, there will not, in general, be a unique extremum. Compactness has many other important consequences, some of which will be described below.

Now for a summary of the material in this chapter. Section 1 explains the basic concepts: convergence and compactness. Everything else is based on this discussion. Section 2 uses "convergence" to motivate the introduction of the Lebesgue integral. This is not the most efficient or popular approach, but it has a certain intuitive appeal that outweighs its technical awkwardness, at least when the intent is to summarize the theory rather than to develop all its details. Compactness in function spaces is used in Section 3 to prove the fundamental existence theorem for ordinary differential equations and to establish existence of geodesics in metric spaces. In Section 4, elementary facts about Banach and Hilbert spaces are discussed. Those concepts distill the common features of the function spaces (L^p, l^p, C) introduced in earlier portions of the text.

Section 1 Convergence and Compactness

Convergence is a very basic concept. We met it long ago, in Definition I.1.2.5. Without "convergence", it is difficult to think about "continuity", and without those two notions it would be rather pointless to introduce metric spaces or general topological spaces. But it is now time to change the point of view. Mostly, so far, convergence and continuity have been tools that helped us to appreciate abstract spaces and general constructions. Mathematical practice usually reverses the order. One has a specific sequence, and one wants to understand its limit — if it exists. One has a continuous function, and one wants to infer a property of interest (the function has a maximum, the function has a zero, ...). In such typical problems, the abstract spaces and general constructions will be the tools, and the emphasis is placed on the finer details of sequences and functions.

We begin with a fundamental question: can one tell, by looking at a sequence, whether or not it converges? Of course, a nontrivial sequence consists of infinitely many objects whose construction may be quite complicated, so "looking at it" may not be easy. Still, the *Cauchy criterion*, discussed in Subsection 1, reveals, in principle, when a sequence "wants to converge". It says, roughly speaking, that a sequence ought to converge if its elements x_n get closer and closer to each other as $n \to \infty$ (for then they should also be getting close to a limit x). Such sequences are called *Cauchy sequences*.

Even when a sequence wants to converge, however, there may not be a limit if the problem is set in too narrow a framework. A sequence of rational numbers may converge to an irrational number, a sequence of continuous functions may converge to a discontinuous function, and a sequence of integrable functions may converge to a non-integrable function. This leads to the notion of *completeness*. A metric space is "complete" if every Cauchy sequence does in fact converge, to an element of that same space. If a metric space is not "complete", one enlarges it, adding new "ideal" points until every Cauchy sequence has been assigned a unique limit. This construction, outlined in Subsection 2 and discussed at greater length in Section 2, is characteristic of twentieth century analysis. To complete the metric space $C([0,1])$ of continuous functions on $[0,1]$ in the L^1 metric, one must throw in functions that are far from continuous; they are merely *measurable*, and the calculus integral

$$\int_0^1 |f(x) - g(x)|\, dx$$

must be reinterpreted: this is the Lebegue theory of integration. There are many other "completions" of classical calculus-type objects. The theory of generalized functions of Section I.4 can be thought of in this light: in order to make sequences like

$$\frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}$$

converge, one introduces new objects, such as the Dirac delta distribution, as limits.

It is not always reasonable to expect convergence of the sequences encountered in the analysis of complicated problems. The following question will be studied later, in Section 3: given two points p and q on a surface S, find a curve of shortest length (a *geodesic*) connecting p and q. To each curve Γ from p to q associate its length, $\ell(\Gamma)$. The goal is to minimize the function ℓ over the set M of all curves joining the given points. (M turns out to be a complete metric space). We will try to establish existence of a geodesic by finding curves Γ_n whose lengths approach the minimum length as $n \to \infty$. But a simple example reveals a serious difficulty. If p and q are, respectively, the north and south poles on a sphere, there will be infinitely many "shortest curves" from p to q. Can one be sure that the approximate geodesics Γ_n all approximate the same semicircle on the sphere? One cannot be sure, and all we will be able to conclude is that some *subsequence* Γ_{n_j} of the Γ_n approaches one of the many possible geodesics.

The topological mechanism underlying this example (and many others) is *compactness*, introduced in Subsection 3. A topological space is compact (roughly speaking, again) if every sequence has a convergent subsequence. Compactness is much more restrictive a condition than completeness. It confers the moral equivalent of "finiteness" on infinite sets, and is the source not only of existence theorems such as the one just sketched, but also of important *uniformity* results. This is the subject of Subsection 4. "Uniformity" is a catch-all term that signifies

properties like: the δ in "$|f(x) - f(x_0)| < \epsilon$ whenever $|x - x_0| < \delta$" is the same for all x_0, or the N_ϵ in "$|f_n(x) - f(x)| < \epsilon$ whenever $n > N_\epsilon$" is the same for all x. Such conditions (these two are called *uniform continuity* and *uniform convergence*) afford crucial control over limiting processes and approximation techniques.

The material in this first section is the theoretical underpinning of Chapter 2. Subsequent sections develop completeness and compactness in various directions (sometimes beyond recognition), but the basic ingredients remain unchanged.

Subsection 1.1 Cauchy Sequences.

The fundamental concept of *convergence* makes sense in any topological space (recall Definition 3.2.10):

1.1.1 Definition. *Let (X, T) be a topological space. Let $\{x_n\}$ be a sequence in X. The sequence is said to converge to $x \in X$ if for every neighborhood U of x, there exists an integer N_U, such that $n \geq N_U \Rightarrow x_n \in U$.* ∎

Example I.3.2.15 shows, by example, that in the more bizarre topological spaces, this definition may be completely useless: there may be no non-trivial convergent sequences at all. That is one good reason to specialize to metric spaces, but there is a more important one: the Cauchy criterion, to be introduced presently, is quantitative in nature — one needs a numerical measure of distance.[1] Therefore, we will speak only about metric spaces; we begin by recalling Definition I.1.2.5 of convergence in a metric space:

1.1.2 Definition. *Let (M, d) be a metric space. Let $x_n \in M$ for $n \geq 1$, and let $x \in M$. One says that the sequence $\{x_n\}$ converges to x if, given any $\epsilon > 0$, there is an integer N_ϵ (depending on ϵ), so that*

$$d(x_n, x) < \epsilon \quad \text{for} \quad n > N_\epsilon. \ \blacksquare$$

Definition 1.1.2 of convergence is, so to speak, "external": given the sequence, you can presumably tell whether or not it converges to x, but that is not efficient. To test whether a given sequence converges at all, you would have to check Definition 1.1.2 for all possible x. The Cauchy criterion comes to the rescue: it is "internal".

1.1.3 Definition. *A sequence $\{x_n\}$ in a metric space (M, d) is called a Cauchy sequence if, for every $\epsilon > 0$, there exists an N_ϵ such that (the "Cauchy criterion")*

$$n, m > N_\epsilon \Longrightarrow d(x_n, x_m) < \epsilon. \ \blacksquare$$

The following result explains the meaning of this definition: if the x_n get close to x, they must eventually get close to each other.

---

[1] As always, there is a generalization. In the theory of *uniform topological spaces*, one can find a non-quantitative version of metric spaces (EDMII, Article 436). Fortunately, it will not be needed.

1.1.4 Proposition. *If $x_n \to x$, then $\{x_n\}$ is a Cauchy sequence.* ∎

Proof. Prescribe $\epsilon > 0$. By the definition of convergence, there is an $N_{\epsilon/2}$ such that $d(x_n, x) < \epsilon/2$ and $d(x, x_m) < \epsilon/2$ whenever $n, m > N_{\epsilon/2}$. Now the Cauchy property follows immediately from the triangle inequality,

$$d(x_n, x_m) \leq d(x_n, x) + d(x, x_m) < \epsilon. \blacksquare$$

1.1.5 Examples

A. The definition of Cauchy sequence requires that $d(x_n, x_m) < \epsilon$ for *all* $n, m > N_\epsilon$. It is *not* enough to have $d(x_n, x_{n+1}) < \epsilon$ for $n > N_\epsilon$. Here is an example. If $x_n = \sqrt{n}$, then $x_{n+1} - x_n = 1/(\sqrt{n+1} + \sqrt{n}) \longrightarrow 0$. The sequence $\{x_n\}$ does not converge to a finite limit, however. So, even if *adjacent* terms get close, a sequence need not converge.

B. Let $\{a_j\}$ be a sequence of complex numbers, and set $s_n = \sum_{j=0}^{n} a_j$. We say that the series $\sum_{j=0}^{\infty} a_j$ converges if the sequence $\{s_n\}$ of partial sums has a limit. By the Cauchy criterion, if this is the case then for every $\epsilon > 0$, there will be an N_ϵ such that $|s_n - s_m| < \epsilon$ whenever $n, m > N_\epsilon$. Equivalently,

$$\Big| \sum_{j=n+1}^{m} a_j \Big| < \epsilon \quad \text{for} \quad n, m > N_\epsilon.$$

C. For $n \geq 1$, set $f_n(t) = t^n$, $t \in [0, 1]$. When $m \geq n$, then $t^m \leq t^n$, and so

$$\|f_n - f_m\|_1 = \int_0^1 |t^n - t^m|\, dt = \frac{1}{n+1} - \frac{1}{m+1}.$$

Given $\epsilon > 0$, choose $N_\epsilon > 2/\epsilon$. Then for $n, m > N_\epsilon$,

$$\|f_n - f_m\|_1 < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

and $\{f_n\}$ is a Cauchy sequence in $L^1([0, 1])$. As it must be: the estimate

$$\|f_n - 0\|_1 = \int_0^1 |t^n|\, dt = \frac{1}{n+1}$$

shows that f_n tends to the zero function in the L^1 metric, and so Proposition 1.1.4 applies.

The sequence, however, is *not* Cauchy in the L^∞ metric. To prove this, we should find *some* $\epsilon > 0$, so that for every integer N there are indices $n, m > N$ for which

$$\|f_n - f_m\|_\infty > \epsilon.$$

Let $\epsilon = .1$, and pick an $N > 0$. Fix $m = N + 1$. Because

$$\lim_{n\to\infty} (1 - \frac{1}{n})^m = 1.$$

we will have

$$(1.1.6) \qquad (1 - \frac{1}{n})^m > .9$$

for all sufficiently large n (> some $N_1 > N$, say). On the other hand,

$$(1.1.7) \qquad \lim_{n\to\infty} (1 - \frac{1}{n})^n = e^{-1},$$

and therefore

$$|(1 - \frac{1}{n})^n - e^{-1}| < .1$$

for $n >$ some $N_2 > N_1$. Fix such an n. Estimates (1.1.6) and (1.1.7) show that at the point $t_0 = 1 - 1/n$, $|t_0^n - t_0^m| > .9 - (e^{-1} + .1) > .1 = \epsilon$. Hence, for n and m as chosen above (both $> N$),

$$\|f_n - f_m\|_\infty = \sup_{0\le t\le 1} |t^n - t^m| > \epsilon.$$

and the Cauchy criterion does not hold. It follows, in particular, that *this sequence $\{f_n\}$ does not converge in $L^\infty([0,1])$.* (If it did, it would be a Cauchy sequence.) ∎

• **1.1.8 Exercise.** Prove that the sequence $s_n = \sum_1^n 1/j$ is not a Cauchy sequence. ∎

• **1.1.9 Exercise.** Let (M, d) be a metric space, let $\{x_n\}$ be a sequence in M, and suppose that there are numbers k, $0 < k < 1$, and $C > 0$ such that $d(x_{n+1}, x_n) \le Ck^n$ for all $n \ge 1$. Show that the sequence is Cauchy. ∎

1.1.10 Exercise. Define a sequence of real numbers by $x_1 = 2$,

$$x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}.$$

(This is the Newton-Raphson method for finding $\sqrt{2}$.) Prove that $\{x_n\}$ is a Cauchy sequence, *without* using the fact that $\lim x_n = \sqrt{2}$. (Hint: use the preceding exercise.) ∎

The main point of the concept "Cauchy sequence" is that *a Cauchy sequence should converge.* This intuition comes from experience with the real numbers.

1.1.11 Fact. *Let $\{x_n\}$ be a Cauchy sequence of real numbers. Then there is an $x \in \mathbb{R}$ such that $\lim_{n\to\infty} x_n = x$.* ∎

This is a profound statement. When we try to identify the basic properties of the real numbers, the *axioms* from which all other results follow as corollaries, we must postulate, in one form or another, that the number line is "unbroken": there are no holes. One way of saying this is to require that sequences that want to converge (Cauchy sequences) don't try to converge, in vain, to a hole in the number line. Another way is to ask that every bounded set of real numbers have a supremum, or least upper bound (defined in I.2.3.55). If this were not the case, there would be a hole where the boundary number should be. A third possibility is to require the topological property called *connectedness* (Definition I.3.2.29):

1.1.12 Definition. *A topological space (X, T) is said to be **connected**, if it is impossible to write $X = U_1 \cup U_2$ with disjoint, nonempty open sets U_1, U_2.* ∎

If the real numbers (in the usual topology) are required to be connected, one cannot have $\mathbb{R} = (-\infty, a) \cup (a, \infty)$, for example, and there is unlikely to be a hole where the number a should be.

These three properties of $\mathbb{R}$ —convergence of Cauchy sequences, existence of least upper bounds, and connectedness— are equivalent. If one is postulated, the others follow as corollaries. I prefer to take all of them as given, and move on to the consequences (for more discussion of the foundations of the real number system, see R. C. Buck, *Advanced Calculus*, particularly Appendix 2, and also the beginning of Section 2 below).

The next definition introduces the class of metric spaces in which Cauchy sequences always converge. As will become clear in the examples, a proof that a particular metric space has this property is usually based on axiom (1.1.11) about the real numbers.

1.1.13 Definition. *A metric space (M, d) is called **complete** if every Cauchy sequence converges*[2]*. A metric space that is not complete is called **incomplete**.* ∎

There now follow the two (expected) kinds of examples: of incomplete metric spaces, and of complete metric spaces. In Subsection 2, I will explain how to turn incomplete metric spaces into complete ones.

1.1.14 Examples

A. Let Q be the set of rational numbers, with the usual metric. We consider this set to be a metric space in its own right, without reference to the rest of the

[2] The limit must be in M – because there are no other points around once one has to work with a specific metric space M – and the limit is necessarily unique.

real numbers. There are sequences of rationals that converge to an irrational; such sequences are Cauchy sequences in Q that have no limit in Q. (E.g.: $x_1 = 1, x_2 = 3/2, \ldots, x_{n+1} = x_n/2 + 1/x_n, \ldots$, has limit $\sqrt{2}$).

B. The sequence $f_n(t) = t^n$ in Example 1.1.5C is Cauchy in the L^1 metric on the space $C([0,1])$ of continuous functions on $[0,1]$. We saw that $\{f_n\}$ approaches the continuous function $g(t) \equiv 0$ in the L^1 metric. Note that the *pointwise* limit of this sequence of functions is not continuous:

$$\lim_{n\to\infty} t^n = f(t) \stackrel{\text{def}}{=} \begin{cases} 0, & \text{if } 0 \le t < 1, \\ 1, & \text{if } t = 1. \end{cases} \tag{1.1.15}$$

The moral is that one must use the convergence concept that is appropriate to the metric.

C. Here is an example that shows the metric space from part B to be incomplete. Let $p = 1$, and for $n \ge 2$ let

$$f_n(t) = \begin{cases} -1, & \text{for } 0 \le t \le \frac{1}{2} - \frac{1}{n}; \\ nt - \frac{n}{2}, & \text{for } \frac{1}{2} - \frac{1}{n} \le t \le \frac{1}{2} + \frac{1}{n}; \\ +1, & \text{for } \frac{1}{2} + \frac{1}{n} \le t \le 1. \end{cases}$$

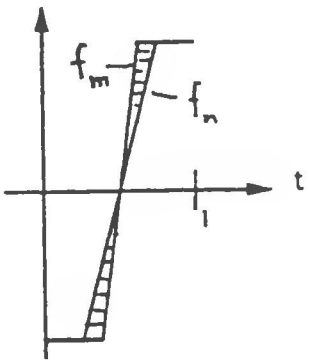


Figure 1. *A Cauchy sequence in L^1*

This is a Cauchy sequence in the L^1 metric. The distance $\|f_n - f_m\|_1$ is just the total area between the graphs of f_n and f_m (see Figure 1), and equals $2 \times |\frac{1}{n} - \frac{1}{m}|$, which becomes arbitrarily small for sufficiently large n, m. We now ask: can this Cauchy sequence have a limit f *in the space* $C[0,1]$?

Clearly, the *pointwise* limit, $\lim_{n\to\infty} f_n(t)$, is the discontinuous step function

$$\psi(t) = \begin{cases} -1, & \text{for } 0 \le t < 1/2, \\ 0, & \text{for } t = 1/2, \\ +1, & \text{for } 1/2 < t \le 1. \end{cases}$$

But, as we saw in part B, the fact that $f_n(t) \longrightarrow \psi(t)$ for each individual t does not exclude the possibility that f_n tends to a completely different, possibly continuous, function *in the L^1-metric*. We should prove that *no continuous function can be the L^1-limit of f_n*.

Suppose, to the contrary, that $f_n \xrightarrow{L^1} f$, with continuous f. By Minkowski's inequality (which holds for piecewise continuous functions, see Section I.1.4),

$$\|f - \psi\|_1 \leq \|f - f_n\|_1 + \|f_n - \psi\|_1.$$

By assumption, the first term on the right tends to zero. The second term tends to zero as well, because it is just the area shaded in Figure 2.

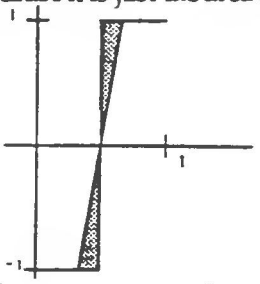


Figure 2. *L^1 convergence to a discontinuous function*

Since the left side, $\|f - \psi\|_1$, of the inequality does not depend on n, but can be made arbitrarily small, we must have $\|f - \psi\|_1 = 0$. This, however, is a contradiction. Because f is continuous, there must be a $t_0 \in [0, 1/2)$ such that $f(t_0) \neq -1$ or a $t_0 \in (1/2, 1]$ such that $f(t_0) \neq +1$. Suppose the latter, without loss of generality. By continuity of f, $|f(t) - 1| >$ some $\eta > 0$ in a neighborhood of t_0. This implies that the integral of $|f(t) - \psi(t)|$ must be positive. ∎

• **1.1.16 Exercise.** Use the same sequence f_n and modify the last argument to prove that $C([0,1])$ is not complete in the L^2–metric. ∎

1.1.17 Exercise. Prove that $C([0,1])$ is not complete in any L^p–metric, for $1 \leq p < \infty$. (It will be proved presently that $C([0,1])$ *is* complete in the L^∞-metric.) ∎

We now turn to examples of *complete* metric spaces. Notice, in each case, how the completeness of $\mathbb{R}$ is essential.

1.1.18 Proposition. *$\mathbb{R}^2$ is complete in the Euclidean metric.* ∎

Proof. Let $\{(x_n, y_n)\}$ be a Cauchy sequence in $\mathbb{R}^2$. I.e., given $\epsilon > 0$, there is an N_ϵ such that $n, m > N_\epsilon \Rightarrow (x_n - x_m)^2 + (y_n - y_m)^2 < \epsilon^2$. Hence, $(x_n - x_m)^2 < \epsilon^2$, or $|x_n - x_m| < \epsilon$. So the sequence $\{x_n\}$ of real numbers is a Cauchy sequence, and

by (1.1.11) has a limit, x. Likewise, $\{y_n\}$ has a limit, y. We need to show now that $(x_n, y_n) \longrightarrow (x, y)$ *in the metric of* $\mathbb{R}^2$. This is fairly obvious, but the proof will carry over to the infinite-dimensional case, so we do the details. In

$$(x_n - x_m)^2 + (y_n - y_m)^2 < \epsilon^2,$$

let $m \longrightarrow \infty$, while keeping n fixed. Because the function $f(\xi, \eta) = (a-\xi)^2+(b-\eta)^2$ is continuous, we can conclude that in the limit,

$$(x_n - x)^2 + (y_n - y)^2 \leq \epsilon^2.$$

Hence, (x_n, y_n) is within ϵ of (x, y) in the $\mathbb{R}^2$ metric, whenever $n > N_\epsilon$. According to Definition 1.1.2, $(x_n, y_n) \rightarrow (x, y)$. ∎

1.1.19 Proposition. *Let* $1 \leq p \leq \infty$. *The space* $l^p(\mathrm{R.N})$ *is complete.* ∎

Proof. We do the case $1 \leq p < \infty$. Let $\{\mathbf{x}^{(n)}\}$ be a Cauchy sequence in l^p. So, $\mathbf{x}^{(n)} = (x_0^{(n)}, x_1^{(n)}, \ldots)$, and

$$\sum_{i=0}^{\infty} |x_i^{(n)} - x_i^{(m)}|^p < \epsilon^p$$

whenever $n, m > N_\epsilon$. Just as in (1.1.18), the sequence $\{x_i^{(n)}\}$ is a Cauchy sequence of real numbers, and by assumption it has a limit, x_i. We expect that $\mathbf{x} = (x_0, x_1, \ldots)$ is the limit of the original sequence $\{\mathbf{x}^{(n)}\}$.

Two things must be checked. First, does the limit $(x_0, x_1, \ldots)$ belong to l^p? I.e., is

$$\sum_{i=1}^{\infty} |x_i|^p < \infty\,?$$

Second, is $\|\mathbf{x}^{(n)} - \mathbf{x}\|_p < \epsilon$ for sufficiently large n? I.e., does convergence take place not just entry by entry, but also in the l^p metric?

One would like to say: in

$$\sum_{i=0}^{\infty} |x_i^{(n)} - x_i^{(m)}|^p < \epsilon^p, \tag{1.1.20}$$

fix n and let $m \longrightarrow \infty$ to get

$$\sum_{i=0}^{\infty} |x_i^{(n)} - x_i|^p \leq \epsilon^p, \tag{1.1.21}$$

which would show l^p-closeness. This step is not quite as trivial as it may seem. Define a function F of infinitely many variables by

$$F(a_0, a_1, \ldots) = \sum_{i=0}^{\infty} |x_i^{(n)} - a_i|^p$$

(with the $x_i^{(n)}$ fixed). Passing from (1.1.20) to (1.1.21) amounts to saying that this F is continuous in the infinitely many a_i. This is true, but is also just unfamiliar enough to warrant a little care.

Therefore, we truncate the series in (1.1.20) so that only continuity of a (more familiar) function of finitely many variables is required (an alternative argument is suggested in Exercise 1.1.38):

$$\sum_{i=0}^{M} |x_i^{(n)} - x_i^{(m)}|^p \leq \sum_{i=0}^{\infty} |x_i^{(n)} - x_i^{(m)}|^p < \epsilon^p,$$

and so if $m \to \infty$ in the finite sum on the left, we get

$$\sum_{i=0}^{M} |x_i^{(n)} - x_i|^p \leq \epsilon^p.$$

Now let $M \to \infty$, to conclude that

$$\|\mathbf{x}^{(n)} - \mathbf{x}\|_p \leq \epsilon. \tag{1.1.22}$$

By Minkowski's inequality (I.1.3.2),

$$\mathbf{x}^{(n)} \in l^p \text{ and } \mathbf{x} - \mathbf{x}^{(n)} \in l^p \Rightarrow \mathbf{x} = \mathbf{x}^{(n)} + (\mathbf{x} - \mathbf{x}^{(n)}) \in l^p.$$

So, the hoped-for limit $\mathbf{x}$ is in l^p, and then (1.1.22) shows that our sequence $\{\mathbf{x}^{(n)}\}$ converges to it in the l^p–norm. ∎

In Example 1.1.14C, it was shown that the space of continuous functions on $[0, 1]$ is not complete in the L^1 metric (for the L^p-metrics, $1 < p < \infty$, see Exercise 1.1.17). As we now show, completeness in the L^∞ metric does hold; for this reason, $\|\cdot - \cdot\|_\infty$ is usually considered to be the "natural" norm on $\mathcal{C}([0, 1])$.

1.1.23 Proposition. *$\mathcal{C}[0, 1]$ is complete in the L^∞ metric.*[3] ∎

Proof. Let $\{f_n\}$ be a Cauchy sequence in the L^∞ metric on $\mathcal{C}([0, 1])$: given $\epsilon > 0$, there is an N_ϵ such that

$$\|f_n - f_m\|_\infty < \epsilon \quad \text{whenever } n, m > N_\epsilon.$$

[3] In Subsection 4, this result will be rephrased as follows: a uniform Cauchy sequence of continuous functions has a continuous limit.

We first show that for each $t \in [0,1]$, the limit

$$\lim_{n\to\infty} f_n(t) \stackrel{\text{def}}{=} f(t)$$

exists, and then prove that this limit function f is continuous.

Let $t \in [0,1]$. Given $\epsilon > 0$, we have

$$|f_n(t) - f_m(t)| \le \sup_{y\in[0,1]} |f_n(y) - f_m(y)| = \|f_n - f_m\| < \epsilon$$

when $n, m > N_\epsilon$, by the Cauchy property in the L^∞ metric. Hence $\{f_n(t)\}$ is a Cauchy sequence of real numbers, which has a limit we will call $f(t)$. (Notice how the completeness of the reals again plays a crucial role.)

Once we know that $\lim f_n(t)$ exists for each t, we can get a uniform estimate over $t \in [0,1]$, as follows. Take any $t \in [0,1]$.

$$|f_m(t) - f_n(t)| \le \sup_{y\in[0,1]} |f_m(y) - f_n(y)|, \tag{1.1.24}$$

and so when $m \to \infty$,

$$|f(t) - f_n(t)| \le \sup_{y\in[0,1]} |f(y) - f_n(y)| = \|f - f_n\|_\infty. \tag{1.1.25}$$

For $n, m >$ some N_ϵ, the right side of (1.1.24) will be $< \epsilon$. Hence, the left side of (1.1.25) will be $\le \epsilon$ for $n > N_\epsilon$. Note that N_ϵ does not depend on t.

We now want to prove that the limit function f is continuous. Pick $t \in [0,1]$ and $\epsilon > 0$; we need a $\delta > 0$ such that

$$|f(t) - f(y)| < \epsilon \quad \text{whenever } |t - y| < \delta.$$

The triangle inequality says:

$$|f(t) - f(y)| \le |f(t) - f_n(t)| + |f_n(t) - f_n(y)| + |f_n(y) - f(y)|. \tag{1.1.26}$$

According to (1.1.25), the first and third terms on the right can be made $\le \epsilon/3$ by taking n large enough; this estimate was independent of t and y. Pick such an n. The function f_n is continuous, so that

$$|f_n(t) - f_n(y)| < \epsilon/3$$

if $|t - y| <$ some δ. It now follows from (1.1.26) that

$$|f(t) - f(y)| < \epsilon \quad \text{when } |t - y| < \delta,$$

and so f is continuous.[4] ∎

[4] This kind of argument is sometimes called a "three epsilon over three proof"; it tends to pop up when one deals with sequences of functions.

1.1.27 Example. There are many variations on the theme of the last proposition; I present one more example.

Let $C^1([0,1])$ be the linear space

$$\{f : [0,1] \to \mathbf{R} \mid f \text{ and } f' \text{ are continuous}\}.$$

Define the norm

$$\|f\| \stackrel{\text{def}}{=} \sup_{y\in[0,1]} |f(y)| + \sup_{y\in[0,1]} |f'(y)| = \|f\|_\infty + \|f'\|_\infty. \tag{1.1.28}$$

(Check that this is indeed a norm.) I claim that the space $C^1([0,1])$ is complete.

Let $\{f_n\}$ be a Cauchy sequence. Since, for each $t \in [0,1]$,

$$|f_n(t) - f_m(t)| \le \|f_n - f_m\| < \epsilon$$

and

$$|f_n'(t) - f_m'(t)| \le \|f_n' - f_m'\| < \epsilon$$

for n, m sufficiently large, both $\{f_n(t)\}$ and $\{f_n'(t)\}$ are Cauchy sequences of real numbers, and so have limits $f(t)$ and $g(t)$. By Proposition 1.1.23, f and g are continuous functions. To finish the proof, we must show that the limit function f belongs to $C^1([0,1])$, i.e., has a continuous derivative, and that the convergence of f_n to f takes place in the norm (1.1.28). Since

$$\|f_n - f\|_\infty + \|f_n' - g\|_\infty \to 0 \quad \text{as} \quad n \to \infty, \tag{1.1.29}$$

it suffices to check that (as one hopes) the limit g of the functions f_n' is in fact the derivative f' of the limit f of the f_n. Once this is established, (1.1.29) says precisely that $\|f_n - f\| \to 0$.

The fundamental theorem of calculus applies to the f_n:

$$f_n(t) - f_n(0) = \int_0^t f_n'(s)\,ds. \tag{1.1.30}$$

Let $n \to \infty$ in (1.1.30). On the left side, we get $f(t) - f(0)$, since $f_n(t) \to f(t)$ for every $t \in [0,1]$ (including $t = 0$). The limit of the right side is $\int_0^t g(s)\,ds$ — this is seen as follows.

$$\Big|\int_0^t f_n'(s)\,ds - \int_0^t g(s)\,ds\Big| \le \int_0^t |f_n'(s) - g(s)|\,ds \le \int_0^t \|f_n' - g\|_\infty\,ds \le \|f_n' - g\|_\infty,$$

and the last term tends to zero with n.

Hence, (1.1.30) yields, in the limit,

$$f(t) - f(0) = \int_0^t g(s)\,ds,$$

and the fundamental theorem of calculus applies again and says that f' exists and equals g. ∎

I conclude this subsection with two of the many important applications that rely on the Cauchy criterion and/or the notion of completeness.

1.1.31 Theorem (Contraction Mapping Theorem). *Let (M, d) be a complete metric space. Let $f : M \to M$ be a continuous function, and suppose there is a k, $0 < k < 1$, such that*

$$d(f(x), f(z)) \leq k d(x, z) \quad \text{for all } x, z \in M \tag{1.1.32}$$

*(one says that f is a **contraction**). Then there exists a unique $y \in M$ for which $f(y) = y$.* ∎

Proof. We find y as the limit of a sequence of "approximate" solutions of the equation $f(x) = x$. Choose any x_0. Then define

$$x_1 = f(x_0),\ x_2 = f(x_1), \ldots, x_{n+1} = f(x_n), \ldots.$$

By the contraction property, we have

$$\begin{aligned} d(x_{n+1}, x_n) &= d(f(x_n), f(x_{n-1}) \leq k d(x_n, x_{n-1}) \\ &= k d(f(x_{n-1}), f(x_{n-2})) \leq k^2 d(x_{n-1}, x_{n-2}) \\ &\ \ \vdots \\ &\leq k^n d(x_1, x_0). \end{aligned} \tag{1.1.33}$$

Hence, when $m > n$,

$$\begin{aligned} d(x_m, x_n) &\leq d(x_m, x_{m-1}) + \ldots + d(x_{n+1}, x_n) \\ &\leq (k^{m-1} + \ldots + k^n)\, d(x_1, x_0) \\ &= k^n \frac{1 - k^{m-n}}{1 - k}\, d(x_1, x_0) \\ &< \frac{k^n}{1 - k}\, d(x_1, x_0). \end{aligned}$$

Because $k < 1$, the last term can be made arbitrarily small if n is taken large enough. The sequence $\{x_n\}$ is therefore Cauchy, and by the assumed completeness of M, has a limit y.

Now let $n \to \infty$ in the defining relation

$$f(x_n) = x_{n+1},$$

and find, since continuity of f gives $f(x_n) \to f(y)$ as $x_n \to y$, that $f(y) = y$ as desired.

Uniqueness of y again follows from the contraction property. If $f(y_1) = y_1$ and $f(y_2) = y_2$, then

$$0 \leq d(y_1, y_2) = d(f(y_1), f(y_2)) \leq k d(y_1, y_2),$$

and since $k < 1$, it must be that $d(y_1, y_2) = 0$. Hence $y_1 = y_2$. ∎

1.1.34 Examples

A. The hypotheses of the Contraction Mapping Theorem can be weakened as follows. Suppose that the function f (still assumed continuous) maps a closed (hence complete) subset D of M back into D, and that the contraction property (1.1.32) holds on D (but perhaps not on all of M). Then, as long as the initial point x_0 lies in D, the proof goes as before, and produces a unique solution $y \in D$ of $f(y) = y$. This modification is often necessary, since the functions f encountered in applications may not be globally contracting.

B. *Newton's Method.* Let $g : \mathbb{R} \to \mathbb{R}$ be twice continuously differentiable. Suppose that $g(y) = 0$. We want to find a way to approximate the root y of g. Newton's method, as explained in every calculus text, approximates the intersection of the graph of g with the horizontal axis by the x-intercept of a tangent line.

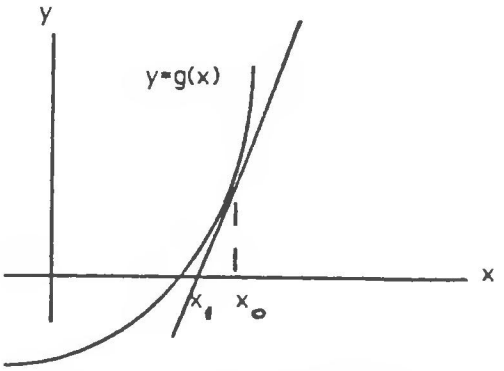


Figure 3. *Newton's method*

Once a "starting guess" x_0 is made, successive (and, one hopes, better) approximations are defined by

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}. \tag{1.1.35}$$

Define

$$f(x) = x - \frac{g(x)}{g'(x)};$$

then Newton's iteration (1.1.35) is simply $x_{n+1} = f(x_n)$, as in the proof of the Contraction Mapping Theorem. A solution y of $f(y) = y$ is automatically a solution of $g(y) = 0$, as long as $g'(y) \neq 0$. Let us assume that $g'(y) \neq 0$, so that y is a *simple* zero of g (as drawn in Figure 3), and see how Theorem 1.1.31 applies.

Since $f'(x) = g(x)g''(x)/(g'(x))^2$ and $g(y) = 0$ by assumption, we have $f'(y) = 0$. It follows from continuity of f' that there is a k, $0 < k < 1$, and a $\delta > 0$ such that $|f'(x)| \leq k$ when $|x - y| < \delta$. By the mean value theorem,

$$|f(x) - f(x')| \leq k|x - x'|$$

as long as $|x-y| < \delta, |x'-y| < \delta$. Since also $|f(x)-y| = |f(x)-f(y)| \leq k|x-y|$, the function f maps the interval $D = \{x \mid |x - y| < \delta\}$ into itself and has the contraction property on D. If the initial guess x_0 lies in the region D, the modified Contraction Mapping Theorem from part A applies, and shows that the Newton iterates tend to the solution y of $f(y) = y$, i.e. to the root y of g. ∎

1.1.36 Proposition (Baire Category Theorem). *In a complete metric space, every residual set is dense.* ∎

Proof. Recall that a residual set is a countable intersection of open dense sets (Definition I.5.1.3). Let $U_1, U_2, \ldots$ be open and dense. We must show that every open set V contains a point belonging to all U_n.

Because U_1 is dense, the open set V must contain a point x of U_1. Therefore the open set $V \cap U_1$ is nonempty, and it contains an open ball B_1 of radius < 1. By the same argument, there is a ball B_2 of radius $< 1/2$ contained in $V \cap U_2 \cap B_1$. By induction, one finds for each n a ball B_n of radius $1/n$, such that $B_n \subset V \cap U_n \cap B_{n-1}$. Pick a point $x_n \in B_n$ for each n. Since $B_m \subset B_n$ when $m > n$, and $x_m \in B_n$ for $m \geq n$, we have $d(x_n, x_m) < 2/n$.

It follows that the sequence $\{x_n\}$ is Cauchy, and by completeness it has a limit x. This limit must be contained in every B_n, so that

$$x \in \bigcap_{n=1}^{\infty} B_n \subset (\bigcap_{n=1}^{\infty} U_n) \cap V \subset V. \quad ∎$$

1.1.37 Corollary. *A complete metric space cannot be the countable union of nowhere dense sets*[5]. ∎

Proof. Suppose $M = \bigcup F_n$, where each F_n is nowhere dense. Then $\emptyset = M^c = \bigcup F_n^c$ by deMorgan's law, but since F_n^c is open dense (Exercise I.3.1.5), this contradicts Baire's Category Theorem. ∎

[5] This is sometimes stated as follows: "A complete metric space is not of first category in itself", or "A complete metric space is of second category in itself".

1.1.38 Corollary. *The rationals are not a G_δ set.* ∎

Note: compare Example I.5.1.20A and Exercise I.5.1.26.

Proof. Suppose $\mathbf{Q} = \bigcap U_n$ where each U_n is open. Then $\mathbf{Q} \subset U_n$, and so $\mathbf{R} = \overline{\mathbf{Q}} \subset \overline{U}_n$, which means that each U_n is dense as well. DeMorgan's law implies that $\mathbf{Q}^c = \bigcap U_n^c$, so

$$\mathbf{R} = \mathbf{Q} \cup \mathbf{Q}^c = \mathbf{Q} \cup \bigcap U_n^c.$$

On the right, Q is the countable union of nowhere dense sets (the singletons $\{r_n\}$), and each U_n^c is nowhere dense. Hence R is the countable union of nowhere dense sets, which contradicts the preceding corollary. ∎

• **1.1.39 Exercise.** Prove: if (M, d) is a complete metric space, and if a set A and its complement A^c are both dense, then they cannot both be F_σ sets. ∎

• **1.1.40 Exercise.** Prove: the irrationals are not a meager subset of R. ∎

• **1.1.41 Exercise.** Prove that $l^\infty(\mathbf{R}, \mathbf{N})$ is complete. ∎

• **1.1.42 Exercise.** Let (M, d) be a metric space. Let $\mathcal{C}(M)$ be the linear space of continuous real-valued functions on M for which the norm

$$\|f\|_\infty \stackrel{\text{def}}{=} \sup_{x \in M} |f(x)|$$

is finite. Prove that $\mathcal{C}(M)$ is complete in the metric derived from this norm. ∎

• **1.1.43 Exercise.** A particular case of the following fact was used in the proof of Proposition 1.1.19. Let V be a normed linear space, fix $\mathbf{x} \in V$, and define $F : V \to \mathbf{R}$ by $F(\mathbf{a}) = \|\mathbf{x} - \mathbf{a}\|$. Then F is continuous. Prove this using the triangle inequality. ∎

• **1.1.44 Exercise.** The functions

$$f_n(t) \stackrel{\text{def}}{=} \sqrt{(t - \tfrac{1}{2})^2 + \frac{1}{n^2}}$$

belong to the space $\mathcal{C}^1([0, 1])$ defined in Example 1.1.27. Do they form a Cauchy sequence? ∎

1.1.45 Exercise. Prove that the Sobolev space $l^{p,k}$ in Example (I.1.1.35) is complete. ∎

1.1.46 Exercise. Is the metric space of sequences of 0's and 1's in Example (I.1.1.45) complete? ∎

1.1.47 Exercise. Let $\mathcal{C}^k([0, 1])$ be the linear space of functions defined on $[0, 1]$ whose first k derivatives exist and are continuous. Equip it with the norm

$$\|f\| \stackrel{\text{def}}{=} \sup_{y \in [0,1]} |f(y)| + \sup_{y \in [0,1]} |f'(y)| + \ldots + \sup_{y \in [0,1]} |f^{(k)}(y)|.$$

Prove completeness. ∎

• **1.1.48 Exercise.** Let $\mathbf{A}$ be a real $n \times n$ matrix, and suppose that $\|\mathbf{A}\| < 1$ for some natural matrix norm $\|\cdot\|$. Use the Contraction Mapping Theorem to prove that the equation $\mathbf{x} + \mathbf{Ax} = \mathbf{b}$ has a unique solution $\mathbf{x}$ for every given $\mathbf{b} \in \mathbf{R}^n$. ∎

The next six exercises deal with the important notion of *connectedness*, Definitions I.3.2.29 or 1.1.12. In the exercises below, the topology on $\mathbf{R}^n$ is always the usual metric topology.

1.1.49 Exercise. Assume that every bounded set of real numbers has a least upper bound and a greatest lower bound. Show that $\mathbf{R}$ is connected. ∎

1.1.50 Exercise. Same assumption as in the preceding exercise. Show that an interval $(a, b) \subset \mathbf{R}$ is connected. ∎

1.1.51 Exercise. Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces, with X connected, and let $f : X \to Y$ be continuous. Show that the range $f(X) \subseteq Y$ is connected. ∎

1.1.52 Exercise. Let $A \subset \mathbf{R}^2$ be the set

$$\{(x, y) \mid x > 0, y = \sin(1/x)\} \cup \{(x, y) \mid x = 0, -1 \leq y \leq 1\}.$$

Show that A is a connected subset of $\mathbf{R}^2$. ∎

1.1.53 Exercise. ***The Intermediate Value Theorem.*** Let $(X, \mathcal{T})$ be a connected topological space, and let $f : X \to \mathbf{R}$ be continuous. Suppose that $f(x) = a$ and $f(y) = b$, with $a < b$. Let $a < c < b$. Then there is a $z \in X$ such that $f(z) = c$. ∎

1.1.54 Exercise. Let f be a continuous real-valued function defined on the unit circle

$$\{(\cos t, \sin t) \mid 0 \leq t < 2\pi\}.$$

Show that there exists a t_0 such that $f(t_0) = f(t_0 + \pi)$. ∎

Subsection 1.2 Completions.

An incomplete metric space has holes, to which some Cauchy sequences try, unsuccessfully, to converge. In most applications, incompleteness is undesirable. We might have the Cauchy sequence $x_{n+1} = f(x_n)$ generated from a contraction mapping f, but the limit, which should solve $y = f(y)$, may not exist — because the sequence converges to a point that is not in our metric space, so to say. It might not help to change metric spaces: in a different metric, the contraction property could be lost, and with it the sequence that at least tried to solve $y = f(y)$. Often, the cure is to enlarge the metric space, and create elements where once there were only holes.

This process is called *completion*. It is, as we shall see shortly, quite easy in principle. If we have an abstract metric space, its points are only letters anyway, and we might as well stick in a few more letters to fill the holes. In practice, however, the elements of the original incomplete metric space are perfectly concrete objects: functions, linear transformations, sequences, etc. The things we add to fill the holes, however, remain thoroughly abstract, not much more concrete than letters from another alphabet. To profit from the process of completion, one must learn to think about and calculate with these new objects. This can be quite complicated, and often requires the development of new ideas. All of Section 2 is devoted to an illustration of this point. In the present subsection, only the abstract construction (which fits naturally at this stage) is described.

The one metric space that is sufficiently simple and still natural enough to motivate the construction is the set of real numbers, with its usual topology. I said earlier (cf. (1.1.11)) that the completeness of the reals could be taken as axiom or emerge as corollary of another, equivalent, axiom. I will now treat it as an axiom — to be precise, I will construct (in outline) the real numbers as the completion of an incomplete metric space. Perhaps there is too much motivation, but the abstract nature of the material sometimes causes difficulties and a little extra talk might not hurt.

The real numbers can be built up in a sequence of steps. First come the natural numbers, then their additive inverses (the negative integers); then come the rationals Q, often defined as equivalence classes of pairs of integers. Finally, the irrationals are introduced, inevitably by a very subtle construction.

To draw attention to the degree of abstraction attached to this step, I mention an approach that will *not* be of further relevance for us. Dedekind[6] defined irrational numbers in the following indirect way. Choose two disjoint subsets $A, B \subset \mathbb{Q}$, with $A \cup B = \mathbb{Q}$, so that every element of A is less than every element of B. (This is called a *Dedekind cut*.) Such a pair of sets defines a real number: intuitively, the number is the irrational separating A and B; rigorously, that irrational is a new object, and it is *defined* to be the pair (A, B). One must now derive the rules of arithmetic and topology for these rather unappealing gadgets.

[6] Julius W. R. Dedekind, 1831-1916.

Cantor, on the other hand, introduced irrationals by the process of completion. This is of interest in the present context. Let us suppose that we have somehow obtained the set Q of rational numbers, together with the usual metric $d(\cdot,\cdot) = |\cdot - \cdot|$.

This is an incomplete metric space. How could one prove this? One needs to find a Cauchy sequence of rational numbers that cannot converge to any rational number. Consider, for example, the sequence

$$x_0 = 2, x_1 = \frac{3}{2}, \ldots, x_{n+1} = \frac{x_n}{2} + \frac{1}{x_n}. \tag{1.2.1}$$

It is the sequence of Newton iterates for the function $g(x) = x^2 - 2$. It is a Cauchy sequence — this follows from Example 1.1.34B (or from Exercise 1.1.10). Its limit x would have to satisfy $x = (x/2)+(1/x)$ or $x^2 = 2$. But one knows from elementary number theory that no rational x satisfies $x^2 = 2$.[7] Ergo, Q cannot be complete.

Let us fix this defect, starting with the square root of 2. We invent a new number, called $\sqrt{2}$. It is the limit of Newton's sequence (1.2.1), and solves $x^2 = \sqrt{2}$. We can also define $r \pm \sqrt{2}$ for all rational r, as the newly created limits of the Cauchy sequences $r \pm x_n$. (This step is peculiar to the real numbers, since elements of metric spaces cannot, in general, be added.) There is an obvious way to extend the metric from Q to the larger set $\mathrm{Q} \cup \{r \pm \sqrt{2} \mid r \in \mathrm{Q}\}$: for example, if $q \in \mathrm{Q}$, define $d(q, \sqrt{2}) = \lim_{n\to\infty} d(q, x_n)$.[8]

I am now happy, having added a pile of points to Q, when you come along and point out that the sequence $z_n = y_n^2$, where

$$y_0 = 2, \quad y_{n+1} = \frac{3y_n}{4} + \frac{1}{2y_n^3} \tag{1.2.2}$$

is also Cauchy, and cannot have a rational limit. Being generous, I am willing to introduce another new number. While studying your sequence (in order to think of an appropriate name), I notice that your z_n and my x_n satisfy

$$\lim_{n\to\infty} |x_n - z_n| = 0. \tag{1.2.3}$$

Your z_n's get closer and closer to my x_n's, which in turn tend to my $\sqrt{2}$. Well! Your new number is really the same as mine, even though all of your rational

---

[7] Suppose p and q are integers with no common factor, satisfying $(p/q)^2 = 2$ or $p^2 = 2q^2$. Since $2q^2$ is even, so is p^2, but then p is even and p^2 has a factor 2^2. Thus q^2 must have another factor 2, whence q has 2 as factor, and p and q have a common factor. Contradiction.

[8] By trying to use the real numbers as prototypical completed metric space, I am creating a logical difficulty for myself: I am claiming to construct, simultaneously, the real numbers as elements of a metric space *and* the real numbers as the ticks on the ruler with which I measure distance. Don't get upset about this. If possible, distinguish the objects x and y in $d(x, y)$ from the numbers $d(x, y)$. The former are the focus of the discussion.

approximations are different from mine. The only way out is to accord equal status to both sequences: everything that needs to be known about this new number can be deduced from either Cauchy sequence.

The creation of the rest of the real numbers could not possibly be done so explicitly. There are far too many irrationals, and most of them don't solve any simple equations. The only realistic strategy is to add, all at once, all the potential limits of all conceivable Cauchy sequences. Suppose $\{u_n\}$ is such a sequence. What shall we call the new number? Given that it probably is nothing as simple as $15^{1/37}$, it is most efficient to call it $\{u_n\}$ - we use the sequence itself as a label for the number, since our information about the number will likely come from the sequence. But other sequences may try to converge to the same limit: that would give rise to too many different "names". This flaw can be fixed by introducing the symbol $[\{u_n\}]$ to mean "the new number defined by the sequence $\{u_n\}$ or by any other sequence $\{u'_n\}$ that satisfies $|u_n - u'_n| \to 0$".

Now to the official development. A new point added to the incomplete metric space will be, as just explained, a collection of Cauchy sequences, all of which try to converge to the same limit. The first bit of vocabulary deals with this operation of "lumping things together".

1.2.4 Definition. *Let X be a set. An **equivalence relation** on X is a subset of $X \times X$, i.e. a subset of the set of pairs $\{(x_1, x_2) \mid x_1, x_2 \in X\}$, satisfying three conditions. (It is customary to write $x_1 \sim x_2$, rather than (x_1, x_2), for those pairs.)*

i) $x_1 \sim x_1$;

ii) if $x_1 \sim x_2$, then $x_2 \sim x_1$;

iii) if $x_1 \sim x_2$ and $x_2 \sim x_3$, then $x_1 \sim x_3$. ∎

To understand the meaning of this, read it aloud, saying "is equivalent to" for every "$\sim$".

1.2.5 Definition. *Let X be a set equipped with an equivalence relation $\sim$. For $x \in X$, the **equivalence class** of x is the set*

$$[x] := \{y \mid y \sim x\}.$$

*A member of an equivalence class is called a **representative** of that class.* ∎

• **1.2.6 Exercise.** Prove: if $x \in [y]$, then $y \in [x]$. ∎

• **1.2.7 Exercise.** In a metric space (M, d), say that $x \sim y$ if $d(x, y) < 1$. Is this an equivalence relation? ∎

1.2.8 Exercise. Let X be the set of 2×2 complex matrices. Say that $A \sim B$ if $B = C \cdot A \cdot C^{-1}$ for some invertible matrix C. Prove that $\sim$ is an equivalence relation, and produce a list of equivalence classes by choosing one representative from each equivalence class. Prove that the function $f(A) = \text{trace } A$ is defined unambiguously on the set of equivalence classes. ∎

The definition of the completion of a metric space comes next, followed by a forbidding-looking proposition which says, with precision, what was sketched above. Look for the main points: the elements of the completion are defined, the distance between the new points is unambigously defined, the "completion" is complete (as it had better be), and the completion is minimal in the sense that it is the closure of the original M.

1.2.9 Definition. *Let (M, d) be a (possibly incomplete) metric space. Its completion is the set $(\overline{M}, \overline{d})$ defined as follows: If $\{x_n\}, \{y_n\}$ are two Cauchy sequences in M, say that they are equivalent, written $\{x_n\} \sim \{y_n\}$, if $d(x_n, y_n) \longrightarrow 0$ as $n \longrightarrow \infty$. $\overline{M}$ is the collection of equivalence classes of Cauchy sequences. For $[\{x_n\}], [\{y_n\}] \in \overline{M}$, define*

$$\overline{d}([\{x_n\}], [\{y_n\}]) = \lim_{n \to \infty} d(x_n, y_n). \quad ∎$$

1.2.10 Proposition.

i) The limit $\lim_{n \to \infty} d(x_n, y_n)$ in the definition of $\overline{d}$ exists.

ii) The function $\overline{d}$ depends only on the equivalence classes

$$[\{x_n\}], [\{y_n\}] \in \overline{M},$$

and not on the choice of representatives ($\{x_n\}$ and $\{y_n\}$).

iii) $\overline{d}$ is a metric.

iv) The metric space $(\overline{M}, \overline{d})$ is complete.

v) For $x \in M$, let $\{\overline{x}_n\}$ be the sequence defined by $\overline{x}_n = x$ for all n. Identify M with the subset $\{[\{\overline{x}_n\}] \mid x \in M\}$ of $\overline{M}$. Then the new metric coincides with the old one on M: $\overline{d}([\{\overline{x}_n\}], [\{\overline{y}_n\}]) = d(x, y)$.

vi) M is dense in $\overline{M}$. ∎

The completion $(\overline{M}, \overline{d})$ of (M, d) has one other important property: it is the *unique* completion in which M is dense. That last requirement is natural: one does not want the completion to be too large. The rational numbers $\mathbf{Q}$, for example, can be embedded in the complete metric space $\mathbf{R}^2$ (usual metric) along the x-axis, but adding a whole plane's worth of points that have nothing to do with Cauchy sequences in $\mathbf{Q}$ is artificial.

Two metric spaces are considered to be the same if it is possible to identify points and distances between points:

1.2.11 Definition. *Let (M, d) and (M', d') be metric spaces. An isometry from (M, d) to (M', d') is a function $F : M \to M'$ satisfying*

$$d'(F(x), F(y)) = d(x, y) \quad \text{for all } x, y \in M.$$

If the range of F is all of M', then the two metric spaces are said to be isometric. ∎

1.2.12 Proposition. *Let (M, d) be a metric space with completion $(\overline{M}, \overline{d})$. Suppose that (M', d') is a complete metric space, and suppose that there is an isometry $F : M \to M'$ whose range $F(M)$ is dense in M'. Then (M', d') and $(\overline{M}, \overline{d})$ are isometric.* ∎

• **1.2.13 Exercise.** Prove Proposition 1.2.10. ∎

• **1.2.14 Exercise.** A metric space (M, d) is called *bounded* if there is a $K > 0$ such that $d(x, y) \leq K$ for all $x, y \in M$. Let (M, d) be bounded, and suppose that (M, d) and (M', d') are isometric. Show that (M', d') is bounded. ∎

• **1.2.15 Exercise.** Show by example: two metrics d and d' may define the same topology on a set M, but (M, d) and (M', d') may fail to be isometric. ∎

• **1.2.16 Exercise.** Prove Proposition 1.2.12. ∎

1.2.17 Exercise. Verify that the sequence (1.2.2) has limit $\sqrt{2}$. ∎

1.2.18 Exercise*. a) Show that

$$d(x, y) = \frac{|x - y|}{\sqrt{(1 + x^2)(1 + y^2)}}$$

defines a metric on $\mathbf{R}$. b) Show that $\mathbf{R}$ is not complete in this metric. c) Find the completion. ∎

Subsection 1.3 Compactness

The Cauchy criterion is often too restrictive. One might have a sequence of approximate solutions to an equation, but the method for obtaining that sequence might not allow one to decide whether or not it converges. At this point, additional structure may be useful: one wants conditions that ensure some sort of convergence, but are weaker than the Cauchy criterion (it may be difficult to check, or may not even be satisfied). The notion of *compactness* can help in this situation. It has many other ramifications, some of which will be described in this chapter; the present subsection is devoted to basic definitions and propositions.

Like many basic topological ideas, compactness can be defined in different ways depending on the type of topological space under consideration. The metric space definition can be made quite intuitive; we give it first.

1.3.1 Terminology. *Let $\{x_n\}$ be a sequence. Choose an infinite subset $\{n_j\}$ of $1, 2, \ldots$, so that $n_1 < n_2 < \ldots$. The sequence $\{y_j\}$ defined by $y_j = x_{n_j}$, $j = 1, 2, 3, \ldots$ is said to be a subsequence of $\{x_n\}$.* ∎

1.3.2 Definition. *Let (M, d) be a metric space. A subset $A \subset M$ is said to be compact, if every sequence $\{x_n\} \subset A$ has a convergent subsequence, $\{x_{n_j}\}$, whose limit $\lim_{j\to\infty} x_{n_j}$ lies in A. If every sequence in M has a convergent subsequence, then M is said to be a compact metric space.* ∎

1.3.3 Examples

A. The sequence $x_n = (-1)^n$ has the subsequence $x_{2n} = +1$ which (trivially) converges to 1. The countable set of rational numbers can be thought of as a sequence $\{r_n\}$ in many different ways; it has uncountably many convergent subsequences, at least one for every irrational number.

B. The set of real numbers, with the usual metric, is not compact: the sequence $x_n = n$ does not have a subsequence that converges to a point of R.

C. *A compact metric space must be complete.* Let $\{x_n\}$ be a Cauchy sequence. Compactness implies that a subsequence converges: $\lim_{j\to\infty} x_{n_j} = x$. Now pick $\epsilon > 0$. If j is large enough (or equivalently, n_j is large enough), $d(x_{n_j}, x) < \epsilon/2$. By the Cauchy property, if m and n_j are large enough, then $d(x_m, x_{n_j}) < \epsilon/2$. The triangle inequality

$$d(x_m, x) \leq d(x_m, x_{n_j}) + d(x_{n_j}, x) < \epsilon$$

now shows that the original sequence $\{x_m\}$ also converges to x.

By part B, a complete metric space need not be compact.

D. *A compact subset A of a metric space must be closed.* Indeed, if x is a point of closure of A (Definition I.3.5.8), then for every integer $n > 0$ there is an $x_n \in A$ such that $d(x, x_n) < 1/n$. By compactness, this sequence $\{x_n\}$ has a convergent subsequence, whose limit is evidently x (in fact, the sequence itself converges to x). Compactness also requires that the limit x be in A. Hence A contains all its points of closure and is therefore closed.

The converse is not true. For example, R is a closed subset of itself, and is not compact.

E. *A compact subset A of a metric space must be bounded,* i.e., there is a $K > 0$ such that $d(x, y) < K$ for all $x, y \in A$. For the proof, first note that if A is unbounded, then for every x_0 and every $K_1 > 0$, there will be an x such that $d(x_0, x) > K_1$. For, if there were an x_0 and K_1 such that $d(x_0, x) < K_1$ for all $x \in A$, the triangle inequality would give

$$d(x, y) \leq d(x, x_0) + d(x_0, y) < K_1 + K_1 \stackrel{\text{def}}{=} K$$

for all $x, y \in A$, implying boundedness.

Suppose now that A is compact but not bounded. Pick any x_0. For each integer $n > 0$, there is an $x_n \in A$ with $d(x_0, x_n) > n$. By compactness,

the sequence $\{x_n\}$ has a convergent subsequence, $x_{n_j} \to x \in A$. But then $d(x_{n_j}, x) < 1$ for n_j large enough, whence

$$d(x_0, x_{n_j}) \le d(x_0, x) + d(x, x_{n_j}) < d(x_0, x) + 1 \stackrel{\text{def}}{=} K_1,$$

which contradicts $d(x_0, x_{n_j}) > n_j \to \infty$.

The converse is not true: a bounded subset of a metric space need not be compact, not even if it is closed (as compact sets must be, by part D). There cannot be a counterexample in $\mathbf{R}^n$ because, as will be shown in Corollary 1.3.6, closed bounded subsets of $\mathbf{R}^n$ *are* compact. Here is an infinite-dimensional counterexample.

The unit ball in $l^2(\mathbf{R}, \mathbf{N})$ is closed and bounded, but it is not compact. Indeed, the sequence $\mathbf{e}^{(n)} = (0, 0, \ldots, 1, 0, \ldots)$, with a 1 in the n^{th} place and 0's elsewhere, has no convergent subsequence, since $\|\mathbf{e}^{(n)} - \mathbf{e}^{(m)}\|_2 = \sqrt{2}$ for all n, m. ∎

So far, we have not seen any examples of compact sets or metric spaces. Compactness is like completeness in the following sense: the description of compact subsets of $\mathbf{R}$ is very much tied in with the real number axioms, and the characterization of compact subsets of function spaces relies ultimately on compactness in $\mathbf{R}$.

1.3.4 Proposition. *A subset $A \subset \mathbf{R}$ (in the usual metric) is compact if, and only if, it is closed and bounded.* ∎

Proof. If A is compact, it must be closed and bounded by Examples 1.3.3D,E. Conversely, suppose that A is bounded and closed, and let $\{x_n\} \subset A$. We must find a convergent subsequence. The idea is the following. We expect the x_n to pile up somewhere (perhaps at many points). To find just one accumulation point, which will turn out to be the limit of a subsequence, we start at the far right on the real axis and move to the left, asking at each point: are there still infinitely many x_n's to the left? The boundary between the region where the answer is "yes" and the region where the answer is "no" is the desired accumulation point.

Because A is bounded, there is a $K > 0$ such that $-K < x_n < K$ for all n. Let

$$E = \{y \mid x_n < y \text{ for infinitely many } n\}.$$

The set E is bounded below by $-K$, and hence (this is basically an axiom for the reals) it has a greatest lower bound L.

Step 1. Let $x_{n_1} = x_1$.

Step 2. $L - 1/2$ is a lower bound of E, and there are only finitely many x_n less than $L - 1/2$. On the other hand, $L + 1/2$ is not a lower bound of E, and so there are infinitely many x_n between $L - 1/2$ and $L + 1/2$. Pick such an x_n with $n > n_1 = 1$, and call it x_{n_2}.

Step 3. Likewise, $(L - 1/3, L + 1/3)$ contains infinitely many x_n; pick one with $n > n_2$, and call it x_{n_3}.

Proceeding inductively, we obtain a subsequence $\{x_{n_j}\}$ of $\{x_n\}$. By construction, $x_{n_j} \in (L - 1/j, L + 1/j)$, which implies that $x_{n_j} \to L$. Since A is closed, $L \in A$, and the compactness condition is verified. ∎

1.3.5 Corollary. *Every bounded sequence in* $\mathbf{R}$ *has a convergent subsequence.* ∎

Proof. Let $A = \overline{\{x_n\}}$. A is a closed and bounded set, and therefore $\{x_n\} \subseteq A$ has a convergent subsequence with limit in A. ∎

1.3.6 Corollary (Heine-Borel Theorem)[9]. *A subset* A *of* $\mathbf{R}^n$ *(in the usual metric) is compact if, and only if,* A *is closed and bounded.* ∎

Proof. The proof is only given for $\mathbf{R}^2$, where the notation remains manageable. The straightforward generalization to $\mathbf{R}^n$ is left for the reader to write out. Suppose that A is closed and bounded, and let $\{\mathbf{x}^{(n)}\} \subset A$. The bounded sequence $\{x_1^{(n)}\}$ of first coordinates has a convergent subsequence $\{x_1^{(n_j)}\}$ by the preceding corollary. Set $\mathbf{y}^{(j)} = \mathbf{x}^{(n_j)}$. The bounded sequence $\{y_2^{(j)}\}$ of second coordinates has a convergent subsequence $\{y_2^{(j_i)}\}$. Set $\mathbf{z}^{(i)} = \mathbf{y}^{(j_i)}$. This is a subsequence of $\{\mathbf{x}^{(n)}\}$. Since both coordinates converge in $\mathbf{R}$, the sequence $\{\mathbf{z}^{(i)}\}$ converges in the l^∞ metric of $\mathbf{R}^2$, and, by equivalence of the l^∞ and l^2 metrics, also in the latter. Since A is closed, the limit belongs to A, and the compactness condition is verified. ∎

1.3.7 Corollary (Bolzano-Weierstrass Theorem). *Every bounded sequence in* $\mathbf{R}^n$ *has a convergent subsequence.* ∎

Proof. Exercise. ∎

One of the most important consequences of compactness is probably familiar in Euclidean space from calculus or advanced calculus, but it holds equally well in abstract metric spaces, as long as they are compact:

1.3.8 Proposition. *Let* (M, d) *be a compact metric space, and let* $f : M \to \mathbf{R}$ *be a continuous function. Then* f *assumes its maximum and minimum values: there are points (not necessarily unique)* $x_\pm \in M$ *such that*

$$a \stackrel{\text{def}}{=} f(x_-) \leq f(x) \leq f(x_+) \stackrel{\text{def}}{=} A$$

[9] Heinrich Eduard Heine (1821-1861), Emile Borel (1871-1956)

for all $x \in M$. ∎

Proof. We first show that the range of f is bounded above. Suppose not: then, for each integer n, there is an $x_n \in M$ such that $f(x_n) > n$. The sequence $\{x_n\}$ has a convergent subsequence $x_{n_j} \to x \in M$. By continuity, $f(x_{n_j}) \to f(x)$, but this contradicts $f(x_{n_j}) > n_j \to \infty$.

Since the set $\{f(x) \mid x \in M\}$ is bounded from above, it has a least upper bound A. For each integer $n > 0$, there is an $x_n \in M$ such that $A - 1/n \leq f(x_n) \leq A$ (if not, then $A - 1/n$ would be an upper bound smaller than the least one). The sequence $\{x_n\}$ has a convergent subsequence: $x_{n_j} \to x_+$. Then

$$A - \frac{1}{n_j} \leq f(x_{n_j}) \leq A \quad \text{for all } n_j$$

implies that $f(x_{n_j}) \to A$. By continuity of f, $f(x_+) = A$.

The argument for the minimum is similar. ∎

1.3.9 Example. A continuous function on a non-compact space need not be bounded: take $f(x) = x$ on $\mathbf{R}$. Standard functions, like $\sin x$, $1/(1+x^2)$, and $\tanh x$ show that bounded continuous functions on a non-compact space may or may not assume their maximum and/or minimum values. ∎

• **1.3.10 Exercise.** Prove that a closed subset of a compact metric space is compact. ∎

• **1.3.11 Exercise.** Recall from Example 1.3.3E that the unit ball B in $l^2(\mathbf{R}, \mathbf{N})$ is not compact. Find a continuous function $f : B \to \mathbf{R}$ whose range has least upper bound 1, but which satisfies $f(\mathbf{x}) < 1$ for all $\mathbf{x} \in B$. (Hint: try something like $f(\mathbf{x}) = \sum_0^\infty a_j^2 x_j^2$, with suitable coefficients a_j.) ∎

• **1.3.12 Exercise.** Prove: if A is a compact subset of a metric space, then there are points $a, b \in A$ whose distance is maximal, i.e. $d(a, b) \geq d(x, y)$ for all $x, y \in A$. ∎

• **1.3.13 Exercise.** Prove: if A and B are disjoint closed subsets of a metric space, and A is also compact, then the distance between A and B is strictly positive, i.e. there exists a $\delta > 0$ such that $d(a, b) \geq \delta$ for all $a \in A$ and $b \in B$. Show by example that this need not be the case if neither A nor B is compact. ∎

• **1.3.14 Exercise.** Let f be a continuous function from a compact metric space (M, d) to a metric space (M', d'). Show that the range $f(M)$ is a compact subset of M'. ∎

• **1.3.15 Exercise.** Use the preceding exercise and the characterization of compact subsets of $\mathbf{R}$ in Proposition 1.3.4 to give a more streamlined proof of Proposition 1.3.8. ∎

• **1.3.16 Exercise.** Let A be a compact subset of a metric space (M, d). Prove that for every $\epsilon > 0$, there exists a finite subset $S_\epsilon = \{a_1, \ldots, a_{N_\epsilon}\}$ of A with the following property: for every $x \in A$ there is an $a_i \in S_\epsilon$ such that $d(x, a_i) < \epsilon$. (Such a set S_ϵ is called a *finite ϵ-net.*) ∎

• **1.3.17 Exercise.** This continues the preceding exercise. Suppose A is a closed subset of a complete metric space (M, d) that admits a finite ϵ-net for every $\epsilon > 0$. Show that A is compact. ∎

1.3.18 Exercise. Let (M, d) be a compact metric space. Prove that there exists a countable dense subset of M. (Such a space is called *separable*, cf. Exercise I.3.5.39.) ∎

The principal ingredient in the definition of compactness is convergence, and we know that this notion depends only on the metric topology of a metric space (M, d), and not on the particular metric that happens to induce the topology. It should therefore be possible to express compactness more intrinsically, in terms of open sets — much as Definition I.3.2.27 translated the $f(x_n) \to f(x)$ definition of continuity into the sparse statement "f^{-1}(open) is open". There is indeed an adaptation of "compactness" to general topological spaces, which is often simpler to use even when one deals with metric spaces. I did not start with it, because, being more abstract, it is more difficult to appreciate. Now I give the topological definition and deduce some consequences. In the end, I will show that it is equivalent to (1.3.2) for metric spaces.

1.3.19 Terminology. *Let $(X, \mathcal{T})$ be a topological space. A collection $\mathcal{U}$ of open sets is said to be an* **open covering** *of X if the union $\bigcup_{O \in \mathcal{U}} O$ equals X. A collection $\mathcal{U}$ of open subsets of X is called an* **open covering** *of $A \subset X$ if the union contains A. If there is a finite subset $\{O_1, \ldots, O_n\}$ of $\mathcal{U}$ whose union is X (respectively, contains A), one says that it is a* **finite subcovering** *of X (respectively, A).* ∎

1.3.20 Definition. *If every open covering of a topological space $(X, \mathcal{T})$ has a finite subcovering, one says that the space X is* **compact**. *If every open covering of a subset $A \subset X$ contains a finite subcovering of A, one says that A is a* **compact subset of** *X.* ∎

1.3.21 Examples

A. $\mathbf{R}$ is not compact. For, let $O_n = (-n, n)$, and let $\mathcal{U} = \{O_n | n \geq 1\}$. Clearly, $\mathcal{U}$ is an open covering of $\mathbf{R}$, but no finite subset of $\mathcal{U}$ covers $\mathbf{R}$.

B. The open interval $(0, 1)$ is not compact. For, let $O_n = (1/n, 1 - 1/n)$, $n \geq 3$. The union of all these intervals covers $(0, 1)$, but no finite subset does. ∎

1.3.22 Remark. The open coverings in Definition 1.3.20 of compactness are allowed to contain uncountably many open sets. To prove that a space (or a subset) is compact, one must prove that every open covering, no matter how huge and ugly, has a finite subcovering. To prove that a space (or subset) is not compact, one need only exhibit a particular $\mathcal{U}$ that has no finite subcovering, and one will try to make this $\mathcal{U}$ as simple as possible. ∎

Definition 1.3.20 of compactness is formulated in terms of open sets, but, by taking complements, one can get an equivalent characterization in terms of closed sets; sometimes, this is a more useful formulation.

1.3.23 Definition. *A collection $\mathcal{F}$ of closed subsets of a topological space $(X, \mathcal{T})$ has the finite intersection property if the intersection of finitely many sets $F_1, \ldots, F_n \in \mathcal{F}$ is always nonempty.* ∎

1.3.24 Proposition. *A topological space $(X, \mathcal{T})$ is compact if, and only if, for every collection $\mathcal{F}$ of closed sets that has the finite intersection property, the following is true:*

$$\bigcap_{F \in \mathcal{F}} F \neq \emptyset. \text{ ∎}$$

Proof. Suppose X is compact, and that the intersection $\bigcap_{F \in \mathcal{F}}$ is empty. Then $\bigcup_{F \in \mathcal{F}} F^c = X$ by de Morgan's law. So $\{F^c \mid F \in \mathcal{F}\}$ is an open covering of X and has a finite subcovering $\{F_1^c, \ldots, F_n^c\}$. But

$$\bigcup_{j=1}^{n} F_j^c = X \Longrightarrow \bigcap_{j=1}^{n} F_j = \emptyset,$$

which contradicts the finite intersection property. The converse is proved in a similar way. ∎

• **1.3.25 Exercise.** Prove the converse of Proposition 1.3.24. ∎

1.3.26 Remark. In Definition 1.3.20 and Proposition 1.3.24, the open covering and the family of closed sets must be allowed to contain *uncountably* many members. There are (somewhat exotic) examples of topological spaces in which every countable open covering has a finite subcovering, but some uncountable open covering does not. ∎

• **1.3.27 Exercise.** A certain economy of language is achieved by the following construction. Let $(X, \mathcal{T})$ be a topological space, and let A be a subset of X. One makes A into a topological space in its own right, by defining the open subsets of A to be

$$\mathcal{T}_A = \{U \cap A \mid U \in \mathcal{T}\}.$$

a) Prove that this collection satisfies the axioms for a topology. ($\mathcal{T}_A$ is called the *relative topology on A.*)

b) One can now rephrase Definition 1.3.20: a subset $A \subseteq X$ is called *compact*, if it is a compact topological space in the relative topology. Show that this definition is equivalent to the original one.

The advantage of this language is that one need not make a distinction between "compact topological space" and "subset of a compact topological space." The latter concept is now subsumed in the first one. ∎

To illustrate how the new definition (1.3.20) is used, I will now reprove the fundamental Proposition 1.3.4. Notice that, in spite of the new baggage of open sets and coverings, the crucial step is still an application of the least upper (≡ greatest lower) bound axiom.

1.3.28 Proposition. *A finite closed interval* $[a, b] \subset \mathbf{R}$ *is compact.* ∎

Proof. Let $\mathcal{U}$ be an arbitrary open covering of $[a, b]$. Let E be the set of $x \in [a, b]$ with this property: the interval $[a, x]$ can be covered by finitely many sets from $\mathcal{U}$. This set E is nonempty; it contains a itself, since some $O \in \mathcal{U}$ contains a and hence covers the "interval" $[a, a] = \{a\}$. Let c be the *least upper bound* of E. (Recall: the existence of a least upper bound is an axiom for the real number system; completeness of the reals is an equivalent form of that axiom). Because b is *an* upper bound of E, we must have $c \leq b$.

Now, some $O \in \mathcal{U}$ contains c ($\mathcal{U}$ covers $[a, b]$), and this open O contains an interval $I = (c - \epsilon, c + \epsilon)$ for some ϵ. Because c is a *least* upper bound of E, there must be a point $y \in E$ in that interval I. Otherwise, $c - \epsilon/2$ would be a smaller upper bound for E, contradicting the definition of c. But $[a, y]$ is covered by a finite collection $O_1, \ldots, O_n$ of sets from $\mathcal{U}$, and the finite collection made up of those sets *plus* the one additional set O covers the larger interval $[a, c + \epsilon/2]$.

This argument shows that c cannot be strictly less than b, because then c cannot be an upper bound for E. So $c = b$, but then all of $[a, b]$ can be covered by finitely many sets $O_1, \ldots, O_n, O$ from $\mathcal{U}$. ∎

1.3.29 Corollary. *A closed, bounded subset* A *of* $\mathbf{R}$ *is compact.* ∎

Proof. Suppose that A is closed and bounded. It is therefore contained in a finite interval $[a, b]$. Let $\mathcal{U}$ be an open covering of A, and let $\mathcal{U}^* \stackrel{\text{def}}{=} \mathcal{U} \cup \{A^c\}$; this is an open covering of $[a, b]$. We know from the proposition that there is a finite subcovering of $[a, b]$ by sets $O_1, \ldots, O_n \in \mathcal{U}^*$. If none of these O_j is the extra set A^c, we are done: the O_j come from $\mathcal{U}$, cover $[a, b]$ and so also cover $A \subset [a, b]$. If one of the O_j is the extra set A^c, throw it out: $A^c \cap A = \emptyset$, so while A^c may be needed to cover $[a, b]$, it cannot help with covering A, and the remaining sets O_j from $\mathcal{U}$ form a finite covering of A. ∎

In $\mathbf{R}^n$, compact sets are the same as closed and bounded sets. In more general metric spaces, those conditions are necessary, but no longer sufficient, for compactness (cf. Example 1.3.3E). In topological spaces, "boundedness" no longer makes sense, but there is still some connection between compactness and closedness.

1.3.30 Proposition. *A closed subset* F *of a compact topological space* (X, T) *is compact.* ∎

Proof. (See also Exercise 1.3.10, where the argument was supposed to be based on convergence of subsequences.) Let $\mathcal{U}$ be an open covering of F. As in the proof of the last corollary, let $\mathcal{U}^* = \mathcal{U} \cup \{F^c\}$. $\mathcal{U}^*$ covers X, hence there is a finite subcovering. If it doesn't include F^c, good – we are done. If it does include F^c, throw that set out: it is not needed to cover F. ∎

1.3.31 Proposition. *A compact subset F of a Hausdorff space is closed.* ∎

Proof. We need to show that F^c is open. Fix $x \in F^c$. For every $y \in F$, there are disjoint neighborhoods O_y of y and $O_y^{(x)}$ of x (that is the Hausdorff property). The collection $\{O_y \mid y \in F\}$ is an open covering of F, and by compactness of F there is a finite subcovering $\{O_{y_1}, \ldots O_{y_n}\}$. The union of these n sets contains F. Correspondingly, the *finite* intersection $U = O_{y_1}^{(x)} \cap \ldots \cap O_{y_n}^{(x)}$ is open, contains x, and is disjoint from F (since it is disjoint from all O_{y_j}). The set U is an open set containing x and contained in F^c, which shows that F^c is open.

See Exercise 1.3.38 for a counterexample in a non-Hausdorff space. ∎

Remark. This is a typical compactness argument. Each point y of a space is assigned an open neighborhood O_y according to some rule. This gives an open covering of the space, usually by a huge number of sets. Compactness allows one to choose a finite subcovering. ∎

We could go on for a while, using Definition 1.3.20 to prove every result previously derived from Definition 1.3.2. It is probably time to establish, once and for all, the equivalence of those two definitions.

1.3.32 Proposition. *Let (M, d) be a metric space. Every open covering of M has a finite subcovering if, and only if, every subsequence has a convergent subsequence.* ∎

Remark. The proof works equally well for subsets $A \subset M$: every open covering of A has a finite subcovering if, and only if, every sequence in A has a subsequence that converges to a point of A. ∎

Proof. If every open covering of M has a finite subcovering, it is fairly easy to deduce the existence of convergent subsequences — see Exercise 1.3.33, with hint.

The converse is more difficult. Assuming that every sequence in M has a convergent subsequence, we first prove that every *countable* covering of M has a finite subcovering.

Let $O_1, O_2, \ldots$ be an open covering, and suppose that there is no finite subcovering. Then, for each n, the union

$$\bigcup_{k=1}^{n} O_k$$

does not cover M, and there is a point

$$x_n \in M - \bigcup_{k=1}^{n} O_k.$$

The x_n are all distinct,*) and by hypothesis there is a convergent subsequence $x_{n_j} \to x \in M$. This x cannot belong to any $\cup_{k=1}^{n} O_k$: if it did, that union would be a neighborhood of x, and would contain all x_{n_j} for $j >$ some J by definition of convergence. So the O_n do *not* cover M, a contradiction.

We must now prove that every uncountable open covering of M also has a finite subcovering (cf. Remark 1.3.26). To this end, we will show that the hypothesis (convergent subsequences) implies: there is a countable set $Y = \{y_1, y_2, \ldots\} \subseteq M$ which is dense in M. Let us see how the conclusion follows from this.

Let $\mathcal{U}$ be an open covering of A. Let $x \in M$. There is an O_α containing x, and hence a ball $B(x, 1/m) \subseteq O_\alpha$. Because Y is dense, there is a $y_n \in B$ such that

$$y_n \in B(x, \frac{1}{4m}) \subset B(x, \frac{1}{m}) \subseteq O_\alpha,$$

and then

$$x \in B(y_n, \frac{1}{2m}) \subset B(x, \frac{1}{m}) \subseteq O_\alpha$$

(see Figure 4).

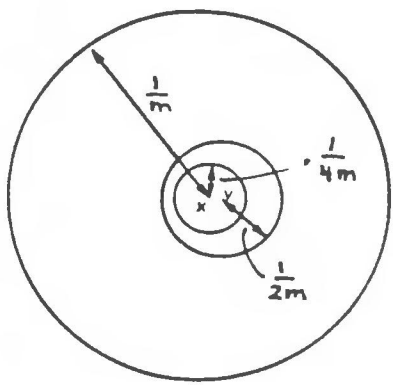


Figure 4. *Extracting a countable subcovering*

For each pair (n, k), pick one of the sets O_α containing $B(y_n, 1/k)$, and call it $O_{(n,k)}$. There may be some (n, k) for which the ball $B(y_n, 1/k)$ is contained in no O_α. We have seen, however, that every $x \in M$ is contained in at least one such ball which in turn lies entirely within some O_α. Hence, there is a countable subcollection $\{O_{(n,k)}\}$ of $\{O_\alpha\}$, indexed by a subset of all pairs of positive integers, which still covers M. The first part of the proof now applies: this countable subcovering has a finite subcovering, as was to be shown.

*) - the x_n may not all be distinct, but this is irrelevant.

We now prove existence of a countable dense subset of M. Fix an integer $m > 0$. I claim that there exists a finite set of points $Y_m = \{z_1, \ldots, z_N\}$ (N depending on m), such that: for every $x \in M$, there is a z_j with $d(x, z_j) < 1/m$. (This is a finite $1/m$-net, as defined in Exercise 1.3.16; see Exercises 1.3.17 and 1.3.18 for related problems.) Indeed, pick any $z_1 \in M$. Either all $x \in M$ satisfy $d(x, z_1) < 1/m$, in which case we are done, or there is a z_2 such that $d(z_1, z_2) > 1/m$. Next, either all $x \in M$ satisfy $d(x, z_1) < 1/m$ or $d(x, z_2) < 1/m$ and we are done, or there is a z_3 with

$$d(z_1, z_3) > 1/m, \quad d(z_2, z_3) > 1/m.$$

If the process does not terminate after a finite number of steps, we have an infinite sequence $\{z_j\}$ such that $d(z_i, z_j) > 1/m$ for all $i \neq j$. By assumption, there is a convergent subsequence $\{z_{k_j}\}$, and by the Cauchy criterion, $d(z_{k_i}, z_{k_j}) < 1/m$ for i, j large enough. This is a contradiction.

Let $Y = \bigcup_{m=1}^{\infty} Y_m$. This is a countable set, and it is dense in M: given $x \in M$ and $\epsilon > 0$, there is an m with $1/m < \epsilon$ and a $y \in Y_m \subset Y$ such that $d(x, y) < 1/m < \epsilon$. This finishes the proof. ∎

• **1.3.33 Exercise.** Show that every infinite subset A of a compact topological space has at least one accumulation point. (Hint: Take a sequence $\{x_1, x_2, \ldots\} \subset A$. Let F_n be the closed set $\overline{\{x_n, x_{n+1}, \ldots\}}$ (the bar denotes closure). The family $\mathcal{F} = \{F_1, F_2, \ldots\}$ has the finite intersection property (1.3.23). Now use (1.3.24) to find a point $y \in \bigcap F_n$, and show that y is a limit point of A. ∎

• **1.3.34 Exercise.** Let f be a continous function from a compact topological space $(X, \mathcal{T})$ to a topological space $(Y, \mathcal{S})$. Then the range of f is a compact subset of Y.

Special case: the range of a continuous function from a bounded interval $[a, b]$ to $\mathbf{R}$ is a closed, bounded subset of $\mathbf{R}$. ∎

• **1.3.35 Exercise.** Let $F_1 \supset F_2 \supset F_3 \ldots$ be a sequence of nonempty closed, bounded subsets of $\mathbf{R}^n$. Show that

$$\bigcap_{n=1}^{\infty} F_n \neq \emptyset.$$

Show by example that the result may fail when the sets are not bounded. ∎

1.3.36 Exercise. Definition. A topological space $(X, \mathcal{T})$ is said to be *locally compact* if every point has a neighborhood whose closure is compact.
Prove that $\mathbf{R}$ is locally compact. Prove that $l^2(\mathbf{R}, \mathbf{N})$ is not locally compact. ∎

1.3.37 Exercise*. Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces, with X compact and Y Hausdorff. Let f be a one-to-one continuous function from X onto Y. Show that the inverse function f^{-1} is also continuous.

If $(X, \mathcal{T})$ is not compact, or $(Y, \mathcal{S})$ is not Hausdorff, this conclusion may not be true. For instance, consider the map $f(x) = x$ from $\mathbf{R}$ with the usual topology to $\mathbf{R}$ with the Zariski topology. Find other examples. ∎

1.3.38 Exercise°. The set $\mathbf{R} - \{0\} \subset \mathbf{R}$ is not closed in the Zariski topology on $\mathbf{R}$, but it is compact. This example shows that the Hausdorff hypothesis in Proposition 1.3.31 is essential. ∎

1.3.39 SUPPLEMENT. Compactness of products.

Product topologies were defined in Supplement I.3.4.25. It is an important fact, used frequently in analysis, that compactness of the factors is inherited by the product.

1.3.40 Tihonov's Theorem. *Let $\{(X_\lambda, \mathcal{T}_\lambda)\}$ be a (possibly uncountable) family of compact topological spaces. Then the product $\prod X_\lambda$, with the product topology, is again compact.* ∎

We will not prove this theorem, but the special case of a product of *two* topological spaces is easy, and is stated as an exercise.

• **1.3.41 Exercise.** Let $(X_1, \mathcal{T}_1)$ and $(X_2, \mathcal{T}_2)$ be compact topological spaces. Show that $X_1 \times X_2$ is compact in the product topology defined in (I.3.4.26). (Hint: use Exercise 1.3.34 and the fact that open set of the form π_i^{-1}(open) are a subbase for the topology on $X_1 \times X_2$, by Definition I.3.4.26. ∎

• **1.3.42 Exercise.** Show that the "one-point compactification" of $\mathbf{R}$ defined in Exercise I.3.2.37 is a compact topological space. (This explains the terminology). Note: product topology does not enter; the example is just a prelude to the next exercise. ∎

• **1.3.43 Exercise.** Let $\dot{\mathbf{R}}$ be the one-point compactification of $\mathbf{R}$. For $\lambda \in \mathbf{R}$, let $X_\lambda = \dot{\mathbf{R}}$. Let X be the product $\prod_{\lambda \in \mathbf{R}} X_\lambda$. Show that this space is compact.

The huge space X may be interpreted as the set of all functions (continuous or not) defined on $\mathbf{R}$, with values in $\mathbf{R} \cup \{\infty\}$. Its compactness is crucial in one of the constructions of Brownian motion. ∎

Subsection 1.4 Uniformity

A relation between convergence and compactness is built into the very definition of compactness. Proposition 1.3.8 shows that there is also interaction between continuity and compactness. In this subsection, we develop more connections between those three concepts. In one example, compactness influences continuity, while other examples prepare the ground for compactness arguments in later sections. The unifying thread is the notion of *uniformity*; the word is not technical, and is meant only to suggest common features of the definitions given below.

1.4.1 Definition. *Let $f : (M, d) \longrightarrow (M', d')$ be a continuous function from one metric space to another. We say that f is* ***uniformly continuous*** *if for every $\epsilon > 0$ there is a $\delta > 0$ such that $d'(f(x), f(y)) < \epsilon$ whenever $d(x, y) < \delta$.* ∎

In the definition of (plain old) continuity of f at a point x_0, the δ depends on ϵ *and in general* also on x_0. "Uniform continuity" means that, once ϵ is chosen, the same δ works for *all* x_0's.

1.4.2 Examples

A. The function $f(x) = 1/x$ is continuous on $(0, \infty)$, but it is not uniformly continuous. Fix an $\epsilon > 0$, and take $x_0 = 1/n$. To check continuity of f at $1/n$, we should make $|f(1/n) - f(y)| < \epsilon$, i.e., we need $|n - 1/y| < \epsilon$, or $1/(n+\epsilon) < y < 1/(n-\epsilon)$. The largest possible δ is $1/n - 1/(n+\epsilon) = \epsilon/(n(n+\epsilon))$. This tends to zero as $n \longrightarrow \infty$: when we take x_0 closer and closer to 0, we must make δ ever smaller. There is no one δ that will work for all points x in the domain $(0, \infty)$.

B. Let f be a continuously differentiable function defined on $\mathbf{R}$. Suppose that the derivative of f is bounded: $|f'(y)| \leq$ some K for all $y \in \mathbf{R}$. Then, by the mean value theorem,

$$\begin{aligned} |f(x) - f(x_0)| &= |f'(\xi)|\,|x - x_0| \quad \text{for some } \xi \text{ between } x \text{ and } x_0. \\ &\leq K|x - x_0|. \end{aligned}$$

It follows that $|f(x) - f(x_0)| < \epsilon$ whenever $|x - x_0| < \epsilon/K$. This choice, $\delta = \epsilon/K$, depends only on ϵ, but not on x_0.

A function f with bounded derivative is therefore uniformly continuous. Examples: $f(x) = x$, $f(x) = \sin x$, and so forth.

C. A function on $\mathbf{R}$ can be uniformly continuous even if its derivative is not bounded (and even if it has no derivative at all). Consider [10] $f(x) = \sqrt{x}$ on $[0, \infty)$. Let us suppose that $|x - x_0| < \delta$, and see how big $|\sqrt{x} - \sqrt{x_0}|$ can get. By reversing the argument, we then find δ for given ϵ.

Suppose first that $0 \leq x_0 < \delta$, so that $0 \leq x < 2\delta$. Then

$$|\sqrt{x} - \sqrt{x_0}| \leq \sqrt{x} + \sqrt{x_0} < \sqrt{\delta} + \sqrt{2\delta} < 3\sqrt{\delta}.$$

If $\delta \leq x_0$, then

$$|\sqrt{x} - \sqrt{x_0}| = \frac{|x - x_0|}{\sqrt{x} + \sqrt{x_0}} \leq \frac{|x - x_0|}{\sqrt{x_0}} < \frac{\delta}{\sqrt{\delta}} = \sqrt{\delta} < 3\sqrt{\delta}.$$

We have shown that $|x - x_0| < \delta$ implies $|\sqrt{x} - \sqrt{x_0}| < 3\sqrt{\delta}$. Given ϵ, pick δ so that $3\sqrt{\delta} = \epsilon$ (so $\delta = \epsilon^2/9$): then $|x - x_0| < \delta$ forces $|\sqrt{x} - \sqrt{x_0}| < \epsilon$. This δ depends on ϵ but not on x_0, and $\sqrt{x}$ is uniformly continuous on $[0, \infty)$. ∎

• **1.4.3 Exercise.** Prove or disprove: the function $f(x) = \sin(1/x)$ is uniformly continuous on $(0, 1]$. ∎

---

[10] This example comes from R. C. Buck's text *Advanced Calculus*, which contains much interesting material relevant to this subsection.

1.4.4 Exercise. Prove or disprove: the function

$$f(x) = \begin{cases} -1/\ln x, & \text{for } 0 < x \le 1/2, \\ 0, & \text{for } x = 0, \end{cases}$$

is uniformly continuous on $[0, 1/2]$. ∎

1.4.5 Proposition. *A continuous function from a compact metric space (M, d) to a metric space (M', d') is uniformly continuous.* ∎

Proof. By contradiction. What is the negation of "f is uniformly continuous"? There is an $\epsilon > 0$, such that for any $\delta = 1/n$, we can find points $x_n, y_n \in M$ within $1/n$ of each other, $d(x_n, y_n) < 1/n$, but with $d'(f(x_n), f(y_n)) \ge \epsilon$. By compactness of M, the sequence $\{x_n\}$ has a convergent subsequence $\{x_{n_k}\}$. Let x be its limit. Because $d(x_n, y_n) < 1/n$, the corresponding subsequence of $\{y_n\}$ also has limit x. But now

$$\epsilon \le d'(f(x_{n_k}), f(y_{n_k})) \le d'(f(x_{n_k}), f(x)) + d'(f(x), f(y_{n_k})).$$

So, either

$$d'(f(x_{n_k}), f(x)) \ge \epsilon/2, \text{ or } d'(f(x), f(y_{n_k})) \ge \epsilon/2, \text{ or both}$$

for infinitely many indices n_k. This contradicts continuity of f:

$$\text{if } x_{n_k} \longrightarrow x, \text{ then } d'(f(x_{n_k}), f(x)) \longrightarrow 0,$$

and likewise for the y_{n_k}'s,

$$\text{if } y_{n_k} \longrightarrow x, \text{ then } d'(f(y_{n_k}), f(x)) \longrightarrow 0. \text{ ∎}$$

1.4.6 Example. Uniform continuity is used as a matter of course in the theory of Riemann integrals of continuous functions. The general situation is treated in Section 2; a special example will suffice for now. Let f be a continuous function on $[0, 1]$. We show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(k/n) = \int_0^1 f(x)\,dx. \tag{1.4.7}$$

The left side is the kind of approximating sum one meets in the first calculus course; such limits also appear when one discretizes a continuous problem.

Write the right side of (1.4.7) as a sum,

$$\sum_{k=1}^{n} \int_{\frac{k-1}{n}}^{\frac{k}{n}} f(x)\,dx.$$

We should prove that

$$\lim_{n\to\infty} \left| \sum_{k=1}^{n} \left(\int_{\frac{k-1}{n}}^{\frac{k}{n}} (f(x) - f(k/n))\, dx \right) \right| = 0. \tag{1.4.8}$$

Let $\epsilon > 0$ be given. By Proposition 1.4.5, f is uniformly continuous, so there is a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. Choose $1/n < \delta$; then $|f(x) - f(k/n)| < \epsilon$ when $|x - k/n| < 1/n$, for all $k = 1, \ldots, n$. Thus the sum in (1.4.8) is dominated by

$$\sum_{k=1}^{n} \int_{\frac{k-1}{n}}^{\frac{k}{n}} |f(x) - f(k/n)|\, dx \;\leq\; \sum_{k=1}^{n} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \epsilon\, dx \;=\; \epsilon.$$

The limit (1.4.8) is established. ∎

• **1.4.9 Exercise.** Let $f : (M, d) \longrightarrow (M', d')$ be uniformly continuous. Show that if $\{x_n\}$ is a Cauchy sequence in (M, d), then $\{f(x_n)\}$ is a Cauchy sequence in (M', d').

Note: the function in (1.4.2A) is not uniformly continuous. While the sequence $\{1/n\} \subset (0, 1]$ is Cauchy, the sequence $\{f(1/n)\} = \{n\} \subset (0, \infty)$ is not. ∎

• **1.4.10 Exercise.** Suppose g is uniformly continuous on $\mathbf{R}$, and f is a continuous function for which $|f(x) - g(x)| < .1$ for all $x \in \mathbf{R}$. Is f necessarily uniformly continuous? ∎

• **1.4.11 Exercise.** Let f be continuous and bounded on $\mathbf{R}$. Is f necessarily uniformly continuous? ∎

• **1.4.12 Exercise.** Suppose f is continuous on $\mathbf{R}$, and for every integer $n > 0$, there is a function g_n, uniformly continuous on $\mathbf{R}$, such that $|f(x) - g_n(x)| < 1/n$ for every $x \in \mathbf{R}$. Is f necessarily uniformly continuous? ∎

1.4.13 Exercise. Let f be continuous on $[0, 1]$. The *modulus of continuity* of f is defined by

$$\omega(f; \delta) = \sup_{|x-y| \leq \delta} |f(x) - f(y)|.$$

Prove that $\lim_{\delta \to 0} \omega(f; \delta) = 0$. ∎

1.4.14 Exercise. Let (M, d) and (M', d') be metric spaces, and suppose that $f : M \to M'$ is uniformly continuous. Prove: if $d(x_n, y_n) \to 0$, then $d'(f(x_n), f(y_n)) \to 0$. Is this true if f is not uniformly continuous? ∎

The next uniform property to be described is *uniform convergence* of a sequence of functions. We have already met it in the guise of L^∞ convergence, but it is worthwhile to spell out the significance of this property in a different way.

1.4.15 Definition. *Let (M, d) and (M', d') be metric spaces, let $\{f_n\}$ be a sequence of functions from M to M', and let $f : M \to M'$. If*

$$\lim_{n\to\infty} f_n(x) = f(x) \quad \text{for all } x \in M,$$

*one says that the f_n **approach** f **pointwise**.* ∎

1.4.16 Definition. *Let (M, d) and (M', d') be metric spaces. Let $\{f_n\}$ be a sequence of functions from M to M', and let $f : M \to M'$. One says that the f_n **converge uniformly** to f if for every $\epsilon > 0$, there is an N depending on ϵ but not on x, such that*

$$n > N \Longrightarrow |f_n(x) - f(x)| < \epsilon \quad \text{for all } x \in M.$$ ∎

Remark. The "uniform" in "uniform convergence" refers to the fact that the large integer N depends on ϵ but not on x: it is *uniform* in x. Of course, if $f_n \to f$ uniformly, then necessarily $f_n \to f$ pointwise. The converse is not true, as the examples below demonstrate. ∎

Uniform convergence should be pictured as in Figure 5. If $f_n \to f$ uniformly, then given a strip of arbitrary width ϵ about the graph of f, *all* f_n from some n on must lie inside that strip.

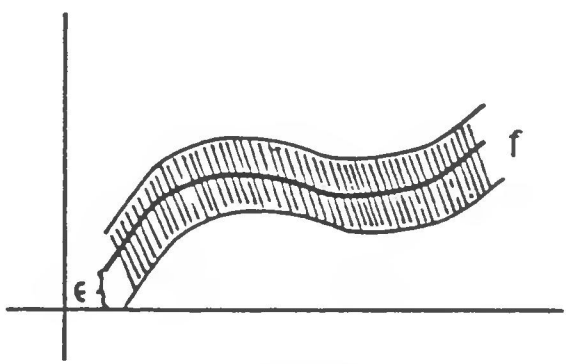


Figure 5. *Uniform convergence to f*

A particularly ubiquitous special case of pointwise and uniform convergence of functions is *convergence of series of functions*.

1.4.17 Definition. *Let (M, d) be a metric space, and let $\{f_n\}$ be a sequence of real- or complex- valued functions. Set*

$$s_n(x) = \sum_{j=1}^{n} f_j(x).$$

*We say that $\sum_{j=1}^{\infty} f_j$ **converges pointwise (resp. uniformly)** to f if the partial sums s_n converge pointwise (resp. uniformly) to f.* ∎

1.4.18 Examples

A. The functions $f_n(x) = x^n$ converge to the zero function pointwise on $(0, 1)$, but they do not converge uniformly. If they did, then for all sufficiently large n we would have $|x^n| < .1$, but as in Example 1.1.5C, for large n and $x_0 = 1 - 1/n$, the value of $x_0^n = (1 - 1/n)^n$ is very close to e^{-1}.

B. The interval $(0, 1)$ in the preceding example is not compact, but that is not the reason for the failure of uniform convergence. Let $f_n(x) = n(1 - x)x^n$; then the f_n converge pointwise to zero on $[0, 1]$, but since $f_n(1 - 1/n) = (1 - 1/n)^n$ as before, the functions f_n do not become uniformly small. See Figure 6.

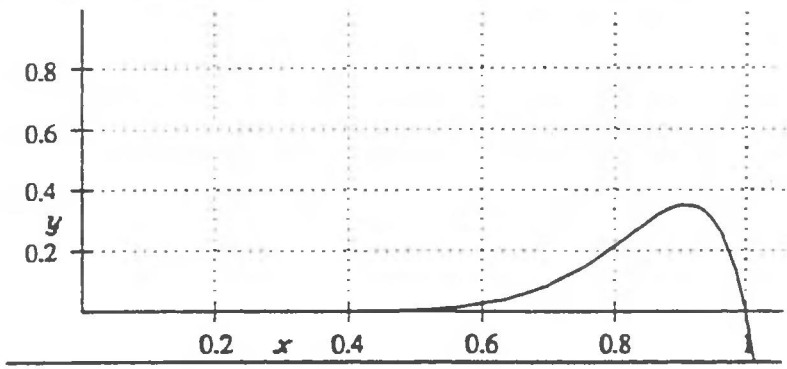


Figure 6. $f_n(x) = n(1 - x)x^n$, *for* $n = 10$

Change the f_n to $f_n(x) = n^2(1 - x)x^n$. They still converge pointwise to zero on $[0, 1]$, i.e., $f_n(x) \to 0$ for each *fixed* $x \in [0, 1]$, but the values of f_n at the *n-dependent* point $1 - 1/n$ tend to infinity!

C. Let f_n be a sequence of functions on (M, d), and suppose that

$$\sup_{x \in M} |f_n(x)| \le a_n, \quad \text{with} \quad \sum_1^\infty a_n < \infty.$$

Then the series $\sum_1^\infty f_n(x)$ converges for each x (i.e., pointwise) by the comparison test. Call the sum $f(x)$. Convergence to f is also uniform:

$$\left| f(x) - \sum_{n=1}^{m} f_n(x) \right| \le \sum_{n=m+1}^{\infty} |f_n(x)| \le \sum_{n=m+1}^{\infty} a_n < \epsilon$$

if m is large enough.

This is the *Weierstrass test for uniform convergence of a series of functions.* It is also referred to as the "Weierstrass M-test", because the constants I called a_n are often called M_n (I did not use that letter, since M is also my metric space). Maybe the "*M*" stands for "majorization"?

D. Let

$$f_n(x) = -2n^2 x e^{-n^2 x^2} + 2(n+1)^2 x e^{-(n+1)^2 x^2}.$$

The partial sums telescope, and one has

$$s_n(x) \stackrel{\text{def}}{=} \sum_{j=1}^{n} f_j(x) = 2(n+1)^2 x e^{-(n+1)^2 x^2} - 2x e^{-x^2}.$$

The s_n converge pointwise to $-2x\exp(-x^2)$ on $[0,\infty)$, but the convergence is not uniform. The function $2(n+1)^2 x e^{-(n+1)^2 x^2}$ has the maximum value

$$(n+1)\sqrt{\frac{2}{e}} \quad \text{at } x = \frac{1}{2(n+1)^2},$$

so that the maximum differences $|-2x\exp(-x^2) - s_n(x)|$, instead of becoming smaller as $n \to \infty$, become arbitrarily large. ∎

Uniform convergence of sequences of functions is important because the limit function inherits various desirable properties of the functions in the sequence. The first result along these lines was already proved in Proposition 1.1.23; we restate it in the language of uniformity.

1.4.19 Definition. *Let (M,d) be a metric space. For a bounded continuous function $f : M \to \mathbf{R}$, set*

$$\|f\| \stackrel{\text{def}}{=} \sup_{x \in M} |f(x)|.$$

The linear space of bounded continuous real-valued functions on M, equipped with this norm, is denoted by $C(M)$. ∎

Remark. This norm is just the familiar $\|\cdot\|_\infty$ norm. In standard usage, that norm has measure-theoretic implications, which will be explained in the next section. Therefore, I now begin to conform to convention, and use a different symbol for the (so-called) sup-norm on continuous functions. Some people write $\|f\|_0$ for this norm. ∎

1.4.20 Proposition. *Let (M,d) be a metric space. Then:*

i) $C(M)$ is complete;

ii) if $\{f_n\} \subset C(M)$ and $f_n \to f$ uniformly, then $f \in C(M)$; in other words, the uniform limit of continuous functions is continuous. ∎

Proof. The proof for completeness of $C(M)$ is exactly the same as the proof for $C([0,1])$ in Proposition 1.1.23.

Part ii) is a consequence of part i). If $f_n \to f$ uniformly, then, given $\epsilon > 0$, there is an N_ϵ so that $m_1, m_2 > N_\epsilon$ implies

$$|f_{m_1}(x) - f_{m_2}(x)| < \epsilon \quad \text{for all } x \in M.$$

Thus $\{f_n\}$ is a Cauchy sequence in the norm $\|\cdot\|$, and by i) it has a limit in $C(M)$. This limit must be f, and so f is continuous. ∎

1.4.21 Corollary. *The sum of a uniformly convergent series of continuous functions is continuous.* ∎

1.4.22 Examples

A. Consider the complex power series $\sum_{n=0}^{\infty} c_n z^n$. Suppose it converges absolutely when $|z| = r$. Then for $|z| < r$, we have $|c_n z^n| < |c_n| r^n$, and by the Weierstrass test (1.4.18C), with $a_n = |c_n| r^n$, it converges uniformly on the open disk $|z| < r$. The sum is therefore a continuous function. (Much more is true, as one learns in complex analysis.)

B. The function defined by

$$\zeta(s) \stackrel{\text{def}}{=} \sum_{n=1}^{\infty} \frac{1}{n^s}$$

is the *Riemann zeta function*. When the complex number s satisfies $\Re(s) \geq 1+\delta$, the Weierstrass test (1.4.18C), with $a_n = n^{-1-\delta}$, shows that the series converges uniformly. Hence, $\zeta(s)$ is continuous in $\Re(s) > 1$. (Actually, it is an analytic function there, but other techniques are needed to deduce this.)

C. The notion of uniform continuity makes it possible to give a much more elegant proof of a strenghtening of Corollary I.4.0.4. Recall that we had set

$$E(t,x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t},$$

and showed that

$$\lim_{t \downarrow 0} \int_{-\infty}^{\infty} E(t,y) f(x-y)\, dy = f(x) \tag{1.4.23}$$

for continuous, bounded functions f. In distribution language,

$$\lim_{t \downarrow 0} E(t,x) = \delta(x).$$

For ease of notation, write $E(t,x) = E_t(x)$, and set

$$E_t \star f(x) = \int_{-\infty}^{\infty} E_t(y) f(y-x)\, dy = \int_{-\infty}^{\infty} E_t(y-x) f(y)\, dy.$$

The new function $E_t \star f$ is called the *convolution* of E_t and f. I will prove:

Proposition. *If f is continuous and bounded on $\mathbb{R}$, then $E_t \star f$ approaches f as $t \downarrow 0$uniformly on every interval (which may be $\mathbb{R}$ itself) on which f is uniformly continuous.* ∎

Proof. Suppose f is uniformly continuous on the interval I. It is to be shown that, given $\epsilon > 0$, there is a t_0 depending on ϵ *but not on* x, such that

$$|E_t \star f(x) - f(x)| < \epsilon \quad \text{for } 0 < t < t_0,\ x \in I. \tag{1.4.24}$$

As in the proof of Proposition I.4.0.3, we use the fact that

$$\int_{-\infty}^{\infty} E_t(y)\,dy = 1$$

for all $t > 0$, so that the difference in (1.4.24) can be written

$$E_t \star f(x) - f(x) = \int_{-\infty}^{\infty} (f(x-y) - f(x))E_t(y)\,dy.$$

After the substitution $y = 2\sqrt{t}z$, the last integral becomes

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (f(x - 2\sqrt{t}z) - f(x))e^{-z^2}\,dz. \tag{1.4.25}$$

We estimate this integral. The general strategy is as in the proof of (I.4.0.3).

Choose N so large that

$$\frac{1}{\sqrt{\pi}} \int_{|z|>N} e^{-z^2}\,dz < \frac{\epsilon}{4\|f\|_\infty}. \tag{1.4.26}$$

On the interval I, f is uniformly continuous. There is a δ such that $|f(x+h) - f(x)| < \epsilon/2$ whenever $|h| < \delta$, for all $x \in I$. When $|z| \le N$, $|2\sqrt{t}z| \le 2\sqrt{t}N$, and so for $2\sqrt{t}N < \delta$ we have

$$\frac{1}{\sqrt{\pi}} \int_{-N}^{N} |f(x - 2\sqrt{t}z) - f(x)|e^{-z^2}\,dz < \frac{1}{\sqrt{\pi}} \int_{-N}^{N} \frac{\epsilon}{2} \cdot e^{-z^2}\,dz < \frac{\epsilon}{2}. \tag{1.4.27}$$

Now use (1.4.26) in

$$\begin{aligned} \frac{1}{\sqrt{\pi}} \int_{|z|>N} |f(x - 2\sqrt{t}z) - f(x)|e^{-z^2}\,dz & \\ \le \frac{1}{\sqrt{\pi}} \int_{|z|>N} (\sup_{z\in\mathbf{R}} |f(x - 2\sqrt{t}z) - f(x)|)\, e^{-z^2}\,dz & \\ \le \frac{1}{\sqrt{\pi}} \int_{|z|>N} 2\|f\|_\infty e^{-z^2}\,dz < \frac{\epsilon}{2}; & \end{aligned}$$

This estimate and (1.4.27) show that the difference (1.4.25) is less than ϵ, independently of $x \in I$, as long as $2\sqrt{t}N < \delta$, or $0 < t < \delta^2/4N^2$. ∎

The result is particularly significant for its application to the heat equation, $u_t = u_{xx}$. It was pointed out in Subsection I.4.0 that $u(t,x) = E_t \star f(x)$ solves this partial differential equation, and that $u(0,x)$, the initial "temperature distribution", is the function f. It now follows from the proposition just proved that the convergence of $u(t,x)$ to $u(0,x)$ is uniform for x in bounded intervals. Physically: for small time t, the temperature in the rod is near the initial temperature in every bounded spatial region; there are no sudden hot or cold spots. ∎

Three further implications of uniform convergence are developed in the final three examples. This is only a brief summary of a rather large chapter of advanced calculus.

1.4.28 Example. *Term-by-term differentiation of series.*

A. The series

$$\sum_{n=0}^{\infty} \frac{\sin 2^n x}{2^n}$$

converges uniformly by the Weierstrass test, (1.4.18C). Let $f(x)$ be the sum; it is a continuous function. One hopes that the derivative of f can be obtained in the obvious way:

$$f'(x) = \frac{d}{dx} \sum_{n=0}^{\infty} \frac{\sin 2^n x}{2^n} = \sum_{n=0}^{\infty} \cos 2^n x.$$

Unfortunately, the differentiated series does not converge. It is difficult to tell where the problem lies: does f not have a derivative at any x? or is there just no derivative at some particular x? or is f differentiable, and the Fourier cosine series is for some reason not appropriate? See part D for a related example, where this question is answered.

B. Example A can be phrased in the following general terms. The partial sums $s_n(x)$ converge uniformly (for $x \in \mathbf{R}$) to $f(x)$. The derivatives $s_n'(x)$, however, do not converge, uniformly or even pointwise. In the language of Example 1.1.27: the sequence $\{s_n\}$ is not Cauchy in $C^1(\mathbf{R})$. That example also suggests hypotheses that will guarantee the validity of term-by-term differentiation of an infinite series.

1.4.29 Proposition. *Suppose the series $\sum_{n=1}^{\infty} f_n(x)$ converges uniformly to $f(x)$ for x in an interval $J \subset \mathbf{R}$, and suppose the differentiated series $\sum_{n=1}^{\infty} f_n'(x)$ converges uniformly to $g(x)$. Then $f'(x) = g(x)$, $(x \in J)$.*

Proof. See p. II - 410. ∎

C. Continue Example 1.4.22A: power series. Assume that $\sum_{n=0}^{\infty} c_n x^n$ converges (uniformly) on the real interval $-r \le x \le r$. Then, for every $r_1 < r$, the differentiated series $\sum_{n=0}^{\infty} n c_n x^{n-1}$ converges uniformly on $|x| < r_1$, by comparison with $\sum_{n=0}^{\infty} |c_n| r^n$. It follows from part B that term-by-term differentiation of the power series is permitted on an interval of uniform convergence.

The conclusion remains valid when x is replaced by the complex variable z, and the derivative d/dx by the complex derivative d/dz. This is done in every complex analysis text. ∎

D. The series

$$\sum_{n=0}^{\infty} \frac{\cos(21^n \pi x)}{3^n}$$

defines a continuous function $f(x)$ on R (comparison test with $\sum 3^{-n}$). It is a very wiggly function: each term in the series superimposes very small, but very rapid, oscillations on the preceding partial sum. The strange choice of constants allows one to prove rigorously that f *is not differentiable at any* x. This example is due to Weierstrass.

Fix $x \in \mathbf{R}$. We will show that the difference quotient

$$\frac{f(x+h) - f(x)}{h}$$

oscillates between $\pm\infty$ as $|h| \to 0$. For $m > 0$, write $f(x) = s_m(x) + r_m(x)$, where s_m is the partial sum $\sum_0^{m-1}$ and r_m the remainder. Define

$$\Delta_h s_m(x) = \frac{s_m(x+h) - s_m(x)}{h}, \quad \Delta_h r_m(x) = \frac{r_m(x+h) - r_m(x)}{h}.$$

By the mean value theorem,

$$|\cos(21^n \pi(x+h)) - \cos(21^n \pi x)| \le 21^n \pi |\sin(21^n \pi \xi)| \times |(x+h) - x| \le 21^n \pi |h|.$$

Hence

$$|\Delta_h s_m(x)| \le \pi \sum_{n=0}^{m-1} 7^n = \pi \frac{7^m - 1}{6} < \frac{7^m \pi}{6}. \tag{1.4.30}$$

Next, for given x and m, there is an integer p such that $|21^m x - p| \le 1/2$; let $q = 21^m x - p$ be the fractional part of $21^m x$. Let h_m be one of the two numbers

$$h_m = \frac{\pm 1 - q}{21^m}:$$

one is positive, the other negative, and

$$|h_m| < \frac{1 + (1/2)}{21^m} = \frac{3}{2 \cdot 21^m}.$$

As $m \to \infty$, $h_m \to 0$. We will estimate the difference quotient of f for this h_m.

Since 21 is odd,

$$\cos[21^n \pi(x + h_m)] = \cos[21^{n-m} \pi(p \pm 1)] = (-1)^{p \pm 1}$$

when $n \geq m$. On the other hand,

$$\begin{aligned} \cos(21^n \pi x) &= \cos[21^{n-m} \pi(p + q)] \\ &= \cos(21^{n-m} \pi p) \cos(21^{n-m} \pi q) \\ &= (-1)^p \cos(21^{n-m} \pi q). \end{aligned}$$

Therefore

$$\cos[21^n \pi(x + h_m)] - \cos(21^n \pi x) = (-1)^{p+1} \left(1 + \cos(21^{n-m} \pi q)\right) \text{ *)}$$

and so

$$\begin{aligned} |\Delta_{h_m} r_m(x)| &= \frac{|(-1)^{p+1}|}{|h_m|} \sum_{n=m}^{\infty} \frac{1 + \cos(21^{n-m} \pi q)}{3^n} \\ &\geq \frac{1}{3^m |h_m|} > \frac{2 \cdot 7^m}{3}. \end{aligned} \tag{1.4.31}$$

We now use (1.4.30) and (1.4.31) to estimate the difference quotient:

$$\begin{aligned} \left| \frac{f(x + h_m) - f(x)}{h_m} \right| &= |\Delta_{h_m} s_m(x) + \Delta_{h_m} r_m(x)| \\ &\geq |\Delta_{h_m} r_m(x)| - |\Delta_{h_m} s_m(x)| \\ &\geq 7^m \left[\frac{2}{3} - \frac{\pi}{6} \right] \\ &> (.1) \times 7^m. \end{aligned} \tag{1.4.32}$$

The last term in (1.4.32) tends to infinity with m, meaning that the absolute value of the difference quotient tends to infinity. The difference quotient itself can go to $+\infty$ or $-\infty$, depending on the signs of p and h_m. One sees in (1.4.31) what happens:

- if p is odd and $h_m > 0$, the limit is $+\infty$, while if $h_m < 0$, the limit is $-\infty$;
- if p is even and $h_m > 0$, the limit is $-\infty$, while if $h_m < 0$, the limit is $+\infty$.

Hence,

$$\lim_{h \to 0} \frac{f(x + h) - f(x)}{h}$$

*) - see the comment on p. II - 410

cannot exist. We have even excluded the possibility that the graph at x is vertical, because that would require the difference quotient to tend either to $+\infty$ or to $-\infty$ on *both* sides.

Note that the seemingly unmotivated choices of 21 and 3 in the definition of $f(x)$ combine to produce the denominator in $\pi/6$ in the next-to-last inequality in (1.4.32), and that makes the difference $(2/3) - (\pi/6)$ positive, which in turn allows one to conclude that the difference quotient becomes infinite along the sequence h_m. Had we started, for example, with $2^{-n}\cos(3^n\pi x)$, the result at this step would have been $(3/2)^m[(2/3) - 2\pi]$, which tends to $-\infty$ and says nothing unfavorable about the difference quotient. See Exercise 1.4.40 for a more general statement.

E. We are now in a position to present the first really nontrivial illustration of the notions of "small" and "large" sets introduced in Subsection I.5.1. The example puts Weierstrass' nondifferentiable function in a wider context. It makes use of the completeness of $\mathcal{C}([0,1])$, the Baire Category Theorem, and basic facts about uniform continuity and uniform convergence.

Proposition. The property of not having a derivative at any point is generic for functions in $\mathcal{C}([0,1])$. ∎

Proof. Recall (Definition I.5.1.27) that *generic* means: the property holds for a dense residual set, i.e., a countable intersection of open dense sets which is still dense. We will in fact show more than the proposition advertises: the generic continuous function does not even have a bounded right difference quotient. The *right difference quotient* at a point x is the ratio

$$\frac{f(x+h)-f(x)}{h}, \text{ for } x < x+h < 1.$$

When $h \geq$some $h_0 > 0$, the function $f(x+h) - f(x)$ is continuous in h and therefore bounded (by a constant we choose to write in the form Ch_0) on the interval $[h_0, 1-x]$. If $f'(x)$ exists, so that $\lim_{h\downarrow 0}(f(x+h)-f(x))/h = f'(x)$, then the right difference quotient also satisfies $|f(x+h)-f(x)| \leq Ch$ when $0 < h < h_0$. We introduce sets of functions for which such an estimate holds somewhere. For $n \geq 2$, let F_n be the subset of $\mathcal{C}([0,1])$ consisting of functions for which

$$(*)\quad |f(x+h)-f(x)| \leq nh, \text{for some } x \in [0, 1-1/n] \text{ and all } h, x < x+h < 1.$$

We will show that F_n is nowhere dense (its closure has empty interior, Definition I.5.1.1). We prove first that F_n is already closed, and then that it has empty interior. Because $\mathcal{C}([0,1])$ is complete (Proposition 1.4.20), the Baire Category Theorem 1.1.36 then asserts that $\bigcap F_n^c$, which is residual, must be dense. This intersection contains all functions that do not have a bounded right difference quotient for any $x \in [0,1)$. The proposition will then have been established.

Let $f_k \in F_n$ and $\|f_k - f\| \to 0$; we prove that $f \in F_n$, which shows F_n to be closed. For each k, there is an $x_k \in [0, 1 - 1/n]$ for which (*) holds,

$$(**) \qquad |f_k(x_k + h) - f_k(x_k)| \le nh. \quad \text{for some } x_k \in [0.1 - 1/n] \text{ and all } h, x_k < x_k + h < 1.$$

By the Bolzano-Weierstrass Theorem, a subsequence of $\{x_k\}$ converges to an $x \in [0, 1 - 1/n]$; for simplicity of notation, we may assume that $\{x_k\}$ itself converges[11]. Fix an h with $x < x + h < 1$. A *five-fold* (!) application of the triangle inequality gives

$$\begin{aligned} |f(x+h) - f(x)| \le\ & |f(x+h) - f(x_k+h)| + |f(x_k+h) - f_k(x_k+h)| \\ & + |f_k(x_k+h) - f_k(x_k)| + |f_k(x_k) - f(x_k)| \\ & + |f(x_k) - f(x)|. \end{aligned}$$

Four of the terms on the right can be made arbitrarily small by taking k sufficiently large: the first and last terms because f is continuous (as follows from the fact that a uniform limit of continuous functions is continuous), and the second and fourth terms because $f_k \to f$ uniformly. As for the third term, we have $x_k < x_k + h < 1$ for large enough k since $x_k \to x$, and (**) applies. Thus, given $\epsilon > 0$,

$$|f(x+h) - f(x)| \le nh + \text{ a term smaller than } \epsilon$$

holds for sufficiently large k. That establishes condition (*) for f.

Now we show that the interior of F_n is empty. Let $f \in F_n$, and prescribe an $\epsilon > 0$. The ϵ-ball about f should not be contained in F_n, i.e. there should be a $\chi \in C([0,1])$, $\chi \notin F_n$, for which $\|f - \chi\| < \epsilon$. To get such a χ, we first approximate f to within a small number η by a polygonal (=continuous piecewise linear) function.

Let $x_1 < x_2$, and set $f(x_i) = y_i, i = 1, 2$. The line segment connecting (x_1, y_1) with (x_2, y_2) has equation

$$y = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1.$$

For $x \in (x_1, x_2)$, the distance between f and this line is estimated by

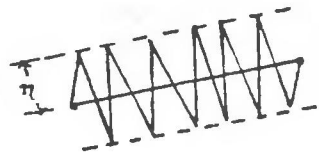
$$|f(x) - y| = \left|(f(x) - y_1) + (y_2 - y_1)\frac{x - x_1}{x_2 - x_1}\right| \le |f(x) - f(x_1)| + |f(x_2) - f(x_1)|.$$

---

[11] If x_k is not restricted to the interval $[0, 1 - 1/n]$ in the definition of F_n, the limit x might well be $= 1$. Since the right difference quotient is not defined at $x = 1$, it would not be possible to argue that $f \in F_n$.

Because f is uniformly continuous (Proposition 1.4.5), the right side will be $< \eta$ if x_1 and x_2 are sufficiently close. By subdividing $[0,1]$ into sufficiently small intervals, and connecting the corresponding points on the graph of f by line segments, we obtain the desired approximation. Call it ψ.

Finally, each line segment making up ψ is in turn approximated to within η by a piecewise linear curve each of whose segments has slope $\geq n$. A picture makes it clear:



(Try to write down the proof, if you don't like pictures.) The right difference quotients of the resulting function (call it χ) must go above the value n because the zigzags were made so steep. Hence χ does not belong to F_n, but the approximation to f can be as close as desired. This shows that no point of F_n is an interior point, and the proposition is proved. ∎

1.4.33 Example. *Interchange of limit and integral*

A. Let $f_n(x) = n^2(1-x)x^n$ on $[0,1]$. We saw in Example 1.4.18B[*)] that $f_n \to 0$ pointwise, but not uniformly, on $[0,1]$. Is it true that

$$\int_0^1 f_n(x)\,dx \to \int_0^1 0\,dx = 0?$$

Evidently not:

$$\int_0^1 f_n(x)\,dx = \frac{n^2}{(n+1)(n+2)} \to 1.$$

This is *the problem of interchange of limit and integral*: if $f_n \to f$, does $\int f_n \to \int f$? If one thinks of integration,

$$g \mapsto \int g \overset{\text{def}}{=} I(g),$$

as a function I whose argument is the integrand g, the question becomes: is I continuous with respect to the chosen form of convergence? I.e., does $I(\lim_n f_n) = \lim_n I(f_n)$?

The example shows that the integral is not continuous with respect to pointwise convergence.

B. Applied to the sequence of partial sums of a series of functions, the problem of interchange of limit and integral becomes *the problem of term-by-term integration of a series*:

$$\int \lim_n s_n = \lim_n \int s_n \iff \int \sum_1^\infty f_n = \sum_1^\infty \int f_n\,?$$

*) - see p. II - 410

As in Example 1.4.18D, set

$$f_n(x) = -2n^2xe^{-n^2x^2} + 2(n+1)^2xe^{-(n+1)^2x^2}.$$

Recall that

$$\sum_{n=1}^{\infty} f_n(x) = -2xe^{-x^2}, \quad \text{for } x \in [0,\infty),$$

but that the convergence is not uniform. A short calculation shows that

$$\int_0^x \sum_{n=1}^{\infty} f_n(s)\,ds = e^{-x^2} - 1,$$

while

$$\sum_{n=1}^{\infty} \int_0^x f_n(s)\,ds = e^{-x^2}.$$

Term-by-term integration of this series of functions does not yield the integral of the sum.

C. Uniform convergence is a sufficient condition for the validity of interchange of limit and integral. It is a very strong restriction, and the Lebesgue theory of integration (next section) gives the same conclusion under far weaker (and more easily applicable) hypotheses.

1.4.34 Proposition. *Suppose that $\{f_n\}$ is a sequence of continuous functions on a bounded interval $[a,b] \subseteq \mathbb{R}$, and that $f_n \to f$ uniformly on $[a,b]$. Then*

$$\lim_{n\to\infty} \int_a^b f_n(x)\,dx = \int_a^b f(x)\,dx.$$

Remark. The casual statement of this proposition obscures what may be the most important fact: the limit function f can be integrated. I have not yet discussed the foundations of integration theory, and you may not realize that the integral you met in calculus ("area under the graph of f") could be undefined for mildly exotic functions f. This matter will be taken up in Section 2. In a nutshell: the pointwise limit f of a perfectly nice sequence $\{f_n\}$ need not have a well-defined, calculus-type area under its graph, and $\lim \int f_n = \int \lim f_n$ may fail for that silly reason alone. For now, just assume (as perhaps you have, for years) that $\int f(x)\,dx$ is meaningful when f is "nice", e.g. continuous. Example 1.4.6 above suggests the rigorous details, and Proposition 2.2.3 below does the rest. ∎

Proof. Uniform convergence means that $|f_n(x) - f(x)| < \epsilon/(b-a)$ for all $x \in [a.\, b]$, if n is greater than some N_ϵ. Thus,

$$\left| \int_a^b f_n(x)\,dx - \int_a^b f(x)\,dx \right| \leq \int_a^b |f_n(x) - f(x)|\,dx < \int_a^b \frac{\epsilon}{b-a}\,dx = \epsilon$$

for $n > N_\epsilon$. ∎

D. Let

$$f_n(x) = \begin{cases} 1/n, & \text{if } 0 \leq x \leq n, \\ 0, & \text{if } x > n. \end{cases}$$

Then $\int_0^\infty f_n(x)\,dx = 1$ for all n (the integral being interpreted in the natural "'area under the graph" way), but since $f_n(x) \to f(x) \stackrel{\text{def}}{=} 0$ (uniformly!) for all $x \geq 0$, we obviously don't have

$$\lim_{n\to\infty} \int_0^\infty f_n(x)\,dx = \int_0^\infty f(x)\,dx.$$

It was therefore essential, in part C, that the interval $[a, b]$ be bounded. ∎

E. Continue Example 1.4.22A: power series. If

$$\sum_{n=0}^{\infty} a_n x^n = f(x)$$

uniformly for $|x| < r$, then by Proposition 1.4.34,

$$\int_0^x \sum_{n=0}^{\infty} a_n t^n\,dt = \int_0^x f(t)\,dt.$$

For example,

$$1 + x + x^2 + x^3 + \ldots = \frac{1}{1-x} \quad \text{for } |x| < 1,$$

and therefore

$$x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots = \ln(1-x),$$

again for $|x| < 1$. ∎

1.4.35 Example. *Improper integrals depending on a parameter*

Proposition 1.4.34 in the preceding example would apply to

$$\lim_{n\to\infty} \int_0^1 e^{-nx^2}\,dx.$$

Let us now consider a more typical case — the integral is improper, extending over an infinite interval, and the parameter n is replaced by a continuous parameter t.

1.4.36 Definition. *The integral*

$$\int_{-\infty}^{\infty} f(t,x)\,dx$$

is said to ***converge uniformly to*** $F(t)$ *for* $t \in [a,b] \subseteq \mathbf{R}$, *if for every* $\epsilon > 0$ *there is an* R_0 *depending on* ϵ *but not on* t, *such that*

$$\left| \int_{-R}^{R} f(t,x)\,dx - F(t) \right| < \epsilon$$

for all $t \in [a,b]$ *and* $R > R_0$.

Remark. The integration interval $(-\infty,\infty)$ can be replaced by $[c,\infty)$ or $(-\infty,c]$. Or, the functions $f(t,\cdot)$ could become infinite on a finite interval (α,β) (e.g., $f(t,x) = t/\sqrt{x}$ for $x \in (0,1)$), as long as the improper integral exists. The proof of the proposition below would be esssentially the same. ∎

1.4.37 Proposition. *Suppose that* $f(t,x)$ *and* $\partial f(t,x)/\partial t$ *are continuous, that* $\int_{-\infty}^{\infty} f(t,x)\,dx$ *converges to* $F(t)$, *and that* $\int_{-\infty}^{\infty} \partial f(t,x)/\partial t\,dx$ *converges uniformly for* $t \in [a,b]$. *Then* F *is differentiable, and*

$$F'(t) = \int_{-\infty}^{\infty} \frac{\partial f(t,x)}{\partial t}\,dx$$

for all $t \in [a,b]$.

Proof. See Exercise 1.4.38, with hint. ∎

• **1.4.38 Exercise.** Prove Proposition 1.4.37. Hint: set

$$G(t) = \int_{-\infty}^{\infty} \frac{\partial f(t,x)}{\partial t}\,dx,$$

and (by changing the order of integration in the appropriate double integral) show that

$$\int_{-\infty}^{X} G(t)\,dt = F(X).$$

If you think of $\int_{-R}^{R}$ as analog of $\sum_{j=1}^{n}$, and of the parameter t as analog of x in $\sum f_n(x)$, the strategy of the proof can be guessed from Proposition 1.4.29. ∎

• **1.4.39 Exercise.** Prove that the function f defined by

$$f(x) = \sum_{n=1}^{\infty} \frac{\sin nx}{n^2}$$

is continuous. Can one conclude from this series representation that

$$f'(x) = \sum_{n=1}^{\infty} \frac{\cos nx}{n} \; ? \qquad (*)$$

(One can find fault with simple-minded "proofs" of equation (*) without knowing whether or not it is in fact true). ∎

1.4.40 Exercise. Let $0 < b < 1$ and let k be an odd integer such that $bk > 1 + (3\pi/2)$. Adapt the argument in Example 1.4.28D to prove that

$$f(x) = \sum_{n=0}^{\infty} b^n \cos(k^n \pi x)$$

is a nowhere differentiable continuous function. ∎

• **1.4.41 Exercise.** Find the pointwise limit of

$$f_n(x) = \frac{nx^2 - 1}{nx^2 + 1}$$

as $n \to \infty$. Is the convergence uniform? ∎

• **1.4.42 Exercise.** Find the pointwise limit of

$$f_n(x) = \left(\frac{nx^2 - 1}{nx^2 + 1}\right)^2$$

as $n \to \infty$. Is the convergence uniform? Find $\lim_{n\to\infty} f_n'(x)$. Comment on the relation to propositions in this subsection. ∎

• **1.4.43 Exercise.** Is the limit

$$\lim_{a \to 0} \frac{\sin ax}{a}$$

uniform for $x \in \mathbf{R}$? For $x \in [0, 1]$? ∎

• **1.4.44 Exercise.** Does x^n converge uniformly on $[0, 1]$? Does

$$\lim_{n\to\infty} \int_0^1 x^n \, dx = \int_0^1 \lim_{n\to\infty} x^n \, dx \; ?$$

Comment on the relation to propositions in this subsections. ∎

• **1.4.45 Exercise.** What is the negation of "$\{f_n\}$ converges uniformly on $[0,1]$"? (I.e., how would you start a proof by contradiction?) ∎

• **1.4.46 Exercise.** Let f be bounded and continuous on $\mathbf{R}$. Show that for every $t > 0$, $E_t * f$ has a continuous derivative for all $x \in \mathbf{R}$. (See Example 1.4.22C for definitions). ∎

• **1.4.47 Exercise.** The integral

$$\Gamma(x) \stackrel{\text{def}}{=} \int_0^\infty e^{-t} t^{x-1}\, dt$$

defines the *Gamma function* (see (I.5.3.29)). Prove that $\Gamma(x)$ is a continuously differentiable function of x when $x > a$, for a certain $a \in \mathbf{R}$. ∎

• **1.4.48 Exercise.** What is the negation of "$\int_0^\infty f(t,x)\, dx$ converges uniformly for $t \in [-1,1]$"? ∎

• **1.4.49 Exercise.** Pretend that you do not know the function $e(x)$ defined by

$$\sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

Prove that this series can be differentiated and integrated, term by term, as often as desired. ∎

• **1.4.50 Exercise.** True or false:

$$\int_0^1 \sum_{n=1}^{\infty} \frac{\cos nx}{n^2}\, dx = \sum_{n=1}^{\infty} \frac{\sin n}{n^3}\ ?$$ ∎

• **1.4.51 Exercise.** Does the improper Riemann integral

$$\int_0^\infty \frac{\sin yx}{x}\, dx$$

converge uniformly in y? ∎

1.4.52 Exercise. For $\mathbf{x} \in \mathbf{R}^n$, define

$$E_t(\mathbf{x}) = \frac{1}{(4\pi t)^{n/2}} e^{-\|\mathbf{x}\|_2^2/4t}.$$

This is the n-dimensional analog of the heat equation solution 1.4.22C. As in that example, let

$$E_t * f(\mathbf{x}) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} E_t(\mathbf{y}) f(\mathbf{x}-\mathbf{y})\, d\mathbf{y}$$

be the *convolution* of E_t and f. Prove: if $f : \mathbf{R}^n \to \mathbf{R}$ is continuous and bounded, then

$$\lim_{t \downarrow 0} E_t * f(\mathbf{x}) = f(\mathbf{x})$$

uniformly on $\mathbf{R}^n$. ∎

1.4.53 Exercise. Write the function

$$\frac{1}{(1-x)(1-y)}$$

as a series in powers $x^n y^m$ by multiplying together the geometric series for the factors $1/(1-x)$ and $1/(1-y)$. Find a region in the (x,y)-plane in which the series converges uniformly. ∎

Section 2 The Lebesgue Integral

The idea of completing a metric space by adding Cauchy sequences as ideal points is elegant in principle, but, because of the abstract nature of the construction, the completed metric spaces are not easy to understand. It is evident that the new objects are things that can be approximated in some sense by the objects in the original, incomplete, M, since they are limits of Cauchy sequences. Once such platitudes are out of the way, every example (or class of examples) requires separate investigation.

This section deals with a fundamental example of completion. According to Example 1.1.14C, the metric space $C([a,b])$ of continuous functions on $[a,b]$, with metric

$$\|f-g\|_1 = \int_a^b |f(x)-g(x)|\,dx$$

is not complete. The elements added in the process of completion turn out to be (equivalence classes of) functions. These functions may be violently discontinuous. The integral used to define the metric will have to be interpreted in a new way; that leads to the theory of Lebesgue integration. The first two subsections introduce the basic ideas in a heuristic way; a complete and rigorous development would take too long, and the interested reader should consult one of the many texts on measure and integration. Lebesgue theory does, however, provide some powerful tools that can be appreciated and used by people who have not slugged through all the underlying theory.

Subsection 1 explains why the calculus intgral is too limited to deal with the standard limiting processes of analysis. The classical (=calculus, dating back to 1850) and modern (1900) definitions of the integral are introduced. Details are presented in Subsections 3 and 4. I cannot stress too often that the discussion is superficial; a complete development in a course of analysis takes the better part of a year. It is, nevertheless, possible to learn to use the results of Lebesgue theory without knowing the subtleties of the constructions. Mathematicians who drive cars without knowing the location of the water pump are, strangely enough, irritated when someone "drives" Lebesgue methods while being ignorant of the inner workings of the theory. In this chapter, I take the attitude that one can drive a car even if one cannot build one. Subsections 6 and 7 develop more fundamental ideas of Lebesgue integration theory. The official definition of the L^p spaces is given (at last), and Fubini's basic theorem about multiple integrals is stated (not proved!) and illustrated.

Subsection 2.1 Completion of $C([0,1])$

The completion of $C([0,1])$ consists of the original, continuous functions, together with all the limits of L^1 Cauchy sequences. What sorts of objects are added? It is best to look at pictures to get a feeling for the situation; the proofs are done by an entirely different method anyway.

2.1.1 Definition. *Let $A \subseteq \mathbb{R}$ be a set. The characteristic function of A is the function c_A defined by*

$$c_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

Note that a characteristic function c_A is not continuous at boundary points of the set A (because there are points from A and A^c arbitrarily close to every boundary point.) Characteristic functions of fairly messy sets wind up in the completion of $C([0,1])$.

The two sequences of functions shown in Figure 7 tend, respectively, to the characteristic functions $c_{(a,b)}$ and $c_{[a,b]}$.

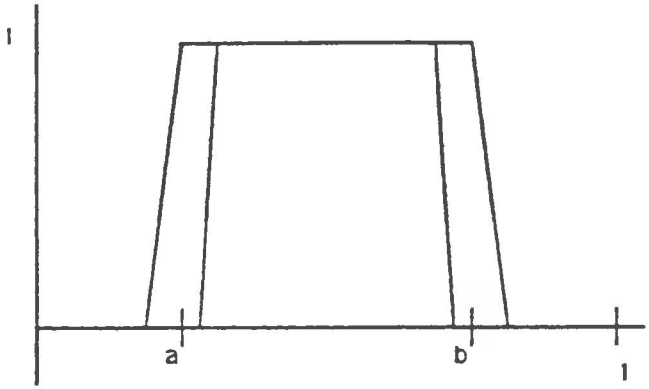


Figure 7. *Approximating characteristic functions*

In the pointwise sense,

$$f_n(x) \to c_{(a,b)}(x) \quad \text{for all } x \in [0,1]$$

and

$$g_n(x) \to c_{[a,b]}(x) \quad \text{for all } x \in [0,1].$$

The pointwise limits of the two sequences are different functions (the limiting values at a, b are not the same). By computing the area under the graphs of $|f_n - f_m|$ and $|g_n - g_m|$ one sees also that both sequences are Cauchy in the L^1 metric. In fact, by looking at areas once more, one sees that

$$\|f_n - c_{(a,b)}\|_1 \to 0, \quad \|g_n - c_{[a,b]}\|_1 \to 0$$

as $n \to \infty$.

But the area under $|f_n - g_n|$ tends to zero as well, so that the Cauchy sequences $\{f_n\}$ and $\{g_n\}$ are *equivalent* in the sense of Definition 1.2.9. Even though the sequences have different *pointwise* limits, they determine just *one* point of the completion. While we may very naturally think about this new "point" of the completion $\overline{C([0,1])}$ as being a discontinuous function, that function is not uniquely determined. One can imagine many other Cauchy sequences equivalent to the two above, all having different pointwise limits. One of them is shown in Figure 8. The pointwise limit of this sequence $\{h_n\}$ is the characteristic function of the set $(a,b] \cup \{1\}$.

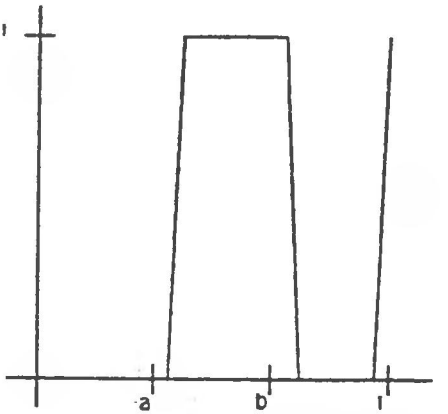


Figure 8. *Another Cauchy sequence*

This simple example reinforces a most important lesson: the points of the completion are not functions, but rather *equivalence classes of functions.*

One can use the characteristic functions $c_{(a,b)}$ to build more complicated characteristic functions, all of which represent elements of the completion. Let

$$A_n = \bigcup_{j=1}^{n} (a_j, b_j)$$

be a union of disjoint open intervals. It should be clear that $c_n \stackrel{\text{def}}{=} c_{A_n}$ can be understood as limit of a Cauchy sequence in L^1 (Figure 9).

Consider next a union $A = \bigcup_{j=1}^{\infty}(a_j, b_j)$ of countably many disjoint intervals. For each n, the characteristic function c_n of the partial union is already in the completion. But now recall from 1.2.10 that when completing a metric space, one also extends the original metric to the new objects. Therefore, we might ask whether $\{c_n\}$ is itself a Cauchy sequence (made up entirely of points not in the original $C([0,1])$). Intuitively, the answer is "yes": the distance between c_n and c_m should be

$$\int_0^1 |c_n(x) - c_m(x)|\, dx,$$

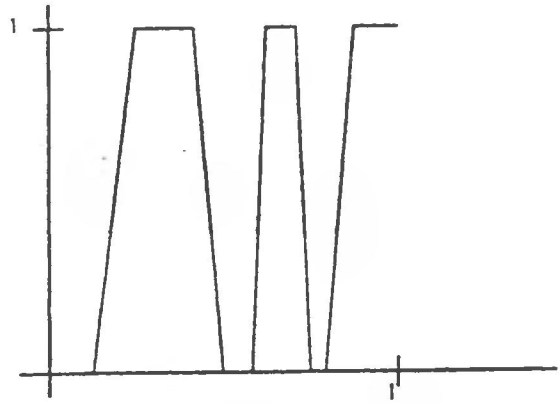


Figure 9. *Approaching the characteristic function of* $\bigcup(a_j, b_j)$

which is just the length of the set $\bigcup_{j=m+1}^{n}(a_j, b_j)$ (if $n > m$, say), and that length gets to be arbitrarily small since the total available length of $[0,1]$ is finite. Thus the characteristic function c_A also represents an element of the completion. Every open subset of R is a countable union of disjoint open intervals (Proposition I.3.1.4), so we have built some pretty complicated characteristic functions by now — think of the (open) complements of Cantor-type sets, for instance.

No need to stop: let $A_1 \supset A_2 \supset A_3, \ldots$ be a decreasing sequence of open subsets of $[0,1]$, and let $B = \bigcap_{n=1}^{\infty} A_n$. B is a G_δ set. You should try to convince yourself that $\{c_{A_n}\}$ is Cauchy in the extended metric, and that therefore c_B will also represent an element of the completion of $C([0,1])$. One can build characteristic functions of F_σ sets in the same way. This is fun! Shall we make even uglier sets? As luck would have it, we can't: every Cauchy sequence in $C([0,1])$ that converges (in the *extended* metric) to the characteristic function of some set is equivalent to one that converges to the characteristic function of a G_δ set. The proof requires different ideas (Lebesgue measure theory), and will not be given, but the result does allow me to turn to other topics.

Characteristic functions of sets are quite dull, as far as functions go: they only take on the values 0 and 1. It would not be difficult to construct more general functions by varying the heights of the bumps in Figure 9. The extra work would not, however, shed more light on the two main points: the completion of $C([0,1])$ really does consist of *equivalence classes of functions*, and the functions in those equivalence classes can be pretty wild. Further progress requires new ideas.

The study of the completion of $C([0,1])$ in the L^1 metric has two aspects. The integral defining the metric must be extended to the elements added to the completion. That is the subject of the next four subsections. Once one understands the elements of the completion, and knows how to define distance between them, one tries to study the completed metric space. A few results in this direction are presented in Subsection 6.

2.1.2 Exercise. The whole discussion involving characteristic functions can be recast

to eliminate all references to L^1 and $C([0,1])$. Let M be the collection of subsets of $[0,1]$ which are unions of finitely many disjoint open intervals. For $A, B \in M$, define

$$d(A,B) = \text{ length of } (A-B)\cup(B-A);$$

here "length" means total length of the intervals making up the set $(A-B)\cup(B-A)$. Show that d is a metric on M. (This is fairly easy if you use the properties of "length" as encoded in the axioms of measure theory, but it will be a nightmare if you write each set as a union of intervals.) Show that

$$d(A,B) = \int_0^1 |c_A(x) - c_B(x)|\,dx.$$

Above, when I took ever more complicated Cauchy sequences of characteristic functions, I was just completing the metric space (M,d). ∎

Subsection 2.2 On Defining Integrals

When f, g are continuous functions, the L^1-metric

$$d(f,g) = \int_0^1 |f(x) - g(x)|\,dx$$

involves only the familiar integral from calculus. Completion of $C([0,1])$, however, has introduced some very discontinuous functions, and one naturally wonders whether the extended metric, $\bar{d}$, on the completion can still be interpreted as an integral. It can, but the definition of the integral must be changed. In this subsection, the calculus definition, due to Riemann and Weierstraß, and the more modern definition due to Lebesgue are outlined and compared at a heuristic level.

The calculus integral is officially known as the *Riemann integral*. It is defined as follows. For simplicity, the integration interval is taken to be $[0,1]$, but any *finite* interval $[a,b]$ could be used instead.

Let f be a (not necessarily continuous function) on $[0,1]$. A *partition* Δ of $[0,1]$ is a finite set of points $0 = t_0 < t_1 < \ldots < t_n < t_{n+1} = 1$. The area under the graph of f is now approximated by rectangles:

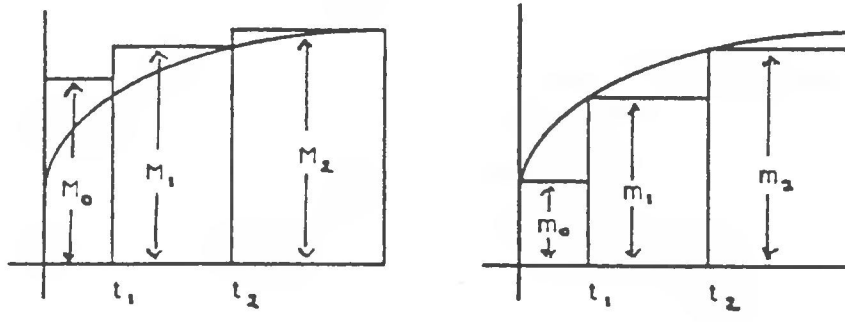


Figure 10. *Circumscribed and inscribed rectangles*

Let $M_i = \max\{f(t) \mid t_i \le t \le t_{i+1}\}$, and let m_i be the minimum of f over that same subinterval. Set

$$U(f;\Delta) = \sum_{i=0}^{n} M_i(t_{i+1} - t_i), \qquad L(f;\Delta) = \sum_{i=0}^{n} m_i(t_{i+1} - t_i).$$

These are the approximations to the area under the graph of f by circumscribed and inscribed rectangles, respectively.

Now define

$$U(f) = \inf_{\text{all } \Delta} U(f;\Delta), \qquad L(f) = \sup_{\text{all } \Delta} L(f;\Delta).$$

These quantities are called the *upper* and *lower Riemann integrals of* f.

2.2.1 Definition. *The function f is said to be* ***Riemann integrable*** *if $U(f) = L(f)$, and in that case the common value of the upper and lower Riemann integrals is called the* ***Riemann integral*** *of f, written*

$$(R)\int_0^1 f(t)\,dt. \tag{2.2.2}$$

Roughly speaking, Riemann integrability requires that the function must not oscillate too much on small enough intervals. When the partition Δ is very fine (meaning that $|t_{i+1} - t_i|$ is small), $U(f;\Delta)$ and $L(f;\Delta)$ should be close to their common value, and hence close to each other. The difference

$$0 \le U(f;\Delta) - L(f;\Delta) = \sum (M_i - m_i)(t_{i+1} - t_i)$$

is small when the differences $M_i - m_i$ between the maxima M_i and the minima of f are very small on the majority of the intervals $[t_i, t_{i+1}]$. For example, when f is continuous, it is uniformly continuous, and $M_i - m_i$ is uniformly small ($< \epsilon$) whenever $|t_{i+1} - t_i| <$ some δ: every continuous function is Riemann integrable. This is the most basic result about the Riemann integral, so I give the official proof.

2.2.3 Proposition. *If f is a continuous function on $[0, 1]$, then it is Riemann integrable.*

Proof. We need to show that $L(f) = U(f)$; to this end, we prove that

$$|U(f) - L(f)| < \epsilon \tag{2.2.4}$$

for arbitrary $\epsilon > 0$. This will follow from two estimates:

$$\text{For all partitions } \Delta, \quad L(f;\Delta) \le L(f) \le U(f) \le U(f;\Delta), \tag{2.2.5}$$

and

(2.2.6) for sufficiently fine Δ, $|U(f;\Delta) - L(f;\Delta)| < \epsilon$.

To prove (2.2.5), first note that by definition of $U(f)$, (resp. $L(f)$) as the infimum $\inf_\Delta U(f;\Delta)$ (resp. as the supremum $\sup_\Delta L(f;\Delta)$), we have

$$L(f;\Delta_1) \le L(f) \quad \text{and} \quad U(f) \le U(f;\Delta_2)$$

for any pair of partitions Δ_1, Δ_2. There remains the inequality

(2.2.7) $$L(f) \le U(f).$$

To get this, let $\Delta_1 : 0 < t_1 < \ldots < t_n < 1$ and $\Delta_2 : 0 < s_1 < \ldots < s_m < 1$ be two partitions, and let Δ be the partition whose division points are the t_i together with the s_j, arranged in increasing order. (Δ is called the *common refinement* of Δ_1 and Δ_2.) Now look at an interval $[t_i, t_{i+1}]$ of Δ_1. It is the union of one or more intervals of Δ, say

$$t_i < s_{j_1} < \ldots < s_{j_k} < t_{i+1}.$$

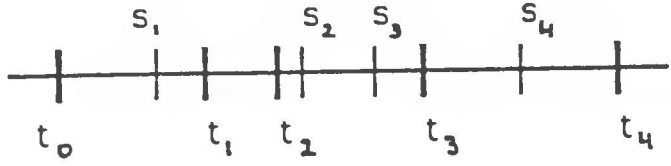


Figure 11. *The common refinement of two partitions*

(See Figure 11.) The minimum m_i of f over $[t_i, t_{i+1}]$ is no greater than the minimum over any subinterval $[s_{j_r}, s_{j_{r+1}}]$. Hence,

$$L(f;\Delta_1) \le L(f;\Delta).$$

A similar argument shows that

$$U(f;\Delta) \le U(f;\Delta_2),$$

and so

$$L(f;\Delta_1) \le L(f;\Delta) \le U(f;\Delta) \le U(f;\Delta_2),$$

or

(2.2.8) $$L(f;\Delta_1) \le U(f;\Delta_2).$$

Now fix Δ_2 and take the supremum of $L(f;\Delta_1)$ over all partitions Δ_1. By (2.2.8) $L(f) \le U(f;\Delta_2)$. Taking the infimum of $U(f;\Delta_2)$ over all Δ_2 gives (2.2.7) and the proof of (2.2.5) is finished.

The proof of the second inequality, (2.2.6), has already been sketched; here it is in full detail. Given $\epsilon > 0$, uniform continuity of f on the compact set $[0,1]$ (Proposition 1.4.5) guarantees existence of a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ whenever $|x - y| < \delta$. Now let Δ be any partition of $[0,1]$ with $|t_{i+1} - t_i| < \delta$ for all i. By compactness, Proposition 1.3.8, there are $x, y \in [t_i, t_{i+1}]$ such that $f(x) = M_i$ and $f(y) = m_i$. Since $|x - y| < \delta$, we have

$$M_i - m_i = f(x) - f(y) = |f(x) - f(y)| < \epsilon.$$

Hence

$$0 \le U(f;\Delta) - L(f;\Delta) = \sum_i (M_i - m_i)(t_{i+1} - t_i) < \epsilon \sum_i (t_{i+1} - t_i) = \epsilon \times 1 = \epsilon.$$

The proof of the proposition is now complete. ∎

It must be emphasized that continuity of f is *not necessary* for existence of the Riemann integral of f. If f has points of discontinuity, one should try to "isolate" them in the following way. Given a partition Δ, subdivide it further and include the discontinuity points in very small intervals $[t_i, t_{i+1}]$. It may then happen that the contribution of $(M_i - m_i)(t_i - t_{i+1})$ to $U(f;\Delta) - L(f;\Delta)$ is small: not because M_i and m_i are close, as they were in the proof of Proposition 2.2.3, but because the intervals enclosing the discontinuity points have arbitrarily small total length. That reminds one of the definition of "set of measure zero," and indeed there is the following fundamental theorem, whose proof will not be given:

2.2.9 Theorem. *A function f on $[0,1]$ is Riemann integrable if, and only if, the set of points of discontinuity of f has Lebesgue measure zero.*

2.2.10 Example. The discontinuity set in Theorem 2.2.9 may be dense in $[0,1]$. Functions with this property are difficult to picture, but surprisingly, they do arise "in nature." The following example requires some elementary probability theory.

I toss a fair coin until a head turns up. The number of tosses required is called X. The first head occurs on toss n if there are $n-1$ tails followed by a head, and this has probability 2^{-n}. We write

$$P(X = n) = 2^{-n}.$$

Suppose you perform the same experiment; the number of tosses needed to get your first head is called Y, and $P(Y = m) = 2^{-m}$.

Now let $Z = X/Y$. Z measures how much sooner (or later) than you I get heads. The possible values of Z are the positive rational numbers. Let n, m

be positive integers with no common factor; I want to compute $P(Z = n/m)$. Now, $Z = n/m$ when $X = \frac{n}{m}Y$. For X to have an integer value, Y must be $m, 2m, 3m, \ldots$, and then $X = n, 2n, 3n, \ldots$. Hence

$$P(X = \frac{n}{m}Y) = P(Y = m \text{ and } X = n) + P(Y = 2m \text{ and } X = 2n) + \ldots.$$

Because my tossing is independent of your tossing,

$$P(Y = km \text{ and } X = kn) = P(Y = km) \cdot P(X = kn) = 2^{-k(n+m)},$$

and

$$P(X = \frac{n}{m}Y) = \sum_{k=1}^{\infty} 2^{-k(n+m)} = \frac{1}{2^{m+n} - 1}.$$

Now define a function f_Z on $(0, \infty)$:

$$f_Z(x) = \begin{cases} 0, & \text{if } x \text{ is irrational,} \\ 1/(2^{n+m} - 1), & \text{if } x = n/m \text{ in lowest terms.} \end{cases}$$

This is called the *probability density of the random variable* Z. I claim that f_Z is *continuous* at all irrationals and *discontinuous* at all rationals. Thus, its set of discontinuities is a dense set of measure zero.

First, let $x = n/m$ be rational. There are irrationals y arbitrarily close to x, and so

$$f_Z(x) - f_Z(y) = \frac{1}{2^{n+m} - 1} - 0$$

cannot be made arbitrarily small. Hence f_Z is not continuous at rational numbers.

Now let x be irrational, so that $f_Z(x) = 0$. Given $\epsilon > 0$, choose M_ϵ so that $1/(2^m - 1) < \epsilon$ when $m > M_\epsilon$. There are only finitely many rational numbers n/m within distance 1 of x that have denominator $m \leq M_\epsilon$. Choose $\delta > 0$ so small that the neighborhood $(x - \delta, x + \delta)$ contains none of these numbers. Then, whenever y is rational and $|x - y| < \delta$, $y = n/m$ with $m > M_\epsilon$. Hence

$$|f_Z(x) - f_Z(y)| = \frac{1}{2^{n+m} - 1} \leq \frac{1}{2^m - 1} < \epsilon.$$

Of course, when $|x - y| < \delta$ and y is irrational, then $|f_Z(x) - f_Z(y)| = 0 < \epsilon$. So f_Z is continuous at irrational points.

According to Theorem 2.2.9, this wild function f_Z is Riemann integrable over any finite interval. The integral turns out to be zero. ∎

So far, we have dealt with functions whose oscillation is controlled enough for the Riemann integral to exist. At the other extreme, there are functions that have large oscillation on every interval.

2.2.11 Example. Let D be the so-called *Dirichlet function*,

$$D(t) = \begin{cases} 1, & \text{if } t \text{ is rational,} \\ 0, & \text{if } t \text{ is irrational.} \end{cases}$$

Every interval $[t_i, t_{i+1}]$ of a partition contains both rationals and irrationals, so $M_i = 1$ and $m_i = 0$. Thus, $U(D) = 1$, $L(D) = 0$, and the Dirichlet function is not Riemann integrable.

This non-integrable function is a limit of continuous functions:

$$D(t) = \lim_{m\to\infty} \lim_{n\to\infty} (\cos(m!\,\pi t))^{2n}. \tag{2.2.12}$$

So, while all continuous functions are Riemann integrable, limits of continuous functions need not be; this is one of the earliest examples of the "incompleteness" of the Riemann integration theory. ∎

You may consider the Dirichlet function to be rather artificial. It is, but it was quite important in the development of modern analysis. The history of concepts like "function" and "integral" gives interesting insight into today's standard definitions.

In the 18th century, a *function* was defined by an explicit formula; *discontinuous functions* were not just functions with jumps, but more commonly functions defined by several formulas ($f(x) = 0$ for $x \le 0$, and $f(x) = x$ for $x > 0$). Integrals were identified with antiderivatives via the fundamental theorem of calculus. The existence of an antiderivative of every function was largely taken for granted.

When Fourier introduced his trigonometric series, the class of functions was enlarged tremendously. It now seemed that very general graphs could be described by a formula, albeit one that involved infinitely many terms. Fundamental problems were posed. Does a Fourier series really mean anything, i.e., does it converge? How can one find the antiderivative—can one integrate each term in the series and sum the individual antiderivatives?

These are almost impossibly difficult questions as long as the only concept of "function" is "that which is defined by the Fourier series." Dirichlet pointed the way to the picture of a function as a correspondence, something that assigns numbers in the range to numbers in the domain, but does not necessarily do so via any explicit formula. This procedure can result in objects far stranger than had ever been studied in analysis—such as the Dirichlet function. Do these general functions have antiderivatives? Do they have Fourier series? What class of functions can one get from Fourier series?

Reasonably complete answers to these and other questions were put together during the century after Dirichlet's work. The crucial ingredient is the integration theory of Lebesgue which, from the point of view I have taken so far, is the completion of Riemann's integration theory in the L^1 metric.

We can now understand, at least intuitively, the inadequacy of the Riemann integral for description of the metric $\int_0^1 |f(x) - g(x)|dx$ on the completion of $C([0,1])$: the functions that are added as "limits" of Cauchy sequences may be so wildly oscillatory that they are no longer Riemann integrable. The concept of

integral must be revised. If you have not seen Lebesgue integration before, the very idea of "new kind of integral" will sound confusing. To break down the first barriers, Lebesgue once told a little parable.

A man wants to compute the total value of a bag of coins. Method 1 is to put the coins in a long row, without regard to denomination, and then to add up the values of the coins while moving from left to right along the row. Method 2 is to group all the 1 cent (or franc) pieces, all the 5 cent pieces, and so forth, to find the subtotals for each denomination, and then to add those. Both methods lead, in principle, to the same result.

Figure 12. *German and French merchants*

Method 1 is Riemann's: to find the area under the graph of f, we move from left to right, adding up areas of little rectangles. A tall rectangle may be followed by a very short one, and there is only as much regularity as the given f will allow. For some functions (Dirichlet's, for example), this procedure does not work.

Method 2 is Lebesgue's. He groups the values of f before trying to approximate the area under its graph. For the Dirichlet function, Lebesgue's procedure is particularly simple. The set $\{t \mid D(t) = 1\}$, i.e. the set of rationals, has measure zero, or zero length. The set $\{t \mid D(t) = 0\}$ has measure 1, or length 1. The area under the graph of D should therefore be $1 \times 0 + 0 \times 1 = 0$.

More generally, suppose that f is bounded, $-M \le f(t) \le M$, and partition the *range* of f rather than its *domain*: $-M-1 = y_0 < y_1 < \ldots < y_n < y_{n+1} = M+1$. The reason for extending the partition beyond $\pm M$ is explained in a moment. Let μ_i be the measure ("length") of the set

$$\{t \mid y_i \le f(t) < y_{i+1}\}. \tag{2.2.13}$$

Define the Lebesgue sum (as opposed to Riemann sum)

$$\sum_{i=0}^{n} y_i \mu_i. \tag{2.2.14}$$

The *Lebesgue integral* of f is the limit of these approximating sums, as the size of the partition (the maximum $|y_i - y_{i+1}|$) tends to zero.

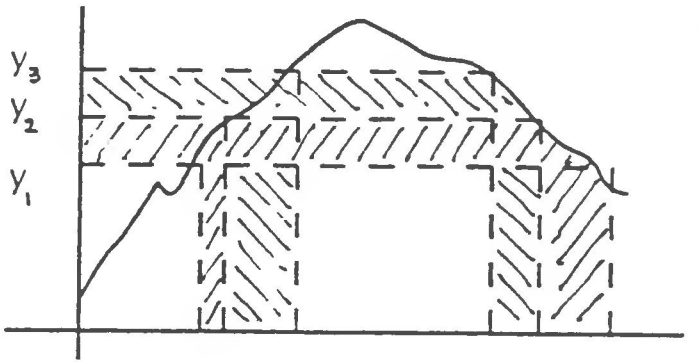


Figure 13. *Lebesgue partition of the y-axis*

2.2.15 Example. Let E_1, E_2, E_3 be disjoint measurable subsets of $[0,1]$ whose union is all of $[0,1]$. Define the function f by $f(t) = i$ for $t \in E_i$. So, f takes on the values $1, 2, 3$ only, but perhaps on very complicated sets. You should draw a picture of such an f; if nothing else comes to mind, take $E_1 = [0, 1/3)$, $E_2 = [1/3, 2/3)$, and $E_3 = [2/3, 1]$.

Now suppose we were to subdivide the y–axis from 0 to 3 only: $0 = y_0 < y_1 < \ldots < y_{n+1} = 3$. Then the sets

$$\{t \,|\, y_i \le f(t) < y_{i+1}\}$$

would *not* include the set E_3, where $f(t) = 3$, because $y_n \le f(t) < 3$ omits the value 3. The Lebesgue approximation for the area under the graph of f would be completely wrong.

But apply the prescription given above. We have $-M \le f(t) \le M$, for $M = 3$. Now subdivide the interval $[-4, 4]$ on the y–axis. By extending the partition beyond the values actually assumed by the function f, we guarantee that the whole range of f is covered. One or another of the intervals $[y_i, y_{i+1})$ will include every possible number $f(t)$.

By letting the largest interval in the partition tend to zero, one will find the integral of f, or area under the graph of f:

$$1 \times \mu(E_1) + 2 \times \mu(E_2) + 3 \times \mu(E_3),$$

where $\mu(E)$ denotes the Lebesgue measure, or length, of the set E. ∎

We will not prove that the Lebesgue approximating sums have a limit. The steps are technical, and of course the existence of the integral depends on properties of the function f. The key requirement is that the sets (2.2.13) should have a length assigned to them. In the language of Chapter I.5.2, they should belong to the σ-algebra of Lebesgue–measurable sets in $[0,1]$. This condition constrains the "wildness" of f permitted in Lebesgue integration theory. The upshot is that Lebesgue's definition of the integral makes sense for a much larger class than does Riemann's. Fortunately, when *both* definitions make sense, then they give the *same values to the integrals*. This is reassuring: it means you don't have to unlearn anything you already know about properties of Riemann integrals.

It is now time to stop being vague, and to begin to collect some precise definitions and propositions. The next few subsections are devoted to this task.

• **2.2.16 Exercise.** If f and g are Riemann integrable, then so is $f+g$. ∎

• **2.2.17 Exercise.** Define a function $f : \mathbf{R} \to \mathbf{R}$ by

$$f(x) = \begin{cases} 0, & \text{if } x \text{ is irrational,} \\ 1/q, & \text{if } x = p/q \text{ in lowest terms.} \end{cases}$$

At what $x \in \mathbf{R}$ is f continuous? ∎

2.2.18 Exercise. Let $\{a_n\}$ be an infinite sequence of distinct points in $(0,1)$, and let $\sum b_n$ be a convergent series of positive terms. For $x \in [0,1]$, define

$$f(x) = \sum_{0 \le a_n < x} b_n.$$

(The notation means: find all a_n in the interval $[0,x)$, and add up the corresponding b_n.) Show that

$$\lim_{h \downarrow 0} f(a_k + h) = f(a_k) + b_k, \quad \lim_{h \downarrow 0} f(a_k - h) = f(a_k),$$

and show that f is continuous at x when x is not one of the a_k.

This f is a nondecreasing function which is continuous except at the a_k, and has prescribed (upward) jump b_k at $x = a_k$. In fact, f is (morally, at least) constant where it does not jump. These functions f are called *pure jump functions*, or *saltus functions*. ∎

2.2.19 Exercise. A function $f : [0,1] \to \mathbf{R}$ is said to be ***nondecreasing*** (on $[0,1]$), if $f(x_1) \le f(x_2)$ whenever $x_1 < x_2$.

a) Let f be nondecreasing. Show that

$$f(x^+) \stackrel{\text{def}}{=} \lim_{h \downarrow 0} f(x+h) \quad \text{and} \quad f(x^-) \stackrel{\text{def}}{=} \lim_{h \downarrow 0} f(x-h)$$

exist (and are finite) for every $x \in (0,1)$.

b) Show that a nondecreasing function has at most countably many points of discontinuity. (Hint: if A is an uncountable set of positive numbers, then there is a sequence $\{c_n\} \subset A$ for which $\sum c_n = \infty$.) ∎

2.2.20 Exercise. This continues the preceding exercise. Show that a nondecreasing function on $[0,1]$ is Riemann integrable. (Hint: if you cover a large but finite number of discontinuity points by a union of intervals of total length $< \epsilon$, the total possible jump of f on any interval outside that set will be $< \epsilon$.)*) ∎

2.2.21 Exercise*. A function $f : [0,1] \longrightarrow \mathbf{R}$ is said to be of *bounded variation*, if there exists a constant $c \geq 0$ such that for every partition of $[0,1]$,

$$\sum_{i=0}^{n} |f(t_{i+1}) - f(t_i)| \leq c.$$

(The number n will vary with the partition chosen.)

a) Show that $f(t) = t$ is of bounded variation.
b) Show that f is of bounded variation if it is continuously differentiable on $[0,1]$.
c) Show that a nondecreasing function is of bounded variation.
d) Show that the difference of two nondecreasing functions is of bounded variation.
e) Let f be of bounded variation. Show that it is the difference of two nondecreasing functions. (Hint: $f = V - (V - f)$, where

$$V(x) \stackrel{\text{def}}{=} \sup_{\text{all partitions of } [0,x]} \sum_{i=0}^{n} |f(t_{i+1}) - f(t_i)| .)$$

f) If f is of bounded variation, then it is Riemann integrable. ∎

2.2.22 Exercise. Establish the limit (2.2.12). ∎

Subsection 2.3 Measurable Functions

When $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ are topological spaces, the interesting functions $f : X \longrightarrow Y$ are the *continuous* ones. *Continuity* is the most basic property of functions that can be expressed within the axiom system of topology. When $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ are *measurable spaces*, the *measurable* functions are fundamental. Measurable functions are "regular" enough to preserve the structures of measure theory and, as we will see, they are appropriate for the generalized integration theory.

2.3.1 Definition. *Let $(X, \mathcal{B})$ and $(Y, \mathcal{C})$ be measurable spaces. A function $f : X \longrightarrow Y$ is said to be* ***measurable*** *if $f^{-1}(A) \in \mathcal{B}$ for all $A \in \mathcal{C}$.*

This definition deals only with membership in a σ-algebra. It is not necessary to have a measure defined on $(X, \mathcal{B})$.

Remarks

A. In Section I.5.2, measurable sets were interpreted as "observable sets" or "events" in the setting of probability theory. Measurable functions, by the same token, are "observables". When the measurable space is a sample space

*) - this exercise is probably easier if you disregard the "Hint".

$(\Omega, \mathcal{B})$, a real-valued function f on Ω can be thought of as a measurement: for each outcome ω, one gets a number, $f(\omega)$. Measurability of $f : \Omega \longrightarrow \mathbf{R}$ means, in particular, that

$$f^{-1}((a,b)) = \{\omega \mid f(\omega) \in (a.b)\}$$

should belong to $\mathcal{B}$ (i.e., should be observable), since (a, b) is a measurable subset of $\mathbf{R}$. In words: one should be able to tell whether or not the measurement performed during the course of one experiment falls between the numbers a and b.

In probability theory, measurable functions are called *random variables*.

B. Compare the definition of *measurability* to the definition of *continuity*: if $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ are topological spaces, then $f : X \longrightarrow Y$ is *continuous* if $f^{-1}(A) \in \mathcal{T}$ for every $A \in \mathcal{S}$.

Recall (Proposition I.5.2.22) that a topological space comes equipped with a natural σ-algebra: the *Borel σ-algebra*, which is the smallest σ-algebra containing all open sets. When the σ-algebras $\mathcal{B}$ and $\mathcal{C}$ in Definition 2.3.1 are the Borel σ-algebras corresponding to topologies $\mathcal{T}$ and $\mathcal{S}$ on X and Y, measurability of f implies that

$$f^{-1}(\text{open set}) \in \mathcal{B}.$$

This asks far less of f than

$$f^{-1}(\text{open set}) \in \mathcal{T}.$$

For example, if the domain X and range Y of f are both $\mathbf{R}$ with its usual topology and Borel σ-algebra, then the difference between continuous and measurable functions is particularly clear. For a continuous f, the set

$$f^{-1}((-\infty, \alpha)) = \{x \in \mathbf{R} \mid f(x) < \alpha\}$$

must be open (a union of countably many disjoint open intervals); for a measurable f, the same set need merely belong to the Borel σ-algebra: it could be the Cantor set, the set of rationals, of irrationals, and so forth. ∎

•**2.3.2 Exercise.** Verify that the Dirichlet function defined in 2.2.11 is Borel measurable on $\mathbf{R}$. ∎

For purposes of integration theory, one is interested in real-valued functions defined on a measure space. More complicated functions (complex-valued, vector-valued) are built from real-valued ones. We therefore want to understand Definition (2.3.1) when the range Y is $\mathbf{R}$ equipped with the usual Borel σ-algebra. Furthermore, it is convenient to allow functions to take on the values $\pm\infty$; this cuts down on lists of exceptional cases.

2.3.3 Definition. *The set*

$$\mathbf{R}_{ex} = \mathbf{R} \cup \{-\infty, \infty\} = [-\infty, \infty]$$

*is called the (set of) **extended real numbers**. It is subject to the following rules of arithmetic:*

i) if $a \in \mathbf{R}$, then $a \pm \infty = \pm\infty$;
ii) if $a > 0$, then $a \times \pm\infty = \pm\infty$;
iii) $0 \times \pm\infty = 0$;
iv) $\infty - \infty$ is undefined.

Rule iii) ensures, for example, that

$$\int_{-\infty}^{\infty} 0 \, dx = 0,$$

since length $\times$ width $= \infty \times 0$ will be 0.

2.3.4 Definition. *Let $(X, \mathcal{B})$ be a measurable space. A function $f : X \longrightarrow \mathbf{R}_{ex}$ is said to be **measurable** if the set $\{t \mid f(t) < \alpha\}$ belongs to $\mathcal{B}$ for each $\alpha \in \mathbf{R}$.*

I now prove one important result about measurable real-valued functions, to show how one works with the definition. Other propositions are left to the exercises. You should read those exercises, even if you don't want to work them.

2.3.5 Proposition. *Let $(X, \mathcal{B})$ be a measurable space, and let f_n, $n \geq 1$ be a sequence of measurable (real-valued) functions which has a limit at every $x \in X$: $\lim_{n\to\infty} f_n(x) = f(x)$ exists for each x. Then f, the limit function, is also measurable.*

Proof. We must show that

$$\{x \mid f(x) < \alpha\} \tag{*}$$

belongs to the σ–algebra $\mathcal{B}$ for every $\alpha \in \mathbf{R}$. It is given that, for each n and β,

$$\{x \mid f_n(x) < \beta\} \in \mathcal{B}. \tag{**}$$

The obvious strategy is to show that the set (*) is obtained from sets of type (**) by σ-algebra operations. It takes some experience to find the correct relation:

$$\{x \mid f(x) < \alpha\} = \bigcup_k \bigcup_n \bigcap_{m>n} \{x \mid f_m(x) < \alpha - \frac{1}{k}\}. \tag{2.3.6}$$

To prove this, we show as usual that each side in (2.3.6) is a subset of the other side.

Suppose that $f(x) < \alpha$. Then, for some k, we have $f(x) < \alpha - 2/k$. On the other hand, $f_m(x) \longrightarrow f(x)$ means that $f_m(x)$ is within $1/k$ of $f(x)$ for $m >$ some n.

Hence, for $m >$ some n, we have $f_m(x) < \alpha - 1/k$. This puts x in the set on the right side of (2.3.6): for *some* k and *some* n and for *all* $m > n$, $f_m(x) < \alpha - 1/k$.

Conversely, suppose that x belongs to the set on the right side of (2.3.6). Then for some k, there is an m such that $f_m(x) < \alpha - 1/k$ for all $m >$ some n. It follows that the limit $f(x)$ is $\leq \alpha - 1/k$, and in particular $f(x) < \alpha$. ∎

Exercises

• **2.3.7 Exercise.** The last result shows that measurability of (real-valued) functions is preserved under pointwise limits. This is a crucial feature of Lebesgue integration theory. Riemann-integrable functions, on the other hand, are not so well behaved. Enumerate the rationals in $[0, 1]$ as $r_1, r_2, \ldots$, and define

$$D_n(x) = \begin{cases} 1, & \text{if } x = r_1, \ldots, r_n, \\ 0, & \text{otherwise.} \end{cases}$$

Show directly that D_n is Riemann integrable. The limit $\lim_{n\to\infty} D_n(x)$ is the Dirichlet function D, which, as we have seen, is not Riemann integrable. ∎

• **2.3.8 Exercise.** Show that the function f_Z from Example 2.2.10 is measurable on $[0, \infty)$. ∎

• **2.3.9 Exercise.** There are many equivalent definitions of measurability of real-valued functions on a measure space. Prove that each of the following properties implies that f is measurable, and that a measurable $f : X \to \mathbf{R}$ satisfies each condition.

i) For every $\alpha \in \mathbf{R}$, $\{x \mid f(x) \leq \alpha\} \in \mathcal{B}$.
ii) For every $\alpha \in \mathbf{R}$, $\{x \mid f(x) > \alpha\} \in \mathcal{B}$.
iii) For every $\alpha \in \mathbf{R}$, $\{x \mid f(x) \geq \alpha\} \in \mathcal{B}$.
iv) For every interval (a, b), $\{x \mid f(x) \in (a, b)\} \in \mathcal{B}$.

Hint: Remember that a σ-algebra is closed under operations like

$$\bigcap_{n=1}^{\infty} (-\infty, \alpha + \frac{1}{n}).$$ ∎

• **2.3.10 Exercise.** Let f be a measurable, extended real-valued function. Show that the sets $f^{-1}(\infty)$ and $f^{-1}(-\infty)$ are measurable. ∎

• **2.3.11 Exercise.** Let X be a set, and let $B_1, \ldots, B_N$ be disjoint sets whose union is X. Let $\mathcal{B}$ be the smallest σ-algebra containing all B_j.

a) Describe $\mathcal{B}$ by listing all the sets it contains.
b) Describe all measurable functions as explicitly as possible. ∎

2.3.12 Exercise. Let $(\Omega, \mathcal{B})$ be the measurable space consisting of Brownian motion sample space equipped with its Borel σ-algebra. For each of the following functions, decide whether or not it is measurable.

a) $f(\omega) = \omega(1)$;
b) $f(\omega) = \sup_{0 \leq t < 1} |\omega(t)|$;
c) $f(\omega) = \inf_{T \geq 0} \{t \mid \omega(t) \neq 0 \text{ for } t \geq T\}$. ∎

• **2.3.13 Exercise.** Let $(X, \mathcal{T})$ be a topological space, and let $\mathcal{B}$ be the Borel σ-algebra. Let $f : X \longrightarrow \mathbf{R}$ be continuous. Show that f is measurable. ∎

• **2.3.14 Exercise.** Let $(X, \mathcal{B})$ be a measurable space. Prove that the sum and product of two measurable functions from X to $\mathbf{R}$ are measurable. ∎

• **2.3.15 Exercise.** Let $(X, \mathcal{B})$ be a measurable space, and let $\{f_n\}$ be a sequence of measurable functions from X to $\mathbf{R} \cup \{\pm\infty\}$. Then $\sup f_n$ and $\inf f_n$ are measurable. ∎

2.3.16 SUPPLEMENT. Measure-preserving transformations

I don't want to leave the impression that measurable functions are only important because they can often be integrated. Functions from a measure space to itself play a big role in the theory of dynamical systems. Since this matter is only distantly related to our main goal of outlining integration theory, I just give a couple of examples.

2.3.17 Definition. *Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f : X \longrightarrow X$ be measurable. The function f is said to be a* ***measure-preserving transformation*** *if $\mu(f^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$.*

You might think that the definition should be: $\mu(f(A)) = \mu(A)$ for all $A \in \mathcal{B}$. It is not, because measurability of f does not imply that $f(A) \in \mathcal{B}$ when $A \in \mathcal{B}$. In many examples, however, f is one-to-one, f^{-1} is measurable *and* measure-preserving, and then $\mu(f(A)) = \mu(A)$.

2.3.18 Example. Let X be the unit circle

$$\{(\cos\theta, \sin\theta) \mid 0 \leq \theta < 2\pi\},$$

with $\mathcal{B}$ = the Borel σ-algebra, and μ = the usual arc length. Let $\alpha \in \mathbf{R}$. Define $f_\alpha : X \longrightarrow X$ by

$$f_\alpha : (\cos\theta, \sin\theta) \longmapsto (\cos(\theta + \alpha), \sin(\theta + \alpha)).$$

This is a measure-preserving transformation. ∎

2.3.19 Example. Let $X = \mathbf{R}^2$, with coordinates (q, p). Again, take $\mathcal{B}$ to be the Borel σ-algebra, and let μ be area. Let H be a continuously differentiable function on $\mathbf{R}^2$. In mechanics, one studies the *Hamiltonian differential equation*

$$\dot{q}(t) = \frac{\partial H(q(t), p(t))}{\partial p}, \quad \dot{p}(t) = -\frac{\partial H(q(t), p(t))}{\partial q}. \tag{2.3.20}$$

For example, with

$$H(q, p) = \frac{1}{2m}p^2 + \frac{k}{2}q^2,$$

the system (2.3.20) becomes

$$\dot{q} = p/m, \quad \dot{p} = -kq, \tag{2.3.21}$$

or

$$m\ddot{q} + kq = 0,$$

which is the equation describing the small oscillations of a mass m attached to a spring with spring constant ("stiffness") k (Exercise I.2.4.34).

Let $\phi_t(q, p)$ be the map that takes the initial condition $q(0) = q, p(0) = p$ to the value $(q(t), p(t))$ of the solution of (2.3.20) at time t. For the system (2.3.21), for example, the map ϕ_t is given by

$$\phi_t(q,p) = \left(q\cos\sqrt{\frac{k}{m}}t + \frac{p}{\sqrt{km}}\sin\sqrt{\frac{k}{m}}t,\ -q\sqrt{km}\sin\sqrt{\frac{k}{m}}t + p\cos\sqrt{\frac{k}{m}}t\right). \tag{2.3.22}$$

For each t, the function $\phi_t : \mathbf{R}^2 \longrightarrow \mathbf{R}^2$ is measurable (Exercise 2.3.13) and measure preserving (see Exercise 2.3.23 below). This result is *Liouville's theorem*. I will not give the general proof. ∎

2.3.23 Exercise*. Show that ϕ_t is measure preserving for each t. (Hint: Compute the Jacobian determinant

$$\frac{\partial\phi_t(q,p)}{\partial(q,p)}.$$

Then change variables in the double integral to deduce that

$$\iint_D dq(t)\,dp(t) = \iint_{\phi_t^{-1}(D)} dq\,dp$$

for open sets D with smooth boundary. This shows that $\mu(\phi^{-1}(D)) = \mu(D)$ for such open sets. The fact that $\mu(\phi^{-1}(A)) = \mu(A)$ for all Borel sets then follows most easily from general measure theory stuff; perhaps you should not worry about that part.) ∎

Subsection 2.4 The Lebesgue Integral

The picture of the Lebesgue integral of a function on $[0,1]$ as being defined by approximations built from a partition of the y-axis is correct (see Exercise 2.5.45), but it is somewhat inconvenient when one tries to prove theorems. A more streamlined approach is outlined in this subsection. The method works in the general setting of measurable extended real-valued functions on a measure space; one gains nothing by restricting attention to functions of a real variable.[12] I present only the bare bones of the construction, followed by a short discussion. For proofs and more details, see a text on Real Analysis. (I recommend *Real Variables*, by A. Torchinsky).

The integral will be defined for measurable functions of increasing complexity. A measure space $(X, \mathcal{B}, \mu)$ is fixed throughout. In the examples involving functions defined on $\mathbf{R}$, the measure is the one discussed (but not rigorously constructed) in Section I.5.2.

Step 1: Nonnegative simple functions.

2.4.1 Definition. *Let $A_1, \ldots, A_n \subset X$ be pairwise disjoint measurable sets. A function*

$$\phi(x) = \sum_{j=1}^{n} \alpha_j c_{A_j}(x)$$

with $\alpha_j \geq 0$ is called a ***nonnegative simple function.*** *(The c_{A_j} are the characteristic functions of the sets A_j).*

The integral over X of a nonnegative simple function ϕ is defined to be

$$\int_X \phi \, d\mu = \sum_{j=1}^{n} \alpha_j \mu(A_j). \tag{2.4.2}$$

2.4.3 Example. The Dirichlet function (2.2.11) is a nonnegative simple function,

$$D(x) = 1 \cdot c_{\text{rationals}}(x) + 0 \cdot c_{\text{irrationals}}(x).$$

Its integral over $X = [0,1]$ is, by definition,

$$\int_{[0,1]} D \, dx = 1 \times \mu(\text{rationals}) + 0 \times \mu(\text{irrationals}) = 1 \times 0 + 0 \times 1 = 0.$$

---

[12] The term *Lebesgue integral* is usually reserved for the integral defined for Lebesgue measure on $\mathbf{R}^n$. The integral constructed for a general measure space is called *integral*. This is about as imaginative as naming your canine pet "Dog", but then mathematicians' literary efforts are said to be limited to stories that begin "Sally was a girl such that . . .".

Step 2: Nonnegative functions.

Next, let $f : X \to [0, \infty]$ be a nonnegative (extended real-valued) measurable function. The integral of f over X is defined to be

$$\int_X f\, d\mu = \sup_{\substack{0 \le \phi \le f \\ \phi \text{ simple}}} \int_X \phi\, d\mu. \tag{2.4.4}$$

2.4.5 Example. Let f be a continuous, nonnegative function on $[0, 1]$. Each partition of $[0, 1]$ defines a finite set of inscribed rectangles, which may be thought of as the graph of a nonnegative simple function. The supremum in (2.4.4) is taken over a set that includes all these inscribed rectangular approximations. Thus, the definition (2.4.4) generalizes the Riemann approximation of integrals by simple areas. The generalization consists in allowing rectangles whose base is a (possibly ugly) measurable set, rather than an interval. ∎

Step 3: General functions.

Finally, for a measurable function $f : X \longrightarrow [-\infty, \infty]$, introduce its positive and negative parts:

$$f_+(x) = \begin{cases} f(x), & \text{if } f(x) \ge 0, \\ 0, & \text{if } f(x) < 0; \end{cases} \qquad f_-(x) = \begin{cases} 0, & \text{if } f(x) > 0, \\ -f(x), & \text{if } f(x) \le 0. \end{cases}$$

Then $f_+, f_- \ge 0$ and $f = f_+ - f_-$. The integral of f over X is defined to be

$$\int_X f\, d\mu = \int_X f_+\, d\mu - \int_X f_-\, d\mu \tag{2.4.6}$$

provided at least one of the two integrals on the right side is finite. This proviso is important enough to be stated separately:

2.4.7 Undefinedness. *If $\int_X f_+\, d\mu = \int_X f_-\, d\mu = \infty$, the integral of f is considered to be undefined.*

Conventions. I usually write integrals as

$$\int_X f\, d\mu.$$

Integrals on the real line will be written in the standard way,

$$\int_a^b f(x)\, dx.$$

When it is important to point out that an integral is interpreted in the Riemann sense, the symbol "(R)" will precede it. Sometimes it may be convenient to write simply

$$\int f,$$

since the X and $d\mu$ are often just decoration, anyway. ∎

2.4.8 Example. It is useful to go through the definitions in a very simple setting. Let $X = \mathbf{Z}_{\geq 0}$, the set of nonnegative integers. Let $\mathcal{B}$ be the σ-algebra of all subsets of $\mathbf{Z}_{\geq 0}$, and let μ be the counting measure,

$$\mu(A) = \text{ number of points in } A.$$

(This is $+\infty$ when A is an infinite set.)

First, what are characteristic functions? If $A = \{m\}$ is the set containing the single point m, then

$$c_A(n) = \begin{cases} 1, & \text{if } n = m, \\ 0, & \text{otherwise.} \end{cases}$$

One should think of c_A as a sequence $(0, 0, \ldots, 1, 0, \ldots)$ with all zeros but for a single 1 in the m^{th} place. Characteristic functions of larger sets have more 1's.

Next, consider some typical nonnegative simple functions. If each A_j is a singleton set, $A_j = \{n_j\}$, the simple function

$$\phi = \sum_{j=1}^{n} \alpha_j c_{A_j} \tag{2.4.9}$$

is a sequence which has the number α_j as n_j^{th} entry, and zeros elsewhere. When the sets A_j are more general (perhaps even infinite), this simple function is a sequence with finitely many distinct entries. The integral

$$\int \phi = \sum_{j=1}^{n} \alpha_j \times \text{ number of points in } A_j$$

is just the ordinary infinite series,

$$\int \phi = \sum_{j=0}^{\infty} \phi(j).$$

It may diverge, but since $\phi(j) \geq 0$, it can only diverge to $+\infty$.

Now turn to the general nonnegative function f on X, which is nothing but a sequence $\{f(j)\}$ of nonnegative numbers. For each $n \geq 1$, define a simple function

ϕ_n: it is the sequence that agrees with $\{f(j)\}$ for indices $1, \dots, n$ and is zero for $j > n$. Clearly, $0 \le \phi_n \le f$. It is now easy to show that

$$\sup_{\substack{0\le\phi\le f \\ \phi \text{ simple}}} \int \phi = \sup_n \int \phi_n = \lim_{n\to\infty} \sum_{j=1}^{n} f(j) = \sum_{j=1}^{\infty} f(j). \tag{2.4.10}$$

Finally, for a general function f, i.e. for a general sequence, we define

$$\int f = \int f_+ - \int f_- = \sum_{f(j)>0} f(j) - \sum_{f(j)<0} (-f(j)).$$

It is required that one of the two series on the right be finite. If this is not the case, the integral remains undefined. For instance, the alternating harmonic sequence, $f(j) = (-1)^j/j$, does not have a sum (="integral") in the Lebesgue measure theoretic sense, since

$$\sum f_+(j) = \sum_{j=1}^{\infty} \frac{1}{2j} \text{ and } \sum f_-(j) = \sum_{j=1}^{\infty} \frac{1}{2j-1}$$

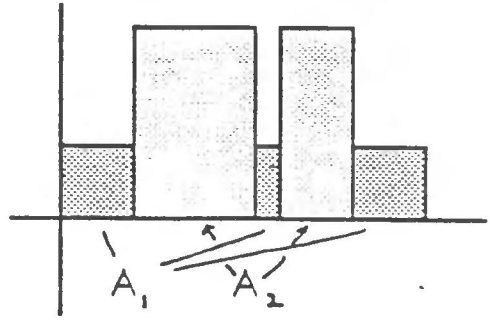
are both $= +\infty$.

The Lebesgue integral, applied to infinite series, is not meant to deal with series that converge conditionally but diverge absolutely. We will see in Example 2.4.17 below that improper Riemann integrals pose similar problems for Lebesgue integration on the real line. ∎

2.4.11 Exercise. Prove (2.4.10) in the preceding example. ∎

Discussion. Since the new definition of the integral is so important, I indulge in a little more discussion, drawing attention to the key ideas and trying to explain the possibly strange notion of "new integral".

Let us start over again with the nonnegative simple functions. The "areas" under their graphs are the building blocks with which more complicated integrals are approximated:



The figure is (of necessity) drawn for $X = \mathbf{R}$, but the measure space X could equally well be something much more complicated, like the Brownian motion sample space. This is already a significant increase in generality over the Riemann integral: without measure theory, one cannot partition a general set X as one does an interval in $\mathbf{R}$, and it is only the Lebesgue integral (and its generalizations) that make integration a universal tool in probability theory.

In Definition 2.4.1 of simple function, the sets A_j need not have finite measure. If $\mu(A_j) = +\infty$ for some A_j, the value of the integral (2.4.2) is $\alpha_j \times \infty$. By convention, this is $+\infty$ when $\alpha_j > 0$, and zero when $\alpha_j = 0$. The restriction $\alpha_j \geq 0$ in (2.4.1) is imposed to exclude functions like

$$1 \cdot c_{(-\infty,0]}(x) - 1 \cdot c_{(0,\infty)}(x).$$

The integral of this simple function would have to be $\infty - \infty$; as was emphasized already, this quantity is considered to be *undefined*.

The intermediate step 2 is perhaps the most clever part of the construction. The best simple function approximants to the "area under the graph of f" are got by partitioning the range of f as described in Subsection 2:

$$\phi(x) = \sum_{j=0}^{n} y_j \, c_{\{x \mid y_j \leq f(x) < y_{j+1}\}}.$$

But, rather than worrying about such partitions, their refinements, the possible unboundedness of f, and such matters, the definition simply admits *all* simple functions whose graphs fit under the graph of f.

You might ask: how can one decide whether or not a definition like (2.4.4) does its approximating in the correct way? After all, while we are still in the process of defining the integral, how do we judge the validity of an approximation to an object that hasn't been defined yet?

First of all, to check that we are on the right track, we might try to prove the following consistency result: if X is a finite interval $[a, b]$, with the usual measure (=length), and if $f \geq 0$ is a continuous function, then the integral defined in (2.4.4) has the same value as the Riemnann integral of f. This is indeed the case, as suggested in Example 2.4.5, but the proof requires tools that have not been developed.

Second, there are certain manipulations we always do with Riemann integrals, and we expect them to be valid in any theory purporting to define integrals. One can prove the following results as consequences of Definition 2.4.9:

2.4.12 Proposition. *Let f, g be nonnegative measurable functions on X, and let a, b be non-negative constants. Then*

$$\text{i)} \quad \int_X (af + bg)\, d\mu = a \int_X f\, d\mu + b \int_X g\, d\mu;$$

ii) $\int_X f\,d\mu \geq 0;$

iii) $\int_X f\,d\mu = 0$ iff $f(x) = 0$ for almost every $x \in X$.

Lebesgue's theory is far from the last word in the study of integrals; throughout the first half of this century, there were numerous extensions of integration theory to more general settings. In all generalizations, two minimal criteria must be satisfied: when the new theory claims to cover a familiar situation, such as integrals over $[a, b]$ of a continuous f, the "new" integral must reduce to the familiar integral, and the linearity and positivity properties i) and ii) of Proposition 2.4.12 should hold.

The final step, Definition 2.4.6, is straightforward, but the proviso that one of $\int_X f_\pm\,d\mu$ be finite is an extremely important wrinkle. *The only time when the integral of a measurable function is not considered to make sense is when $\int f_\pm$ are both infinite.* One of the integrals $\int f_+$ and $\int f_-$ may be $+\infty$ —some of the theorems to be discussed in the next subsection cover this case—, but $\infty - \infty$ is not permitted.

The integral of general measurable functions obeys all the rules we have taken for granted in calculus. For emphasis, I repeat Proposition 2.4.12, this time without requiring positivity:

2.4.13 Proposition. *Let f, g be measurable functions on X, and let a, b be constants. Suppose further that $\int_X |f|\,d\mu < \infty$ and $\int_X |g|\,d\mu < \infty$. Then*

i) $\int_X (af + bg)\,d\mu = a\int_X f\,d\mu + b\int_X g\,d\mu;$

ii) $\int_X f\,d\mu \geq 0$ if $f \geq 0;$

iii) $\int_X |f|\,d\mu = 0$ iff $f(x) = 0$ for almost every $x \in X$.

Comment on the proof: The argument relies directly on the analogous Proposition 2.4.12, which dealt with nonnegative functions, and indirectly, via this proposition, on properties of nonnegative simple functions. The one new item is the assumption

$$\int_X |f|\,d\mu < \infty$$

(the same restriction is placed on g). These hypotheses on f and g ensure that the integral of $af + bg$ cannot be $\infty - \infty$, so that assertion i) of the proposition makes sense.

In fact, one can relax these conditions a little—see Exercise 2.4.21. ∎

2.4.14 Remark. There are many manipulations with integrals we learn in calculus; most of them need to be re-examined when one starts to integrate measurable, rather than contiuous or differentiable, functions. Take the change of variable formula: if $x = g(u)$, then

$$\int_a^b f(x)\,dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(u))g'(u)\,du.$$

If f is measurable and the integral of f is defined, does this also hold for the composition $f \circ g$? If not, the integral on the right might not make sense.

The simplest change of variables is translation, $x = u + k$ for a constant k.

2.4.15 Proposition (Translation Invariance). *Suppose that f is measurable on $\mathbf{R}^n$, and that its integral $\int_{\mathbf{R}^n} f\,d\mu$ (μ=ususal area) is defined. Define $f^k(x) = f(x + k)$. Then f^k is measurable and has an integral, and*

$$\int_{\mathbf{R}^n} f\,d\mu = \int_{\mathbf{R}^n} f^k\,d\mu. \tag{2.4.16}$$

Comments on the proof. Note that this kind of result cannot be stated for a general measure space, because there may not be a concept of addition defined for points of X. On $\mathbf{R}^n$, however, the property is quite obvious. Start with the characteristic function c_I of an interval $I = [a, b]$. Since the length of $[a, b]$ equals the length of $[a + k, b + k]$, we have

$$\int_{-\infty}^{\infty} c_{[a,b]}\,d\mu = \int_{-\infty}^{\infty} c_{[a+k,b+k]}\,d\mu.$$

Property (2.4.16) then follows for nonnegative simple functions by Definition 2.4.2, for nonnegative functions by Definition 2.4.4, and for general functions by (2.4.6). ∎

2.4.17 Examples

A. It is possible for the improper Riemann integral to converge when the Lebesgue integral is not defined.

Consider the function $f(x) = (\sin x)/x$, defined for $x \in \mathbf{R}$. According to Exercise I.4.3.35,

$$\lim_{N\to\infty} \int_{-N}^{N} \frac{\sin x}{x}\,dx = \pi.$$

This limit exists because positive area is constantly cancelled by negative area before either has a chance to build up. According to the Lebesgue definition

(2.4.6), however, we should first take the positive and negative parts of f, and find each integral separately. Now, it is easy to see that

$$\int_{\mathbf{R}} \left(\frac{\sin x}{x}\right)_+ dx = \infty.$$

Indeed,

$$\left(\frac{\sin x}{x}\right)_+ = \begin{cases} (\sin x)/x, & \text{if } x \in [2n\pi, (2n+1)\pi],\ n \in \mathbf{Z}, \\ 0, & \text{otherwise.} \end{cases}$$

The area under the bumps of this function is estimated as

$$\int_{2n\pi}^{(2n+1)\pi} \frac{\sin x}{x}\, dx \geq \frac{1}{(2n+1)\pi} \int_{2n\pi}^{(2n+1)\pi} \sin x\, dx = \frac{2}{(2n+1)\pi}.$$

You can now approximate the areas under ever more bumps by rectangles (simple functions) whose sum diverges like the harmonic series. The details are left as exercise.

Thus, the positive and negative parts both have infinite integral, and the Lebesgue integral of $(\sin x)/x$ remains undefined.

B. In other examples, such as

$$\int_0^\infty e^{-x}\, dx,$$

the improper Riemann integral converges for a different reason: total area is finite without any cancellations. The Lebesgue integral then also exists, and it has the same value. Indeed, one can prove that if f has an *absolutely convergent improper Riemann integral*, meaning that

$$\lim_{N\to\infty} (R) \int_{-N}^{N} |f(x)|\, dx < \infty.$$

then the Lebesgue integral of f is finite, and

$$\lim_{N\to\infty} (R) \int_{-N}^{N} f(x)\, dx = \int_{\mathbf{R}} f\, dx. \blacksquare$$

- **2.4.18 Exercise.** Prove Proposition (2.4.15) for nonnegative simple functions. ∎
- **2.4.19 Exercise.** Let $\phi \in C_0^\infty(\mathbf{R})$. Show that the Lebesgue integral of $\phi(x)/x$ may not be defined. The Cauchy principal value, however, exists: it, like the improper Riemann integral, is a regularization procedure that may apply when Lebesgue integration fails. ∎

• **2.4.20 Exercise.** Suppose that f is measurable, $|f| \leq g$, and g is integrable. Prove that f is integrable. ∎

• **2.4.21 Exercise.** Suppose that the integrals of f and g are defined (but possibly infinite). Is the integral of $f + g$ necessarily defined? Suppose in addition that the integral of g is finite. Is the integral of $f + g$ defined? ∎

• **2.4.22 Exercise.** Prove assertions ii) and iii) of Proposition 2.4.12. ∎

• **2.4.23 Exercise.** Prove or disprove: if

$$\int_{\mathbf{R}} |f(x)|\, dx < \infty,$$

then $\lim_{|x|\to\infty} f(x) = 0$. ∎

2.4.24 Exercise. Let ψ be a nonnegative simple function on a measure space $(X, \mathcal{B}, \mu)$. Show that

$$\nu(E) = \int \psi c_E \, d\mu$$

is a measure on the measurable space $(X, \mathcal{B})$. ∎

2.4.25 Exercise. Let f_Z be the function in Example 2.2.10. Evaluate

$$\int_0^\infty f_Z(x)\, dx. \text{ ∎}$$

Subsection 2.5 Convergence Theorems

Measure theory and the Lebesgue integral on R are important because of the coherence they bring to the problems of interchange of limit and integral or limit and series. The hypotheses imposed in the propositions of Subsection 1.5 are so restrictive that they may not hold in practice, or may be difficult to verify. The Lebesgue integral is much better than the Riemann integral as far as limiting processes are concerned. There are three big theorems: Fatou's Lemma, the Monotone Convergence Theorem, and the Dominated Convergence Theorem (also called the Lebesgue Dominated Convergence Theorem). I will discuss each of them in turn. Emphasis is placed on the last two theorems, because they are used extensively in analysis.

A measure space $(X, \mathcal{B}, \mu)$ is fixed throughout. Functions f, f_n, etc., are defined on X. Again, the measure on $\mathbf{R}^n$ is ordinary length (Lebesgue measure).

2.5.1 Terminology. *A non-negative function f is called* ***integrable*** *if it is measurable and*

$$\int_X f\, d\mu < \infty.$$

A function f is called ***integrable*** *if it is measurable and both f_+ and f_- are integrable. Equivalently, $|f|$ is integrable.*

Remark. This could be confusing! When $\int_X f\,d\mu = \pm\infty$, the integral is still defined, but f is not called "integrable." I didn't think up this strange distinction, but that's how people write. It is usually a good idea to check authors' conventions before reading text books. ∎

Fatou's Lemma

For the statement of Fatou's Lemma, we need to introduce the lim inf (*"limit inferior"*) of a sequence of real numbers.

2.5.2 Definition. *Let $\{y_m\}$ be a sequence of real numbers. If every neighborhood of a point a contains infinitely many of the $\{y_m\}$, then a is said to be a* ***limit point*** *of the sequence. If the sequence has a lower bound, then*

$$\liminf y_m$$

is defined to be the leftmost limit point of $\{y_m\}$ if such a point exists. If $\{y_m\}$ is bounded below, but no finite limit point exists, then $\lim y_m = +\infty$ and one defines $\liminf y_m = +\infty$. Finally, if $\{y_m\}$ is not bounded below, there is a subsequence $y_{m_k} \to -\infty$ and one sets $\liminf y_m = -\infty$. In particular, if $\{y_m\}$ has a finite limit y, then

$$\liminf y_m = \lim y_m = y.$$

If $\{f_n\}$ is a sequence of functions, the notation $\liminf f_n$ denotes the function defined by $f(x) = \liminf f_n(x), x \in X$.

• **2.5.3 Exercise.** a) Suppose that the sequence $\{y_m\}$ is bounded below and has a finite accumulation point. Show that there is a smallest accumulation point.

b) Let the sequence $\{y_1, y_2, \ldots\}$ be bounded below. Define another sequence by $z_k = \inf\{y_k, y_{k+1}, \ldots\}$. Show that the z_k are nondecreasing, that $\lim_{k\to\infty} z_k$ exists or is $+\infty$, and that it is equal to $\liminf y_n$. This explains the notation "lim inf". ∎

• **2.5.4 Exercise.** Let $\{f_n\}$ be a sequence of nonnegative extended real-valued functions. Set

$$g_k(x) = \inf_{n\geq k} f_n(x).$$

Show that

a) $\lim_{k\to\infty} g_k(x) = \liminf f_n(x)$;

b) if the f_n are measurable, then $\liminf f_n$ is also measurable.

2.5.5 Fatou's Lemma. *Let $\{f_n\}$ be a sequence of non-negative measurable functions. Then*

$$\int_X \liminf f_n \, d\mu \leq \liminf \int_X f_n \, d\mu.$$

It is *always* true that the limit of a sequence of measurable functions is measurable (Proposition 2.3.5). It is *not* always true that

$$\lim \int f_n = \int \lim f_n,$$

i.e. limit and integral cannot always be interchanged. That is just a fact of life. Fatou's Lemma gives information about what can happen, at least for non-negative f_n.

2.5.6 Example. Let $f_n(x) = nx^{n-1}$, $x \in [0,1]$. One has

$$\int_0^1 f_n(x)\,dx = 1,$$

and so

$$\lim_n \int_0^1 f_n(x)\,dx = 1.$$

On the other hand,

$$\lim_{n\to\infty} f_n(x) \stackrel{\text{def}}{=} f(x) = \begin{cases} 0, & \text{if } 0 \leq x < 1, \\ \infty, & \text{if } x = 1, \end{cases}$$

so that

$$\int_0^1 \lim f_n(x)\,dx = \int_0^1 f(x)\,dx = 0.$$

Hence,

$$0 = \int \lim f_n < \lim \int f_n = 1,$$

consistent with Fatou's Lemma, since in this example both lim inf's are lim's. ∎

• **2.5.7 Exercise.** Let

$$f_n(x) = \begin{cases} 1. & \text{for } x > n, \\ 0, & \text{otherwise.} \end{cases}$$

Check Fatou's Lemma for this sequence. ∎

• **2.5.8 Exercise.** Check Fatou's Lemma for $f_n(x) = x^n$ on $[0,1]$. ∎

2.5.9 Exercise*. Let $f_n \geq 0$ be measurable. Show that Fatou's Lemma implies

$$\int \sum_{n=0}^{\infty} f_n \, d\mu = \sum_{n=0}^{\infty} \int f_n \, d\mu. \blacksquare$$

2.5.10 Exercise*. In all examples so far, $\liminf f_n$ was equal to $\lim f_n$. Here is a sequence where this is not true: $f_n(x) = \sin^2(n\pi x)$ on $[0,1]$. Evaluate both sides of the inequality in Fatou's lemma for this sequence. ∎

2.5.11 Exercise. Given a sequence $\{y_m\} \subset \mathbf{R}$. Define $\limsup y_m$. Define $\limsup f_n$ for a sequence of functions, and show that it is measurable if the f_n are. ∎

The Monotone Convergence Theorem

2.5.12 Monotone Convergence Theorem. *Let $\{f_n\}$ be a sequence of non-negative measurable functions, and suppose that for almost every x, $\{f_n(x)\}$ is a nondecreasing sequence with limit $f(x)$. Some or all values of the limit function f may be $+\infty$. Then*

$$\int_X f \, d\mu = \lim_n \int_X f_n \, d\mu.$$

2.5.13 Examples

A. Let $\{g_j\}$ be a sequence of nonnegative measurable functions. Set

$$f_n(x) = \sum_{j=1}^{n} g_j(x).$$

The f_n are non-negative and form a nondecreasing sequence. Hence

$$\lim_{n\to\infty} f_n(x) = \sum_{j=1}^{\infty} g_j(x) \stackrel{\text{def}}{=} f(x)$$

exists ($f(x)$ may be $+\infty$ for some x). Now,

$$\int_X f \, d\mu = \int_X \sum_{j=1}^{\infty} g_j \, d\mu,$$

while

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \lim_{n\to\infty} \sum_{j=1}^{n} \int_X g_j \, d\mu = \sum_{j=1}^{\infty} \int_X g_j \, d\mu.$$

The Monotone Convergence Theorem now says that *integral and infinite sum may be interchanged in a series of non-negative functions*:

$$\int_X \sum_{j=1}^{\infty} g_j \, d\mu = \sum_{j=1}^{\infty} \int_X g_j \, d\mu. \tag{2.5.14}$$

(Compare Exercise 2.5.9 above.)

B. To illustrate the use of (2.5.14), I will show that

$$\int_0^1 \frac{1}{1+x^{10}}\,dx = 1 - \frac{1}{11} + \frac{1}{21} - \frac{1}{31} + - \dots . \tag{2.5.15}$$

Formally, the computation is straightforward.

$$\frac{1}{1+x^{10}} = 1 - x^{10} + x^{20} - x^{30} + - \dots , \tag{2.5.16}$$

by the formula for a geometric series. Integration of both sides gives (2.5.15). The question is how to justify term-by-term integration of the infinite series. From Example 1.4.30E, we know that power series may be integrated term-by-term on compact subsets of their interval of convergence. It takes more argument than was given in that example to prove that (in some circumstances) one can extend the integral to the endpoint of the interval of convergence. I will instead use Monotone Convergence and part A.

Note that

$$\frac{1}{1+x^{10}} = (1 - x^{10}) + x^{20}(1 - x^{10}) + x^{40}(1 - x^{10}) + \dots , \tag{2.5.17}$$

and each term

$$g_j(x) = x^{20j}(1 - x^{10}), \quad j \geq 0,$$

is ≥ 0 since $0 \leq x \leq 1$. Appearances in (2.5.16) notwithstanding, we are dealing with a series of non-negative functions! By (2.5.14), one may now compute the integral of the sum as the sum of the integrals. ∎

2.5.18 Example. For $x \in [0, \pi]$, $0 \leq \sin x \leq 1$. Therefore, the sequence $(\sin x)^n$ is nonincreasing for each x, while $1 - (\sin x)^n$ is nondecreasing. Furthermore,

$$\lim_n 1 - (\sin x)^n = \begin{cases} 0, & \text{if } x = \pi/2, \\ 1, & \text{otherwise.} \end{cases}$$

The integral of the limit function is 1. It now follows from the Monotone Convergence Theorem that

$$\lim_n \int_0^\pi \left(1 - (\sin x)^n\right) dx = \pi,$$

and then by a simple manipulation with limits that

$$\lim_n \int_0^\pi (\sin x)^n\,dx = 0.$$

This result can be proved by direct evaluation of the integral or by simple estimates, but I wanted to show how easily it follows from the general theory. Actually, the Dominated Convergence Theorem is even better suited to this little problem. ∎

• **2.5.19 Exercise.** Find, with proof, a series expansion for the definite integral

$$\int_0^1 \frac{x^a}{1+x^b}\,dx,$$

where $a, b > 0$. Does the series expansion of the integrand converge uniformly on $[0,1]$? ∎

• **2.5.20 Exercise.** In the setting of Example 2.4.8, where integral = series, use 2.5.14 to show that the order of summation may be interchanged in a double series with non-negative terms:

$$\sum_{n=1}^{\infty}\sum_{j=1}^{\infty} a_{nj} = \sum_{j=1}^{\infty}\sum_{n=1}^{\infty} a_{nj}.$$

This is true even if the series diverge: if either side is $+\infty$, then so is the other. ∎

2.5.21 Exercise. Prove the Monotone Convergence Theorem, assuming Fatou's Lemma and Proposition 2.4.13. (Hint: if $f_n \le f$, then $\int f_n \le \int f$.) ∎

2.5.22 Exercise. Prove Fatou's Lemma, assuming the Monotone Convergence Theorem. (Hint: use Exercise 2.5.4.) ∎

The Dominated Convergence Theorem

2.5.23 Dominated Convergence Theorem. *Let $\{f_n\}$ be a sequence of measurable functions, and suppose that the limit $f_n(x) \to f(x)$ exists almost everywhere. Suppose there is an integrable function g such that*

$$|f_n(x)| \le g(x)$$

for almost every x. Then

$$\int_X f\,d\mu = \lim_n \int_X f_n\,d\mu.$$

2.5.24 Examples

A. Consider again the sequence $f_n(x) = \sin^n x$ from Example 2.5.18. Since

$$\lim_n f_n(x) \stackrel{\text{def}}{=} f(x) = \begin{cases} 1, & \text{if } x = \pi/2, \\ 0, & \text{otherwise,} \end{cases}$$

since f_n is dominated by the function $g(x) \equiv 1$, and since g is integrable,

$$\int_0^\pi 1\, dx = \pi < \infty,$$

the Dominated Convergence Theorem applies and gives

$$\lim_n \int_0^\pi \sin^n x\, dx = \int_0^\pi f(x)\, dx = 0.$$

B. Here is a less obvious limit: find

$$\lim_{n\to\infty} \int_0^\infty \frac{\ln(x+n)}{n} e^{-x}\, dx. \tag{2.5.25}$$

The integrand is called f_n. For each fixed x,

$$\lim_{n\to\infty} \frac{\ln(x+n)}{n} = 0,$$

so $\lim f_n(x) = 0$. As a first guess, one always wants to try

$$\lim \int f_n = \int \lim f_n,$$

which in this example would give the limit 0 in (2.5.25).

Let us try to find a *non-negative* function g satisfying

$$|f_n(x)| \le g(x) \text{ for almost all } x \in [0,\infty) \tag{2.5.26}$$

and

$$\int_0^\infty g(x)\, dx < \infty. \tag{2.5.27}$$

Since e^{-x} is already integrable, we only need to dominate the $\ln(x+n)/n$ part. Now, for $y \ge 1$, one has

$$\ln y \le y.$$

This is shown as follows. Let $h(y) = \ln y - y$. $h'(y) = 1/y - 1 < 0$ for $y > 1$. Thus, h is decreasing, and since $h(1) = 0$, we have $h(y) < 0$ for $y > 1$. Therefore

$$\frac{\ln(x+n)}{n} < \frac{x+n}{n} = 1 + \frac{x}{n} \le 1 + x \le 2x$$

when $x \ge 1$. If $x < 1$, the last inequality can be replaced by < 2. So, if we take

$$g(x) = \begin{cases} 2xe^{-x}, & \text{for } 1 \le x, \\ 2e^{-x}, & \text{for } 0 \le x < 1, \end{cases}$$

then g dominates all f_n and is integrable. Hence the Dominated Convergence Theorem is applicable, and the limit in (2.5.25) is 0.

C. What is the limit of

$$\int_0^1 x^n \sin x \, dx?$$

We observe that $f_n(x) \stackrel{\text{def}}{=} x^n \sin x \to 0$ for $x \in [0.1)$, while the limit for $x = 1$ is $\sin 1$; the integrals *might* therefore tend to zero. Note that the f_n do not converge uniformly on $[0.1]$, so Proposition 1.4.34 does not apply.

Let's check the conjecture in the old-fashioned way. Pick $\epsilon > 0$. Since

$$|x^n \sin x| < 1,$$

we have

$$\left| \int_{1-\epsilon/2}^1 f_n(x) \, dx \right| \leq \int_{1-\epsilon/2}^1 1 \, dx < \epsilon/2.$$

For $0 \leq x \leq 1 - \epsilon/2$,

$$|f_n(x)| \leq x^n \leq (1 - \epsilon/2)^n.$$

If n is large enough, $(1 - \epsilon/2)^{n+1} < \epsilon/2$, and then

$$\left| \int_0^{1-\epsilon/2} f_n(x) \, dx \right| \leq \int_0^{1-\epsilon/2} (1 - \epsilon/2)^n \, dx < \epsilon/2.$$

Not hard, but long. Here is the Dominated Convergence Proof:

$$|x^n \sin x| \leq |x^n| \leq 1,$$

and

$$\int_0^1 1 \, dx < \infty.$$

End of proof. ∎

2.5.28 Example. *The Fourier Transform.* Let $f : \mathbb{R} \to \mathbb{R}$ be integrable. Recall (from (2.5.1)) that this means:

$$\int_{-\infty}^{\infty} |f(x)| \, dx < \infty.$$

We define the Fourier transform $\hat{f}$ by

$$\hat{f}(k) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f(x) e^{ikx} \, dx. \tag{2.5.29}$$

The integrand on the right is complex; one should interpret it as the sum of the two real integrals of the the real and imaginary parts. (For that matter, the function f could also be complex-valued).

Since

$$|\hat{f}(k)| = \left| \int_{-\infty}^{\infty} f(x) e^{ikx}\, dx \right| \le \int_{-\infty}^{\infty} |f(x)|\, dx < \infty,$$

$\hat{f}(k)$ is finite for every $k \in \mathbb{R}$. More is true: $\hat{f}$ is a continuous function of $k \in \mathbb{R}$. To prove this, consider the difference

$$\hat{f}(k+h) - \hat{f}(k) = \int_{-\infty}^{\infty} (e^{ihx} - 1) \cdot e^{ikx} \cdot f(x)\, dx.$$

The integrand on the right is dominated by the integrable function $2|f|$. Apply the Dominated Convergence Theorem: for every sequence $h_n \to 0$, for fixed k and x,

$$(e^{ih_n x} - 1) e^{ikx} f(x) \to 0.$$

and so

$$\lim_{n\to\infty} \hat{f}(k+h_n) - \hat{f}(k) = 0.$$

This amounts to continuity of $\hat{f}$.

Such results play an important role in Fourier analysis: as a rule of thumb, decay of f at $x = \pm\infty$ translates into smoothness of $\hat{f}$, and vice versa. ∎

2.5.30 Example. There are, of course, problems in which the Dominated Convergence Theorem does not apply. Consider again the Dirac delta limit,

$$\lim_{t \downarrow 0} \int_{-\infty}^{\infty} \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t} f(x)\, dx. \tag{2.5.31}$$

(Replace t by $1/n$ to get a sequence, if you wish.) Suppose f is continuous and bounded. As $t \downarrow 0$, the integrand approaches the function

$$g(x) = \begin{cases} 0, & \text{if } x \neq 0, \\ \infty \cdot f(0), & \text{if } x = 0. \end{cases}$$

This g is certainly measurable and integrable—in fact, it is a simple function:

$$g(x) = 0 \cdot c_{\mathbb{R}-\{0\}} + \infty \cdot f(0) \cdot c_{\{0\}}.$$

Its integral is zero. We know, however, that the limit (2.5.31) equals $f(0)$, which need not be zero. Interchange of limit and integral is not valid in this case. And indeed, we cannot satisfy the hypotheses of Dominated Converge or Monotone

Convergence. It is impossible to find an integrable $h(x)$ that dominates the integrand,

$$\frac{1}{\sqrt{4\pi t}} e^{-x^2/4t} |f(x)| \le h(x) \quad \text{for all } x \in \mathbf{R}.$$

When one needs to determine the validity of

$$\lim \int f_n = \int \lim f_n,$$

and Monotone Convergence or Dominated Convergence don't seem to apply, one should at least entertain the possibility that limit and integration cannot be interchanged. ∎

• **2.5.32 Exercise.** Use the Dominated Convergence Theorem to find the limit as $n \to \infty$ of

$$\int_0^n \left(1 + \frac{x}{n}\right)^{-n} dx.$$

You could compute the integral explicitly, but this is just a warmup for the next problem. ∎

• **2.5.33 Exercise.** Find (with proof) the limit as $n \to \infty$ of

$$\int_0^n (1 + \tfrac{x}{n})^{-n} x^{-1/n}\, dx. \blacksquare$$

2.5.34 Exercise*. Prove that

$$\int_0^\infty e^{-x} \ln x\, dx = \lim_{n\to\infty} \int_0^n (1 - x/n)^n \ln x\, dx.$$

Then show that

$$\frac{d}{dx}\bigg|_{x=1} \int_0^\infty e^{-s} s^{x-1}\, ds = \lim_{n\to\infty} \left\{ \ln n - \sum_{j=1}^{n} \frac{1}{j} \right\}. \tag{2.5.35}$$

The left side of (2.3.35) is the derivative of the Gamma function (I.5.3.29), $\Gamma'(1)$, while the right side is the negative of Euler's constant γ (see (I.4.1.1)). The formula

$$\Gamma'(1) = -\gamma$$

is quite useful in the study of the Gamma function. ∎

• **2.5.36 Exercise.** One of the standard convergence tests is the theory of infinite series is the *comparison test*: if there is a sequence $M_n \ge 0$ with $\sum M_n < \infty$, such that $|a_n| \le M_n$ for all $n \ge 0$, then

$$\sum_{n=0}^{\infty} a_n < \infty.$$

Deduce this from the Lebesgue Dominated Convergence Theorem in the setting of Example 2.4.8. ∎

Miscellaneous Exercises

• **2.5.37 Exercise.** Which of the three big theorems apply to the sequence $c_{[n,\infty)}$? To the sequence $c_{[-n,n]}$? The integrals are to be taken over **R**. ∎

• **2.5.38 Exercise.** Find, with proof,

$$\lim_{n\to\infty} \int_0^1 x^n \sin^{-1} x^2 \, dx. \blacksquare$$

• **2.5.39 Exercise.** (A. Torchinksy) Find, with proof,

$$\sum_{n=0}^{\infty} \int_0^{\pi} \left(1 - \sqrt{\sin x}\right)^n \cos x \, dx. \blacksquare$$

• **2.5.40 Exercise.** Suppose that

$$\sum_{n=-\infty}^{\infty} |a_n| < \infty.$$

Show that the function

$$\sum_{n=-\infty}^{\infty} a_n e^{inx}$$

is continuous. Prove it in two ways: using uniformity, and using one of the convergence theorems in this subsection. ∎

• **2.5.41 Exercise.** Define $f_n(x, y) = (x + y)^n$ on the square

$$S \stackrel{\text{def}}{=} \{(x, y) \mid -1/2 \le x, y \le 1/2\}.$$

a) Find the pointwise limit of f_n.
b) Is the convergence uniform?
c) Find, with proof,

$$\lim_{n\to\infty} \int\int_S f_n(x, y) \, dx dy$$

without evaluating any integrals. ∎

2.5.42 Exercise. Let f be the pure jump function defined in Exercise 2.2.18. Find $\int_0^1 f(x) \, dx$. ∎

2.5.43 Exercise. Give an example of a sequence of simple functions ϕ_n on $[0, 1]$ for which

$$\lim_{n\to\infty} \int_0^1 \phi_n(x) \, dx = 0$$

while

$$\lim_{n\to\infty} \phi_n(x)$$

exists for *no* $x \in [0, 1]$. ∎

2.5.44 Exercise. A sequence $\{f_n\}$ of measurable functions is said to converge to a measurable function f *in measure* if

$$\lim_{n\to\infty} \mu\big(\{x \mid |f(x) - f_n(x)| > \epsilon\}\big) = 0$$

for every $\epsilon > 0$. Suppose that $f_n \to f$ in measure on $[0, 1]$. Does it follow that

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = \int_0^1 f(x)\,dx? \ \blacksquare$$

2.5.45 Exercise*. Here is Lebesgue's original definition of the integral of a bounded non-negative function. Let $0 \le f \le M$. Then

$$\int_X f\,d\mu = \lim_{n\to\infty} \sum_{j=0}^{nM} \frac{j}{n}\,\mu\big(\{x \mid \frac{j}{n} \le f(x) < \frac{j+1}{n}\}\big).$$

We already have a definition of $\int_X f\,d\mu$, so the question becomes: does Lebesgue's formula give the same result?

Prove that the limit on the right does indeed equal the integral. Hint: the sum is, by Definition 2.4.2, the integral of the simple function

$$\phi_n \stackrel{\text{def}}{=} \sum_{j=0}^{nM} \frac{j}{n} c_{A_{jn}},$$

where

$$A_{jn} = f^{-1}\big([j/n, (j+1)/n)\big).$$

Now use one of the convergence theorems. ∎

2.5.46 Exercise*. This continues the preceding exercise. Let f be a non-negative measurable function, not necessarily bounded. Define, for $M = 1, 2, 3, \ldots$:

$$f_M(x) = \begin{cases} f(x), & \text{if } 0 \le f(x) \le M, \\ 0, & \text{otherwise.} \end{cases}$$

Show that

$$\int_X f\,d\mu = \lim_{M\to\infty} \int_X f_M\,d\mu. \ \blacksquare$$

Subsection 2.6 The Lebesgue spaces

This section started with a heuristic discussion of the completion of $C([0,1])$ in the L^1-metric. We can now finish the definition of that particular completion, and of the completion of the other L^p spaces as well. We first met L^p spaces in Section I.1. At that time, only piecewise continuous functions were admitted, and to emphasize the provisional nature of all definitions, I put a little circle under the L^p. Definitions 2.6.2 and 2.6.3 below end the suspense: they are the official ones, and white lies will no longer be necessary.

The Lebesgue integral and its generalization to a general measure space were motivated, in Subsection 1, by the problem of completing $C([0,1])$ in the L^1 metric. We could define a new metric space (normed linear, in fact) called $L^1([0,1])$ to be this completion. That space is of fundamental importance in analysis. It would, unfortunately, be very awkward to study the functions that belong to that space by using L^1 Cauchy sequences of continuous functions. The most practical way to understand those functions relies on the Lebesgue integral. Realizing that all the concepts have not been defined in enough detail to satisfy all readers, I nevertheless offer the proper definition.

2.6.1 Definition. *Let $(X, \mathcal{B}, \mu)$ be a measure space. $L^1(X)$ is the linear space of equivalence classes of measurable, integrable functions from X to the extended reals $[-\infty, \infty]$. Two functions f, g are considered to be equivalent if $f(x) = g(x)$ for almost every $x \in X$ (which is the same as $\int_X |f - g|\, d\mu = 0$).*

You should appreciate that this (standard) definition does not mention completions of incomplete metric spaces; in fact, the set X need not carry a topology, and therefore one may not even have a space $C(X)$ of continuous functions to be completed in the L^1 metric. In any case, the picture of the Lebesgue integral as completion of the Riemann integral is hindsight: the old masters certainly knew that they were filling in the holes of an imperfect theory, but the interpretation of their work in the language of metric spaces and Cauchy sequences is a modern contribution, which is considered, nowadays, to "motivate" the introduction of a new kind of integral. This may only be a reflection of our limited experience—Lebesgue could find all his motivation in Fourier analysis and other, very concrete, problems of analysis.

So, we use "completion" to gain a preliminary understanding of measurable functions and their integrals, and then we switch to the established approach, where all constructions and propositions flow from measure theory. It is ironic that the density of continuous functions in $L^1(X)$ (our starting point) then becomes a rather deep theorem, combining topology with our new measure-theoretic orientation.

Our definition of the L^p spaces ($p > 1$) relies on measure rather than completion. One should *think* of $L^p([0,1])$ as the completion of $C([0,1])$ in the L^p metric, because one can then picture the functions in this space quite easily and reliably: they are approximated, in the L^p norm, by continuous functions, but they can have

all sorts of discontinuities, and they can become infinite at certain "p-integrable" rates. One should, however, *work* with the measure-theoretic definition, because it is usually more efficient.

2.6.2 Definition. *Let $(X, \mathcal{B}, \mu)$ be a measure space. For $1 \le p < \infty$, define $L^p(X, d\mu)$ to be the set of equivalence classes of measurable functions for which*

$$\int_X |f|^p \, d\mu < \infty,$$

two functions being equivalent if they differ at most on a set of measure zero.

The space $L^\infty([0,1])$ (or $L^\infty(X)$, more generally) is quite different. It is not the completion of $C([0,1])$ in the L^∞ norm: by Proposition 1.1.23, $C([0,1])$ is already complete. The definition of $L^\infty(X)$ is measure-theoretic any way you look at it.

2.6.3 Definition. *$L^\infty(X, d\mu)$ is a set of equivalence classes of measurable functions with the following property:*

(2.6.4) there is an $M > 0$ such that $|f(x)| \le M$ for a.e. $x \in X$.

*Two functions are **equivalent in L^∞** if they differ at most on a set of measure zero. The L^∞-norm $\|f\|_\infty$ is defined to be the infimum of the upper bounds M in (2.6.4). One also calls $\|f\|_\infty$ the **essential supremum** of f.*

We begin with examples that illustrate this last definition, because $L^\infty(X)$ is, in many ways, the most peculiar one of all these new spaces.

2.6.5 Examples

A. The Dirichlet function (2.2.11) belongs to $L^\infty([0,1])$. Since $|D(x)| = 0$ for all but the rational x's, which constitute a set of measure zero, we have

$$|D(x)| \le 0 \quad \text{almost everywhere.}$$

Hence, $\|D\|_\infty = 0$.

The zero function, $Z(x) \equiv 0$, differs from D only for rational x, and so $D(x) = Z(x)$ almost everywhere. In the space $L^\infty([0,1])$, therefore, these two functions are considered to be the same: they belong to the same equivalence class.

It is important to become used to this concept of equivalence. In practice, one does not think about equivalence classes and their representatives. One simply pays no attention if the values of a function are changed on a set of measure zero. You can take $\sin x$, change its values on the Cantor set in any way you please (make them all $= \infty$, for example), and you will not have changed the element of $L^\infty([0,1])$.

The moral is important: an element of $L^\infty(X)$ is ***not*** a function—it is an ***equivalence class of functions.*** Still, nobody says "suppose $[f]$ is an equivalence class of functions in $L^\infty(X)$"; everybody says "suppose f is a function in $L^\infty(X)$".

B. One can also define $\|f\|_\infty$ as follows:

$$\text{the set } \{x \mid |f(x)| > \|f\|_\infty\} \text{ has measure } 0,$$

while for any $m < \|f\|_\infty$,

$$\text{the set } \{x \mid |f(x)| > m\} \text{ has positive measure .}$$

C. Let $\{f_n\}$ be a sequence in $L^\infty(X)$. It is a Cauchy sequence if, for every $\epsilon > 0$, there is an N_ϵ such that

$$\|f_n - f_m\|_\infty < \epsilon \quad \text{for all } n, m > N_\epsilon.$$

What exactly does this mean, in terms of specific functions and specific x's?

First, each f_n is really an equivalence class, and we pick one representative, again called f_n. Next, $\|f_n - f_m\|_\infty < \epsilon$ means only that $|f_n(x) - f_m(x)| < \epsilon$ for all x *except* those in a set of measure zero. This set depends on n, m and ϵ. Call it $E(n, m, \epsilon)$. If we set

$$E(\epsilon) = \bigcup_{n,m>N_\epsilon} E(n, m, \epsilon),$$

then $E(\epsilon)$, being a countable union of sets of measure zero, still has measure zero. Hence we can say: given $\epsilon > 0$, there is an N_ϵ and a set $E(\epsilon)$ with $\mu(E(\epsilon)) = 0$, such that

$$|f_n(x) - f_m(x)| < \epsilon \quad \text{for all } x \in X - E(\epsilon) \text{ and } n, m > N_\epsilon. \tag{2.6.6}$$

This is not a strong conclusion. It could be that the sets $E(\epsilon)$ depend on ϵ; they might even be mutually disjoint. Uncountably many sets of measure zero can add up to a set of large measure (for example, the union of all $\{r\}$, as r ranges over the real numbers, is the whole real line). Condition 2.6.6 therefore does not force $\{f_n(x)\}$ to be Cauchy for a single x! This possibility is an undesirable side effect of all the measure-theoretic flexibility.

To cure the problem, we must replace the continuous parameter ϵ by a discrete parameter, say $1/k$. If we want to guarantee $\|f_n - f_m\|_\infty < \epsilon$, it is certainly enough to make $\|f_n - f_m\|_\infty < 1/k$, if $1/k < \epsilon$. The Cauchy condition can therefore be phrased in a thoroughly countable way: given an integer $k > 0$, there is an $N_{1/k}$, such that $\|f_n - f_m\|_\infty < 1/k$ for all $n, m > N_{1/k}$. As argued above, this is equivalent to: $|f_n(x) - f_m(x)| < 1/k$ for $x \in X - E(1/k)$

and $n, m > N_{1/k}$. Finally now, set $E = \bigcup_{k=1}^{\infty} E(1/k)$. Then $\mu(E) = 0$, because the union is countable, and we can say:

2.6.7 Proposition. Let $\{f_n\}$ be a Cauchy sequence in $L^\infty(X)$. Then there is a set $E \subset X$ of measure 0 with the following property: for every $\epsilon > 0$, there is an N_ϵ such that $|f_n(x) - f_m(x)| < \epsilon$ for all $x \in X - E$, whenever $n, m > N_\epsilon$.

Such are the baroque embellishments forced on those who would work with equivalence classes of functions rather than with functions. Mathematicians should worry about these details once, understand the resolution, and then turn to important matters. ∎

The freedom to change functions on sets of measure zero poses a new difficulty. Let (X, B, μ) be a measure space. Suppose we have a measurable function $f : X \to \mathbf{R}_{ex}$, $f \in L^1(X)$, and we redefine it to be ∞ on the set E of measure zero ($\mu(E) = 0$). Take a point $x \in E$. Restore the original value of f at that one point. Is the resulting function equivalent to the original f?

A moment's thought shows that the question reduces to this: is $E - \{x\}$ a set of measure zero? If it is *measurable*, it must have measure zero, because

$$A \subseteq B \Longrightarrow \mu(A) \leq \mu(B).$$

But *it may not be measurable!*

2.6.8 Definition. *Let (X, B, μ) be a measure space. It is said to be* **complete**, *if every subset of a set of μ-measure zero is again measurable (and then, necessarily, has measure zero).*

This notion of completeness of a measure space distinguishes the intuitive concept of length defined for Borel subsets of R from Lebesgue measure:

2.6.9 Definition-Theorem. *Let B be the Borel σ-algebra on* R. *The measure on B induced from $\mu([a, b]) = b - a$ is called the* **standard Borel measure on** R. *There is a unique σ-algebra $\tilde{B}$ generated by B and all subsets of sets of μ-measure zero in B. The measure μ extends to $\tilde{B}$ in a natural way, and the extension is called* **Lebesgue measure on** R. *Lebesgue measure on $\mathbf{R}^n$ is defined in a similar way.*

Remark. The practical differences between the standard Borel measure and the Lebesgue measure on R are explained in texts on Real Analysis. Roughly speaking, the Borel measure is "better" inasmuch as it is based on the topology of R, while Lebesgue measure is "better" because it permits more flexible manipulations with sets of measure zero. But sets of Lebesgue measure zero, as defined in I.5.1.13, need not be constructible via σ-algebra operations on open sets. See H. Royden, *Real Analysis*, p.70, Exercise 28.*) ∎

Having defined L^p and *proposed* a norm on that set, we should next prove that the axioms for a normed linear space are satisfied. This works just as in Subsection I.1.4. I state the result, but leave the details of the proof as exercise.

*) - The edition of Royden's text referred to here is about 30 years old!

2.6.10 Proposition. *The sets $L^p(X)$, as defined in (2.6.2) and (2.6.3), are normed linear spaces.*

Comments on the proof: One needs Minkowski's inequality—but the proof given in I.1.3 carries over with only the obvious change from sum to integral. The main point is this: when one works with integrals, the definition of the integral (Riemann versus Lebesgue, say) is often of secondary importance, as long as the familiar manipulations summarized in Proposition 2.4.12 are available. Minkowski's inequality relies on those general rules.

The nondegeneracy condition

$$\|f\|_p = 0 \Longrightarrow f = 0$$

follows from the equivalence-class definition. If $\|f\|_\infty = 0$, then *every* function $\tilde{f}$ in the equivalence class represented by f satisfies

$$\tilde{f}(x) = 0 \quad \text{for almost every } x \in X.$$

In particular, every such $\tilde{f}$ is equivalent to the zero function Z. This, and nothing more, is the sense in which $\|f\|_\infty = 0$ implies $f \equiv 0$: every function $\tilde{f}$ in the equivalence class of f is zero almost everywhere.

The nondegeneracy condition for $1 \leq p < \infty$,

$$\|f\|_p = 0 \Longrightarrow f = 0,$$

is to be interpreted in the same way. When

$$\int_X |f|^p \, d\mu = 0,$$

it follows from Proposition 2.4.13 that $|f(x)|^p = 0$ for almost every $x \in X$.

You are invited to write out a complete proof, if you think it necessary. ∎

Now comes the fundamental result about L^p spaces: they are complete. This would be a tautology had we defined $L^p(X)$ to be the completion of a space of continuous functions. Our point of view, however, is more general. These spaces were described, in Definition 2.6.2, in a purely measure-theoretic way, and their completeness becomes a property to be proved rather than postulated. We prove completeness for L^1, and leave $1 < p < \infty$ (very similar) and $p = \infty$ (easy) as exercises.

2.6.11 Riesz-Fischer Theorem. [13] *For $1 \leq p \leq \infty$, $L^p(X)$ is a complete metric space.*

First, we prove two lemmas that are of independent interest. The first one, Chebyshev's inequality, has a probabilistic interpretation, see Supplement 2.6.34.

---

[13] Frigyes Riesz (1880-1956), Ernst Fischer (1875-1956).

2.6.12 Lemma (Chebyshev's inequality). *Let f be a measurable, extended real-valued function on a measure space $(X, \mathcal{B}, \mu)$. For every $\epsilon > 0$,*

$$\mu(\{x \mid |f(x)| > \epsilon\}) \leq \frac{1}{\epsilon} \int_X |f| \, d\mu.$$

Proof. Set $E = \{x \mid |f(x)| > \epsilon\}$. Then $\epsilon \, c_E(x) \leq |f(x)|$ for all x. Hence

$$\int_X \epsilon \, c_E \, d\mu = \epsilon \, \mu(E) \leq \int_X |f| \, d\mu. \blacksquare$$

2.6.13 Lemma. *Let $f \in L^1(X)$. Then $|f(x)| < \infty$ for almost every x.*

Proof. By Chebyshev's inequality, applied with $\epsilon = n$,

$$\mu(\{x \mid |f(x) > n\}) \leq \frac{1}{n} \int_X |f| \, d\mu \to 0 \quad \text{as } n \to \infty.$$

Hence (use Proposition I.5.2.13),

$$\mu(\{x \mid |f(x)| = \infty\}) = \lim_{n \to \infty} \mu(\{x \mid |f(x)| > n\}) = 0. \blacksquare$$

Proof of the theorem: Let $\{f_n\}$ be Cauchy in $L^1(X)$. The Cauchy condition makes it possible to find indices $n_1 < n_2 < \dots$ for which $\|f_{n_{j+1}} - f_{n_j}\|_1 < 1/2^j$. By Monotone Convergence applied to series (Example 2.5.13A),

$$\int_X \sum_{j=1}^{\infty} |f_{n_{j+1}} - f_{n_j}| = \sum_{j=1}^{\infty} \int_X |f_{n_{j+1}} - f_{n_j}| < \sum_{j=1}^{\infty} \frac{1}{2^j} = 1. \tag{2.6.14}$$

By Lemma 2.6.13, the integrand on the left side of (2.6.14) is finite almost everywhere. Hence, for almost every x,

$$\sum_{j=1}^{\infty} f_{n_{j+1}}(x) - f_{n_j}(x) \stackrel{\text{def}}{=} g(x) \tag{2.6.15}$$

is finite; furthermore, by Proposition 2.3.5 (applied to the partial sums), g is measurable. The series in (2.6.15) is a telescoping series:

$$\lim_{N \to \infty} \Big(\sum_{j=1}^{N} f_{n_{j+1}}(x) - f_{n_j}(x) \Big) + f_{n_1}(x) = g(x) + f_{n_1}(x) \stackrel{\text{def}}{=} f$$

for almost every x. The limit f is again measurable, and finite almost everywhere.

The subsequence $\{f_{n_j}\}$ of $\{f_n\}$ therefore converges *pointwise* almost everywhere.[14] We will show that it also converges in the L^1 metric,

$$\lim_{j\to\infty} \int_X |f_{n_j} - f|\, d\mu = 0.$$

This will follow from the Dominated Convergence Theorem, if we can bound the integrands:

(2.6.16) $$|f_{n_j} - f| \le |f_{n_j}| + |f| \le \text{ an integrable function.}$$

By (2.6.14), the function

$$\tilde{g} \stackrel{\text{def}}{=} \sum_{j=1}^{\infty} |f_{n_{j+1}} - f_{n_j}|$$

belongs to $L^1(X)$. Since

$$f_{n_j} = f_{n_1} + \sum_{k=1}^{j-1} (f_{n_{k+1}} - f_{n_k}),$$

we have

(2.6.17) $$|f_{n_j}| \le |f_{n_1}| + \sum_{k=0}^{j-1} |f_{n_{k+1}} - f_{n_k}| \le |f_{n_1}| + |\tilde{g}|.$$

Also, since $f = g + f_{n_1}$ by definition,

(2.6.18) $$|f| \le |g| + |f_{n_1}| \le |\tilde{g}| + |f_{n_1}|.$$

The bound (2.6.16) follows from (2.6.17) and (2.6.18). In particular, $f \in L^1(X)$.

We now know that the subsequence $\{f_{n_j}\}$ converges to $f \in L^1(X)$. By the triangle inequality,

$$\|f_m - f\|_1 \le \|f_m - f_{n_j}\|_1 + \|f_{n_j} - f\|_1$$

for all m and n_j. The second term will be $< \epsilon/2$ if n_j is large enough. The first term will be $< \epsilon/2$, if both m and n_j are large enough, because $\{f_n\}$ is Cauchy. Hence the whole sequence $\{f_n\}$ also converges to f. ∎

[14] It is *not* necessarily true that the original sequence $\{f_n\}$ converges pointwise almost everywhere; see Exercise 2.5.43.

Complete normed linear spaces, such as $L^p(X)$, are called *Banach spaces*. Some basic properties of Banach spaces are developed in Section 4. The theory of Banach spaces is extensive, and the body of work about $L^p(X)$, particularly when $X \subseteq \mathbf{R}^n$, is gigantic. From this vast reservoir, I choose only one more result with which to close this subsection: continuous functions are dense in $L^p([0,1])$, for $1 \leq p < \infty$.[15] This takes us back to the starting point of the whole discussion. The fact that functions in $L^1([0,1])$, say, can be approximated by continuous functions is now a consequence of the measure-theoretic definition, whereas earlier this was built into the very definition of $L^1([0,1])$ as a completion. Actually, the proof requires tools that have not been (and will not be) developed, but the result is so useful that it must be stated and discussed even if it is not proved.

2.6.19 Proposition. *Let X be $\mathbf{R}^n$, or the closure of an open subset of $\mathbf{R}^n$, and let μ be Lebesgue measure. Then the compactly supported continuous functions on X are dense in $L^p(X)$ for $1 \leq p < \infty$.*

Comment on the proof: The proposition asserts a connection between three objects—the open set of the topology on $\mathbf{R}^n$, the measurable sets generated by them, and the measure on these sets. The key fact, which emerges from the details of the construction of Lebesgue measure, is the interaction between measure and topology. Namely: given a measurable subset E of $\mathbf{R}^n$, for every $\epsilon > 0$ there is an open set $O_\epsilon \supseteq E$ such that $\mu(O_\epsilon - E) < \epsilon$.

Take $p = 1$. The problem of finding a continuous g with compact support that approximates a given $f \in L^1(X)$ to within ϵ,

$$\int_X |f - g| \, d\mu < \epsilon,$$

is reduced, in a series of steps, to the approximation of a very special kind of function by a continuous one. First, with a suitable simple function ϕ, $\|f - \phi\|_1$ can be made as small as desired (by the very definition of the Lebesgue integral as outlined in Subsection 4, or by Exercise 2.5.45). Next, to approximate a simple function $\sum \alpha_j c_{A_j}$ by a continuous function, it suffices to approximate the characteristic functions c_A of a measurable set of finite measure. By the "key fact", that in turn can be accomplished if the characteristic function of a *bounded open* set O is approximated by a continuous function with compact support. This last step can be done by writing O as a countable disjoint union of parallelopipeds P_k, and approximating each c_{P_k} by a piecewise linear continuous function. ∎

The interaction between measure and topology on which the last proposition is based has an important generalization to other measure spaces.

---

[15] For $p = \infty$, see Exercise 2.6.27.

2.6.20 Definition. *Let $(X, \mathcal{T})$ be a topological space, and suppose that X is* **locally compact** *(Exercise 1.3.36): every $x \in X$ has a neighborhood whose closure is compact. Let μ be a measure defined on the Borel σ-algebra (a* **Borel measure***). Then μ is said to be* **regular***, if, for every measurable set E,*

$$\mu(E) = \inf\{\mu(O) \mid O \supseteq E,\ O \text{ open}\} = \sup\{\mu(K) \mid K \subseteq E,\ K \text{ compact}\}. \blacksquare$$

In words: the measure of a set may be approximated from above by the measure of open sets containing it, and from below by the measure of compact subsets.

2.6.21 Proposition. *If μ is a regular Borel measure on X, then continuous functions with compact support are dense in $L^p(X)$, for $1 \leq p < \infty$.*

2.6.22 Example. Wiener measure on the Brownian motion sample space (Subsection I.5.4) turns out to be a regular Borel measure.[16] ∎

2.6.23 Examples

A. ***Continuity in the mean.*** Measurable functions need not be continuous, so it is almost invariably *not* true that

$$\lim_{h \to 0} |f(x+h) - f(x)| = 0.$$

A kind of averaged continuity, called *continuity in the mean*, does hold.

2.6.24 Proposition. Let $1 \leq p < \infty$, and let μ be Lebesgue measure on $\mathbf{R}^n$. For $f \in L^p(\mathbf{R}^n)$ and $h \in \mathbf{R}$, define $f^h(x) = f(x+h)$. Then

$$\lim_{h \to 0} \int_{\mathbf{R}^n} |f^h - f|^p \, d\mu = 0.$$

Proof. Let $\epsilon > 0$ be given. There is a continuous $g \in L^p(\mathbf{R}^n)$, with compact support, for which

$$\|f - g\|_p < \frac{\epsilon}{3}.$$

By translation invariance of Lebesgue measure, Proposition 2.4.15, one also has

$$\|f^h - g^h\|_p < \frac{\epsilon}{3}.$$

By the triangle inequality,

$$\|f^h - f\|_p \leq \|f^h - g^h\|_p + \|g^h - g\|_p + \|g - f\|_p < \frac{2\epsilon}{3} + \|g^h - g\|_p.$$

---

[16] See E. Nelson, *Feynman Integrals and the Schrödinger Equation*, *J. Math. Phys.* **5** (1964), 332-343.

To prove that $\|f^h - g\|_p < \epsilon$, we must be able to make $\|g^h - g\|_p < \epsilon/3$ by choosing $|h|$ small enough. This goes as follows.

Since g has compact support, it vanishes for $|x| >$ some R. On $|x| \le R$, g is uniformly continuous. Thus, there is a $\delta > 0$ such that $|g(x+h) - g(x)| < (\epsilon^p/3^p)\mu(\{|x| \le R\})$ when $|h| < \delta$. Then

$$\|g^h - g\|_p^p = \int_{\mathbf{R}^n} |g^h - g|^p \, d\mu < \int_{\{|x| \le R\}} \frac{\epsilon^p}{3^p \mu(\{|x| \le R\})} \, d\mu = \frac{\epsilon^p}{3^p}.$$

This finishes the proof. ∎

B. *The Riemann-Lebesgue Lemma.* Example 2.5.28 shows that the Fourier transform connects decay at infinity with regularity. We saw that $\hat{f}$ is continuous if f is integrable: decay at infinity of f translates into continuity of its Fourier transform $\hat{f}$. Now, a function $f \in L^1(\mathbf{R})$ need not be continuous, but nevertheless it has a certain kind of smoothness. It is a nicer function than the Dirac delta, for example—the latter is only a generalized function. Formally, the Fourier transform of δ is

$$\hat{\delta}(k) = \int_{-\infty}^{\infty} \delta(x) e^{ikx} \, dx = e^{ikx}\big|_{x=0} = 1.$$

The constant function 1, of course, does not decay as $|k| \to \infty$. The Fourier transform of $f \in L^1(\mathbf{R})$, however, does decay. This was proved for continuously differentiable f in Example I.3.7.18; as one can now see from Lebesgue integration theory, that smoothness assumption is much too severe.

2.6.25 Riemann-Lebesgue Lemma. If $f \in L^1(\mathbf{R})$, then

$$\lim_{|k| \to \infty} \hat{f}(k) = 0.$$

Proof. Pick $\epsilon > 0$. By Proposition 2.6.19, there is a continuous, compactly supported g for which $\|f - g\| < \epsilon/3$. Suppose the support of g is contained in $(-R, R)$. By Example 1.4.22C, the convolution $E_t * g$ of g with the heat equation solution E_t will satisfy

$$\sup_{x \in \mathbf{R}} |g(x) - E_t * g(x)| < \epsilon/5R$$

when t is small enough. Fix such a t.

By Exercise 1.4.45, $E_t * g$ is continuously differentiable in x. Set $E_t * g = g_t$. Take $\varphi \in C_0^\infty(\mathbf{R})$ so that $0 \le \varphi(x) \le 1$ for all x, and

$$\varphi(x) = \begin{cases} 1, & \text{for } x \in \text{ supp } g_t, \\ 0, & \text{for } |x| \ge R. \end{cases}$$

Set $h = \varphi g_t$; it is continuously differentiable, and

$$|h(x) - g(x)| \le |h(x) - g_t(x)| + |g_t(x) - g(x)| < \frac{2\epsilon}{5R}$$

for all x. We then have

$$\begin{aligned} |\hat{f}(k) - \hat{h}(k)| &\le \|f - g\|_1 + \|g - h\|_1 \\ &\le \frac{\epsilon}{3} + \int_{-R}^{R} |g(x) - h(x)|\, dx < \frac{\epsilon}{5R} + 2R \cdot \frac{2\epsilon}{5R} = \epsilon. \end{aligned} \tag{2.6.26}$$

(The L^1 norm can be replaced by an integral over $[-R, R]$ because both h and g vanish for $|x| > R$.)

According to the special case of the Riemann-Lebesgue Lemma proved in Example I.3.7.18, $|\hat{h}(k)| < \epsilon$ when $|k|$ is large enough. It then follows from (2.6.26) that $|\hat{f}(k)| < 2\epsilon$ for such k. Hence $\hat{f}(k) \to 0$ as $|k| \to \infty$. ∎

• **2.6.27 Exercise.** Show that continuous functions with compact support are not dense in $L^\infty(\mathbf{R})$. ∎

• **2.6.28 Exercise.** If $f \in L^1(X)$, then

$$\lim_{R \to \infty} \int_{\{x \mid |f(x)| > R\}} |f|\, d\mu = 0. \text{ ∎}$$

• **2.6.29 Exercise.** Find the Fourier transform of the characteristic function $c_{[-1,1]}$, and use it to check the Riemann-Lebesgue Lemma. ∎

• **2.6.30 Exercise.** Prove that continuously differentiable functions of compact support are dense in $L^1(\mathbf{R})$ (it is not as simple as the "bullet" might lead you to believe). ∎

• **2.6.31 Exercise.** For $f \in L^1(\mathbf{R})$, find

$$\lim_{A \to \infty} \int_{-\infty}^{\infty} \frac{f(x)}{A^2 + x^2}\, dx. \text{ ∎}$$

2.6.32 Exercise. *Riemann-Lebesgue Lemma for Fourier Series.* Suppose that $f \in L^1([0, 2\pi])$, and set

$$c_n = \int_0^{2\pi} f(x) e^{inx}\, dx.$$

Show that $c_n \to 0$ as $|n| \to \infty$. ∎

2.6.33 SUPPLEMENT: More probability theory

Measurable functions and integrals play a fundamental role in probability theory. Let $(\Omega, \mathcal{B}, P)$ be a sample space — a measurable space of "observable" sets, equipped with a probability measure, meaning that $P(\Omega) = 1$. As explained in the remarks after (2.3.1), a measurable function is to be interpreted as an "observable", a function whose values can be determined from the outcome of the "experiment".

This issue —*what can be observed?*— came up earlier, in our discussion of the probabilistic interpretation of σ-algebras. The problem is really very simple. With a stopwatch, we can measure the average speed of a swimmer, but (probably) not the average velocity of a speeding bullet. The experimental setup is encoded in the choice of σ-algebra. In a coin-toss experiment, the σ-algebra generated by "no heads, one head" would be used if we look at two coins after they have landed on the table. The σ-algebra generated by "HH, HT, TH, TT" would be used if we record the result of each of two tosses. The function that is = 1 if there is one head, and = 0 otherwise, would be measurable (observable) in both experiments; the function that is = 1 if the first toss is heads, and = 0 otherwise, would be measurable only in the second.

2.6.34 Definition. *A measurable function on the probability space $(\Omega, \mathcal{B}, P)$ is called a* ***random variable.***

2.6.35 Example. Toss a fair die[17] lots of times. In N tosses, you expect each face —1 dot, 2 dots, ..., 6 dots— to come up about $N/6$ of the time. The average, or *expected*, value showing on the die is therefore declared to be

$$1 \times \frac{N}{6} + \ldots + 6 \times \frac{N}{6} = 3.5N \,. \blacksquare$$

2.6.36 Definition. *Let f be a random variable defined on a probability space $(\Omega, \mathcal{B}, P)$. The integral*

$$\int_\Omega f \, dP$$

is called the ***expected value of f****, or* ***expectation of f****, and is denoted by $E(f)$.*

The collection of random variables with finite expectation is therefore just the space $L^1(\Omega)$ of integrable functions.

2.6.37 Example. If we toss a die many times, we expect it to show, on the average, the value 3.5. One wants to understand the fluctuations of the observed tosses about this average. The most obvious measure of fluctuation in N tosses,

$$(1 - 3.5) \times \frac{N}{6} + \ldots (6 - 3.5) \times \frac{N}{6}$$

is no good: if

$$1 \times \frac{N}{6} + \ldots 6 \times \frac{N}{6}$$

---

[17] one dice

is close to the expected 3.5, that "measure" of fluctuation is close to zero. We should, instead, study the absolute value of the difference between observation and expectation.

For many practical and theoretical reasons, the mean-square difference is considered to be the most basic measure of fluctuation:

$$(1-3.5)^2\frac{N}{6}+\ldots(6-3.5)^2\frac{N}{6}=17.5N,$$

or, taking the root-mean-square (because it acts so much like Euclidean distance, see the next definition)

$$\sqrt{(1-3.5)^2\frac{N}{6}+\ldots(6-3.5)^2\frac{N}{6}}\sim 4.18N.\ \blacksquare$$

2.6.38 Definition. *Let f be a random variable defined on the probability space $(\Omega, \mathcal{B}, P)$. Suppose its expectation $E(f)$ is finite. Define the* ***variance of*** *f to be*

$$\mathrm{Var}(f)=\int_\Omega |f-E(f)|^2\,dP$$

and the ***standard deviation of*** *f to be*

$$\sigma(f)=\sqrt{\mathrm{Var}(f)}.$$

The random variables with finite variance are just the functions in $L^2(\Omega)$: at this level, probability theory is nothing more than measure theory. Once one mentions the word "independence", however, the subjects begin to diverge. Unfortunately, there is no time to pursue these ideas.

We now turn to Chebyshev's inequality, Lemma 2.6.12. It said:

$$\mu\{x\,|\,|f(x)|>\epsilon\}\le\frac{1}{\epsilon}\int_X |f|\,d\mu.$$

Think of X as Ω, the probability sample space, and of μ as the probability measure P. By Hölder's inequality,

$$\begin{aligned}\int_\Omega |f|\,dP&=\int_\Omega |1||f|\,dP\\&\le\sqrt{\int_\Omega 1\,dP}\sqrt{\int_\Omega |f|^2\,dP}\\&=\sqrt{\int_\Omega |f|^2\,dP},\end{aligned}$$

because $\int_\Omega 1\,dP=1$ (from $P(\Omega)=1$). Therefore,

$$P\big(\{x\,|\,|f(x)|>\epsilon\}\big)\le\frac{1}{\epsilon}\int_\Omega |f|^2\,dP.$$

This is a good estimate for small ϵ's, but not for large ones. If you think of ϵ as a large integer N, then the probability that $|f|>N$ decreases only like $1/N$. We will see presently that this probability can be shown to decay like $1/N^2$. A small modification of Chebyshev's inequality (2.6.12) produces this better result.

2.6.39 Proposition. *Let f be a random variable with finite variance.*[18] *Then for every $\epsilon > 0$,*

$$P(\{x \mid |f(x)| > \epsilon\}) \leq \frac{1}{\epsilon^2} \int_\Omega |f|^2 \, dP.$$

Proof. Let $A = \{x \mid |f(x)| > \epsilon\}$. Then

$$\int_\Omega |f|^2 \, dP \geq \int_A |f|^2 \, dP \geq \epsilon^2 \int_A 1 \, dP = \epsilon^2 P(A). \blacksquare$$

2.6.40 Example. This last version of Chebyshev's inequality allows one to estimate the fluctuation of a random variable about its expected value. Suppose that $E(f)$ and $\mathrm{Var}(f)$ are both finite. Then

$$P(\{x \mid |f(x) - E(f)| > \epsilon\}) \leq \frac{1}{\epsilon^2} \mathrm{Var}(f). \tag{2.6.41}$$

Consider the experiment of tossing a fair coin N times. The sample space Ω is the set of all sequences $\omega = (H \text{ or } T, H \text{ or } T, \ldots)$ of N symbols H, T. The probability of any one sequence is (by definition, based on intuition) 2^{-N}. (Thus, the σ-algebra is the set of all subsets of Ω, and the measure P is defined by $P(\{\omega\}) = 2^{-N}$.) Let $f(\omega)$ be the number of H's (heads) in the sequence ω. Then $E(f) = N/2$ and $\mathrm{Var}(f) = N/4$. It follows from (2.6.41) that

$$P(\{\omega \mid |f(\omega) - \frac{N}{2}| > k\}) \leq \frac{N}{4k^2}. \tag{2.6.42}$$

Take $k = \epsilon N$ in (2.6.42). The right side then becomes $1/4\epsilon^2 N$. The probability that the number of heads deviates from the expected number $(N/2)$ by a fraction ϵ of the total number of tosses is therefore very small, especially when N is large. ∎

The axioms of mathematical probability theory are motivated by intuition. We declare the probability of getting "heads" on a toss of a fair coin to be 1/2, because we feel that, when a coin is tossed many times, "heads" should come up about half the time. Having based the axioms on intuition, we hope to be able to *prove*, from those axioms, that the intuitive picture can be recovered and quantified. The preceding example shows how this works in a simple setting. Stronger results follow from more restrictive assumptions, but that must be left for a course on probability theory.

[18] And hence, by Hölder's inequality, with finite expectation.

Subsection 2.7 Fubini's Theorem

So far, we have *defined* the Lebesgue integral, but we have not yet investigated several basic properties familiar from integral calculus. Is

$$\int_a^b f'(x)\,dx = f(b) - f(a)? \tag{2.7.1}$$

And what about multiple integrals? Is

$$\int_0^1 \int_0^y f(x,y)\,dx\,dy = \int_0^1 \int_x^1 f(x,y)\,dy\,dx? \tag{2.7.2}$$

Those are difficult questions. The "Fundamental Theorem of Calulus" 2.7.1 requires many new ideas, and there is not enough time to discuss them. One must generalize the notion of differentiability: for (2.7.1) to make sense as a Lebesgue integral, f' need only be measurable, but on the other hand, it must be possible to assign unambiguous values to $f(b)$ and $f(a)$ (obviously, one should not change the values of f on the measure zero set $\{a, b\}$).

Multiple integrals are easier to understand than the fundamental theorem of calculus. This subsection summarizes the relevant results, and shows how they are applied.

2.7.3 Example. Multiple integrals

$$\int\int\cdots\int F(x_1,\ldots,x_n)\,dx_1 dx_2 \ldots dx_n,$$

are generally evaluated one integral at a time, working from the inside out: first, the dx_1 integral, then the dx_2 integral, and so on. When that is inconvenient, we change the order of integration. This technique is so common that subtleties are often overlooked.

This next example below is very popular — you can find it in any number of texts.

We try to evaluate

$$\int_0^1 \int_0^1 \frac{x^2 - y^2}{(x^2+y^2)^2}\,dx\,dy. \tag{2.7.4}$$

The integrand can be rewritten as

$$\frac{\partial}{\partial y}\frac{y}{x^2+y^2} = -\frac{\partial}{\partial x}\frac{x}{x^2+y^2}. \tag{2.7.5}$$

By evaluating the inner integral in (2.7.4) first, we find

$$\int_0^1 \frac{\partial}{\partial y}\frac{y}{x^2+y^2}\,dy = \begin{cases} 1/(1+x^2), & \text{if } x \neq 0, \\ \infty, & \text{if } x = 0. \end{cases}$$

This gives

$$\int_0^1 \frac{1}{1+x^2}\,dx = \frac{\pi}{4}$$

for (2.7.4).

On the other hand, if we write (2.7.4) as

$$\int_0^1 \int_0^1 \frac{x^2-y^2}{(x^2+y^2)^2}\,dy\,dx, \,,$$

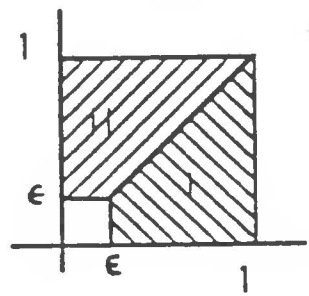
we get $-\pi/4$ (after using (2.7.5)). Something is clearly wrong!

The problem is this: the original integral (2.7.4) is singular. Evaluate it over region I in the figure below:

$$\int_\epsilon^1 \left(\int_0^x \frac{\partial}{\partial y}\frac{y}{x^2+y^2}\,dy \right) dx = -\frac{1}{2}\ln\epsilon.$$

Evaluate it over region II:

$$\int_\epsilon^1 \left(\int_0^y -\frac{\partial}{\partial x}\frac{x}{x^2+y^2}\,dx \right) dy = \frac{1}{2}\ln\epsilon.$$



The sum of the two integrals over the regions I and II, as $\epsilon \to 0$, is $\infty - \infty$. According to Lebesgue rules, this is undefined. By doing the x integral first, and then the y integral (or vice versa), we have *regularized* a divergent expression; the procedure is completely analogous to the evaluation of an improper Riemann integral or of a Cauchy principal value. Unfortunately, the regularization produces different values depending on the order of integration — dx before or after dy. Theorem 2.7.13 (below) about iterated integrals assumes, as is customary in Lebesgue integration theory, that there is no $\infty - \infty$ cancellation. ∎

Some machinery is needed for the discussion of multiple integrals in general measure theory. I record the definitions for sake of completeness, but they are of minor importance to us practical people.

2.7.6 Definition. *A measure space $(X, \mathcal{B}, \mu)$ is said to be σ-finite if X is the union of a countable number of sets of finite measure.*

2.7.7 Example. $\mathbf{R}^n$ can be written as

$$\bigcup_{n=1}^{\infty} \{\mathbf{x} \mid \|\mathbf{x}\|_2 < n\},$$

and is therefore σ-finite. ∎

2.7.8 Definition-Theorem. *If $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ are measure spaces, there is a unique measure space, called the* ***product*** *of X and Y, with the following properties:*

i) The underlying set is $X \times Y$, the set of pairs (x, y) with $x \in X$ and $y \in Y$;

ii) the σ-algebra $\mathcal{B} \times \mathcal{C}$ is generated by the sets $A \times B$, with $A \in \mathcal{B}$ and $B \in \mathcal{C}$;

iii) the measure $(\mu \times \nu)(A \times B)$ is $\mu(A)\nu(B)$.

The theorems that follow relate the integral

$$\int_X \int_Y f(x, y)\, d(\mu(x) \times \nu(y)) \tag{2.7.9}$$

to the iterated integrals

$$\int_X \left(\int_Y f(x, y)\, d\nu(y) \right) d\mu(x) \tag{2.7.10}$$

and

$$\int_Y \left(\int_X f(x, y)\, d\mu(x) \right) d\nu(y). \tag{2.7.11}$$

It is important to understand in what way (2.7.9), (2.7.10), and (2.7.11) are different. You should think of the double integral (2.7.9) as

$$\int\int_{\mathbf{R}^2} f(x, y)\, dx\, dy.$$

It is defined, as was sketched in Section 2, as a limit of approximating sums

$$\sum_{m=1}^{M} f_m\, a(E_m),$$

where f_m is "typical" value of f on the set E_m, whose area is $a(E_m)$. The iterated integral (2.7.10), on the other hand, asks you to compute

$$F(x) \stackrel{\text{def}}{=} \int_{\mathbf{R}} f(x, y)\, dy \tag{2.7.12}$$

for each x, and then to find

$$\int_{\mathbf{R}} F(x)\, dx.$$

(The roles of x and y are interchanged for (2.7.11).) It is implicit in (2.7.10) that the function $y \mapsto f(x, y)$ is measurable (and hence potentially integrable) for each x.

The next theorem summarizes the situation. The restriction to σ-finite measures is necessary; see Torchinsky's *Real Variables*, p.260, for an example.

2.7.13 Fubini's Theorem. [19] *Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be σ-finite measure spaces. Suppose that f is measurable on $(X \times Y, \mathcal{B} \times \mathcal{C})$, and that*

$$\int\int_{X\times Y} |f(x,y)|\, d(\mu \times \nu) < \infty.$$

Then

$$f^y(x) \stackrel{\text{def}}{=} x \mapsto f(x,y)$$

is measurable (with respect to the σ-algebra $\mathcal{B}$) and $\int_X f^y(x)\, d\mu(x)$ exists for ν-almost every y. The function

$$f_x(y) \stackrel{\text{def}}{=} y \mapsto f(x,y)$$

is measurable (with respect to the σ-algebra $\mathcal{C}$ and $\int_Y f_x(y)\, d\nu(y)$ exists for μ-almost every X. One then has equality of double and iterated integrals:

$$\int\int_{X\times Y} f(x,y)\, d(\mu\times\nu) = \int_Y \left(\int_X f^y(x)\, d\mu(x)\right) d\nu(y) = \int_X \left(\int_Y f_x(y)\, d\nu(y)\right) d\mu(x).$$

This is a wonderful theorem, but we (almost always) understand multiple integrals by computing iterated integrals. It is therefore natural to ask: when can one compute

$$\int_X \left(\int_Y f(x,y)\, d\nu(y)\right) d\mu(x),$$

and know that it equals

$$\int_Y \left(\int_X f(x,y)\, d\mu(x)\right) d\nu(y),$$

or

$$\int\int_{X\times Y} f(x,y)\, d(\mu(x) \times \nu(y))?$$

2.7.14 Tonelli's Theorem. [20] *Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be σ-finite measure spaces. Suppose that f is non-negative, extended real-valued, and measurable on $(X \times Y, \mathcal{B} \times \mathcal{C})$. Then*

$$\begin{aligned}\int_X \left(\int_Y f(x,y)\, d\nu(y)\right) d\mu(x) &= \int_Y \left(\int_X f(x,y)\, d\mu(x)\right) d\nu(y) \\ &= \int\int_{X\times Y} f(x,y)\, d(\mu \times \nu).\end{aligned}$$

Equality of these three expressions means: if one of them is $+\infty$, then they are all equal to $+\infty$.

---

[19] Guido Fubini, 1879-1943.

[20] Leonida Tonelli, 1885-1946.

2.7.15 Remark. These two fundamental theorems are generally used in the following way. One shows, somehow, that

$$\int_Y \left(\int_X |f(x,y)|\, d\mu(x) \right) d\nu(y) < \infty. \tag{2.7.16}$$

Tonelli's theorem then implies that

$$\int_X \left(\int_Y |f(x,y)|\, d\nu(y) \right) d\mu(x) < \infty \tag{2.7.17}$$

as well and has the same value as (2.7.16), and that the double integral

$$\int\int_{X\times Y} |f(x,y)|\, d(\mu(x) \times \nu(y))$$

is also finite, and equals (2.7.16) and (2.7.17). Fubini's Theorem now enters: the function f of *two* variables is integrable, and therefore

$$\begin{aligned} &\int\int_{X\times Y} f(x,y)\, d(\mu(x) \times \nu(y)) \\ &= \int_Y \left(\int_X f(x,y)\, d\mu(x) \right) d\nu(y) \\ &= \int_X \left(\int_Y f(x,y)\, d\nu(y) \right) d\mu(x). \blacksquare \end{aligned}$$

2.7.18 Remark. Two of the statements in Fubini's and Tonelli's theorems should be emphasized. Tonelli's theorem asserts the equality of integrals even if they are all infinite (necessarily, $= +\infty$, since the integrand must be nonnegative). Fubini's theorem says that the functions

$$x \longmapsto \int_Y f(x,y)\, d\nu(y) \quad \text{and} \quad y \longmapsto \int_X f(x,y)\, d\mu(x)$$

are measurable and finite almost everywhere (with respect to the measures μ and ν, respectively). This, as we shall see next, implies that the integral representation of certain important functions makes sense. ∎

Convolutions

The convolution of two functions was introduced in 1.4.22C. We saw there (and earlier in Subection 1.4.0) that the integral

$$E_t \star g(x,t) = \int_{-\infty}^{\infty} E(x-y,t) g(y)\, dy \tag{2.7.19}$$

solves the heat equation $u_t = u_{xx}$ with initial condition $u(x,0) = g(x)$. Convolutions occur in many settings, and one defines, quite generally,

$$f \star g(x) = \int_{-\infty}^{\infty} f(x-y)g(y)\,dy. \tag{2.7.20}$$

This is an operation done to functions that makes sense only in special circumstances, because in general measure spaces we don't know how to subtract points (I refer to the $x - y$ in (2.7.20)). The convolution of two functions on $\mathbf{R}^n$ does make sense,

$$f \star g(x) = \int_{\mathbf{R}^n} f(x-y)g(y)\,dy.$$

because vectors in $\mathbf{R}^n$ *can* be subtracted.

The question we take up in this example is the most basic one: for what kinds of functions f, g does the convolution $= f \star g$ exist at all, and to what space does it belong?

Before doing anything, one should show that the function $f(x-y)g(y)$ is measurable in $(x,y) \in \mathbf{R}^{2n}$. If it is not, neither Fubini's theorem nor Tonelli's theorem can get off the ground. The argument is not altogether easy, and is therefore left as exercise (number 2.7.34).

2.7.21 Proposition. *Let $g \in L^1(\mathbf{R}^n)$ ($\mathbf{R}^n$ equipped with Lebesgue measure), and let $f \in L^p(\mathbf{R}^n)$ for some p, $1 \le p \le \infty$. Then $g \star f \in L^p(\mathbf{R}^n)$, and*

$$\|g \star f\|_p \le \|g\|_1 \|f\|_p. \tag{2.7.22}$$

Proof. Suppose that $1 < p < \infty$ (the cases $p = 1$ and $p = \infty$ are left as exercises). Let q be the conjugate exponent: $(1/p) + (1/q) = 1$. Then

$$\begin{aligned} |g \star f(x)| &\le \int_{\mathbf{R}^n} \left(|g(x-y)|\right)^{\frac{1}{q}} \left(|g(x-y)|^{\frac{1}{p}} |f(y)|\right) dy \\ &\le \left[\int_{\mathbf{R}^n} |g(x-y)|\,dy\right]^{1/q} \left[\int_{\mathbf{R}^n} |g(x-y)|\,|f(y)|^p\,dy\right]^{1/p} \\ &\le \|g\|_1^{1/q} \left[\int_{\mathbf{R}^n} |g(x-y)|\,|f(y)|^p\,dy\right]^{1/p}. \end{aligned} \tag{2.7.23}$$

The first inequality is a (standard) trick: it uses $A = A^{1/p}A^{1/q}$, true because $(1/p) + (1/q) = 1$. The second inequality is Hölder. The third one relies on translation invariance, (2.4.15).

Raise the first and last term in (2.7.23) to the power p and integrate over x:

$$\int_{\mathbf{R}^n} |g \star f(x)|^p\,dx \le \|g\|_1^{p/q} \int_{\mathbf{R}^n}\int_{\mathbf{R}^n} |g(x-y)|\,|f(y)|^p\,dy\,dx. \tag{2.7.24}$$

At this point, Fubini-Tonelli enters the picture. The integrand on the right side is nonnegative. We show that one way of computing the iterated integral leads to a finite value:

$$\int_{\mathbf{R}^n} \left(\int_{\mathbf{R}^n} |g(x-y)|\, dx \right) |f(y)|^p\, dy = \|g\|_1 \int_{\mathbf{R}^n} |f(y)|^p\, dy = \|g\|_1 \|f\|_p^p. \tag{2.7.25}$$

Combining (2.7.24) and (2.7.25), we find

$$\|g \star f\|_p^p = \int_{\mathbf{R}^n} |f \star g(x)|^p\, dx \le \|g\|_1^{\frac{p}{q}+1} \|f\|_p^p,$$

or (using $(1/p) + (1/q) = 1$ again).

$$\|g \star f\|_p \le \|g\|_1 \|f\|_p.$$

This was one of the facts that had to be proved. Notice, however, how much else follows from the crucial step (2.7.25). Because the iterated integral in (2.7.25) is finite, the other iterated integral (dy inside, dx outside) is also finite, and so is the double integral, by Tonelli's theorem. Then Fubini's theorem says that the double and iterated integrals are equal *without* any absolute value signs. Furthermore, the function

$$x \mapsto \int_{\mathbf{R}^n} f(x-y) g(y)\, dy$$

is measurable, and is finite for almost every x. That, of course, is just the convolution $f \star g$, and so Fubini's theorem guarantees that $f \star g$ is a well-defined function (up to a set of measure zero). ∎

2.7.26 Example. In (2.5.28), the Fourier transform of a function $f \in L^1(\mathbf{R}^n)$ was defined to be

$$\hat{f}(\xi) = \int_{\mathbf{R}^n} f(x) e^{i\xi \cdot x}\, dx.$$

The Fourier transform of a convolution $f \star g$ (for $f, g \in L^1(\mathbf{R}^n)$) is the product $\hat{f}\hat{g}$ of the Fourier transforms. This is easy to check by purely formal manipulations:

(2.7.27)

$$\begin{aligned} \int_{\mathbf{R}^n} e^{i\xi \cdot x} \int_{\mathbf{R}^n} f(x-y) g(y)\, dy\, dx &= \int_{\mathbf{R}^n} e^{i\xi \cdot (x-y)} f(x-y) \int_{\mathbf{R}^n} e^{i\xi \cdot y} g(y)\, dy\, dx \\ &= \int_{\mathbf{R}^n} e^{i\xi \cdot z} f(z) \int_{\mathbf{R}^n} e^{i\xi \cdot y} g(y)\, dy\, dz \\ &= \hat{f}(\xi)\hat{g}(\xi). \end{aligned}$$

The steps are made rigorous by appeal to Fubini and Tonelli. The iterated integral

$$\int_{\mathbf{R}^n} |e^{i\xi \cdot x}| \left(\int_{\mathbf{R}^n} |f(x-y)|\, |g(y)|\, dy \right) dx \tag{2.7.28}$$

converges, because by (2.7.22), the convolution $f \star g$ belongs to L^1 if f and g do. Once the iterated integral (2.7.28) is known to converge, the other iterated integral and the double integral *with absolute values around the functions* exist, and are equal, according to Tonelli's theorem. Then by Fubini's theorem, the iterated integrals and the double integral converge, and are all equal, *without the absolute values*.

Keeping these facts in mind, one can now justify all the manipulations in (2.7.27). The details are left to the reader. ∎

• **2.7.29 Exercise.** Prove Proposition 2.7.21 when f belongs to $L^1(\mathbf{R}^n)$ or $L^\infty(\mathbf{R}^n)$. ∎

• **2.7.30 Exercise.** Let $f \in L^p(\mathbf{R})$ for some $p \in (1, \infty)$. Show that the convolution $E_t \star f$ from Example 1.4.22C tends to f in the L^p-norm, as $t \downarrow 0$. ∎

• **2.7.31 Exercise.** Find

$$\int_0^1 \int_0^1 \frac{x^2 - y^2}{x^2 + y^2}\, dx\, dy,$$

if the integral exists. Justify all manipulations. ∎

• **2.7.32 Exercise.** If f is Lebesgue integrable on $[0, 1]$, is $g(x, y) = f(x) - f(y)$ Lebesgue integrable over the square $0 \le x, y \le 1$? ∎

2.7.33 Exercise. Let $f \in L^p(\mathbf{R}^n)$ and $g \in L^q(\mathbf{R}^n)$, with $(1/p) + (1/q) = 1$. Show that $f \star g$ is continuous. ∎

2.7.34 Exercise*. Let $f \in L^1(\mathbf{R}^n)$ and $g \in L^p(\mathbf{R}^n)$. Show that $f(x - y)g(y)$ is a measurable function on $\mathbf{R}^{2n}$. (Hint: see Torchinsky, *Real Variables*, p.247.) ∎

Section 3 Arzelà-Ascoli Theorem

In the last section, we investigated the process of completion of a metric space in one very important example. The present section offers another illustration of the basic material in Section 1. We prove the *Arzelà-Ascoli Theorem*,[21] which gives conditions under which a sequence of continuous functions is guaranteed to have a convergent subsequence. This is a *compactness result*: it identifies certain compact subsets of the metric space of continuous functions. Two applications of the theorem, to differential equations and geodesics, are also worked out.

The statement and proof of the Arzelà-Ascoli Theorem require some terminology.

3.0.1 Notation. *The set of continuous functions from a metric space (M, d) to a metric space (M', d') is denoted by $C(M, M')$.*

[21] Cesare Arzelà, 1847-1912; Giulio Ascoli, 1843-1896.

3.0.2 Proposition. *If (M, d) in (3.0.1) is a compact metric space, then*

$$\rho(f,g) = \max_{x \in M} d'(f(x), g(x)) \tag{3.0.3}$$

defines a metric on $C(M, M')$.

Proof. First we prove that the maximum in (3.0.3) exists. To do this, we show that $x \mapsto d'(f(x), g(x))$ is a continuous function on M, which assumes its maximum value by Proposition 1.3.8. The triangle inequality gives

$$d'(f(x), g(x)) \leq d'(f(x), f(y)) + d'(f(y), g(y)) + d'(g(y), g(x))$$

or

$$d'(f(x), g(x)) - d'(f(y), g(y)) \leq d'(f(x), f(y)) + d'(g(y), g(x)). \tag{3.0.4}$$

Interchange x and y:

$$d'(f(y), g(y)) - d'(f(x), g(x)) \leq d'(f(y), f(x)) + d'(g(x), g(y)). \tag{3.0.5}$$

(3.0.4) and (3.0.5) together imply

$$|d'(f(x), g(x)) - d'(f(y), g(y))| \leq d'(f(x), f(y)) + d'(g(y), g(x)). \tag{3.0.6}$$

Given $\epsilon > 0$, there exists a $\delta > 0$ such that $d'(f(x), f(y)) < \epsilon$. $d'(g(x), g(y)) < \epsilon$ whenever $d(x, y) < \delta$. Continuity of the function $d'(f(x), g(x))$ now follows from (3.0.6).

The axioms I.1.0.4 for a metric are left as exercise. ∎

3.0.7 Definition. *Let (X, T) be a topological space, and let (M', d') be a metric space. A family $\mathcal{F}$ of continuous functions from X to M is said to be* **equicontinuous**, *if to every $\epsilon > 0$ and every $x \in X$ there is an open set $O_{\epsilon,x}$ containing x, such that for all $f \in \mathcal{F}$, $d'(f(x), f(y)) < \epsilon$ whenever $y \in O_{\epsilon,x}$.*

The *equi* part is fairly subtle. If $\mathcal{F}$ were just a collection of continuous functions, the neighborhood of x would depend on x, ϵ *and on the function* f. *Equicontinuous* means that the neighborhood is independent of $f \in \mathcal{F}$. All the functions in $\mathcal{F}$ are continuous with the same epsilons and deltas, so to say.

3.0.8 Definition. *A topological space (X, T) is said to be* **separable** *if it contains a countable dense subset.*

(See Exercise I.3.5.39 or the proof of Proposition 1.3.32.)

3.0.9 Example. The rationals are a countable subset of $\mathbf{R}$, and they are dense in the usual topology. Hence, $\mathbf{R}$ is separable. The points with rational coordinates are countable and dense in $\mathbf{R}^n$ and $\mathbf{C}^n$. ∎

3.0.10 Arzelà-Ascoli Theorem. *Let $\{f_n\}$ be an equicontinuous sequence of functions from a separable metric space (M, d) to a complete metric space (M', d'). Suppose that the ranges of all the f_n are contained in a compact subset $K \subset M'$. Then there is a subsequence f_{n_k} which converges to a function $f : M \longrightarrow M'$ for each $x \in M$. Moreover, the sequence converges uniformly on each compact subset of M, and therefore the limit function f is continuous on each compact subset of M.*

The most popular special case of this theorem goes as follows.

3.0.11 Corollary. *Let G be an open subset of $\mathbf{R}^n$, let $\bar{G}$ be its closure, and let $\{f_n\}$ be an equicontinuous sequence of functions from $\bar{G}$ to $\mathbf{R}$ which is uniformly bounded, meaning that $|f_n(x)| \leq$ some R for all n and all $x \in \bar{G}$. Then there exists a subsequence that converges to a continuous function from $\bar{G}$ to $\mathbf{R}$.*

Remarks

A. The limit function whose existence is asserted by the Arzelà - Ascoli theorem is continuous on each compact subset of M. *If* M can be written as a union of an increasing sequence $C_1 \subset C_2 \subset \ldots$ of compact subsets, then the limit function is continuous on all of M. This will be the case, for example, if $M = \mathbf{R}^n$, but there are metric spaces that are *not* unions of compact subsets. The point is not much of a worry in applications.

B. Suppose that the metric space (M, d) is compact. A sequence $\{f_n\}$ of functions (from M to M') that satisfies the hypotheses of the Arzelà-Ascoli Theorem has a subsequence that converges uniformly on the whole compact space M. The limit is a continuous function from M to M'. Therefore, if an equicontinuous family $\mathcal{F}$ with range in a compact subset of M' is a *closed subset of* $C(M, M')$, it is also a *compact subset* (for compact M and complete M'). ∎

Proof. Step 1. The domain M contains a countable dense subset D. We first show that a subsequence of the given $\{f_n\}$ converges on D. Enumerate the points of D as $\{x_1, x_2, \ldots\}$. The set $\{f_n(x_1)\}$ is contained in a compact subset of the range M', and so there is a subsequence, which we denote by $\{f_{1n}(x_1)\}$, that converges. Next, the set $\{f_{1n}(x_2)\}$ is contained in a compact subset of M', and so a subsequence of $\{f_{1n}\}$ converges on the point x_2; call this subsequence $\{f_{2n}\}$. Proceeding in this fashion, we find nested subsequences $\{f_{kn}\}$; the k^{th} one converges on $x_1, \ldots, x_k$.

From the subsequences

$$\{f_{1n}\} \supset \{f_{2n}\} \supset \ldots \supset \{f_{kn}\} \supset \ldots$$

we extract a sub-sub-...-sub-sequence that converges on all of the dense subset D. Namely, the subsequence $\{f_{nn}\}$ converges on all x_j. (This is the Cantor diagonal process, which is used to prove, for example, that the set of real numbers is uncountable; see Exercise I.3.1.8).

Step 2. The subsequence $\{f_{nn}\}$ obtained in Step 1 will henceforth be written as $\{g_n\}$. Being a subset of an equicontinuous family, it is itself equicontinuous. We now show that equicontinuity extends convergence from the dense subset D to all of M.

Let $x \in M$, and let $\epsilon > 0$ be given. There is a neighborhood O_x of x such that when $y \in O_x$ then $d'(g_n(x), g_n(y)) < \epsilon/3$ for all $n = 1, 2, \ldots$ (this by equicontinuity). Since D is dense, there is a $y \in O_x \cap D$, and for this y, $\{g_n(y)\}$ converges. Therefore, that sequence is Cauchy (Proposition 1.1.4), and for $n, m >$ some $N_{\epsilon/3}$, $d'(g_n(y), g_m(y)) < \epsilon/3$. Thus,

$$\begin{aligned} d'(g_n(x), g_m(x)) &\leq d'(g_n(x), g_n(y)) + d'(g_n(y), g_m(y)) + d'(g_m(y), g_m(x)) \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon. \end{aligned}$$

Since, for each $x \in M$, the sequence $\{g_n(x)\}$ is Cauchy, the completeness of M' guarantees the existence of a unique limit which we call $g(x)$. The function g is, at this point in the proof, not known to be continuous.

Step 3. We show that the subsequence $\{g_n\}$ converges uniformly on every compact subset $C \subset M$. This implies, by Proposition 1.4.5, that the limit function g is continuous on each compact subset of M.

Choose $\epsilon > 0$. By equicontinuity, each $x \in C$ has a neighborhood O_x such that $d'(g_n(x), g_n(y)) < \epsilon/3$ for all $n = 1, 2, \ldots$ whenever $y \in O_x$. Let $n \longrightarrow \infty$ to conclude that $d'(g(x), g(y)) \leq \epsilon/3$. Since C is compact, finitely many of these neighborhoods, $O_{x_1}, \ldots, O_{x_K}$, cover C. The sequence $\{g_n(x_i)\}$ converges for each $i = 1, \ldots, M$; choose N so large that $n > N$ implies $d'(g_n(x_i), g(x_i)) < \epsilon/3$. Finally, let $y \in C$. There is an x_i such that $y \in O_{x_i}$ since the O_{x_i} cover C. Then

$$\begin{aligned} d'(g_n(y), g(y)) &\leq d'(g_n(y), g_n(x_i)) + d'(g_n(x_i), g(x_i)) + d'(g(x_i), g(y)) \\ &< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon \end{aligned}$$

whenever $n > N$. Note that N depends on ϵ but not on y; the convergence is indeed uniform. This concludes the proof. ∎

- **3.0.12 Exercise.** Let $\mathcal{F}$ be a family of continuous real-valued function on a metric space (M, d). What is the negation of the statement: "$\mathcal{F}$ is an equicontinuous family"? ∎

- **3.0.13 Exercise.** Is the family $\{x^n\}_{n=0}^{\infty}$ equicontinuous on $(0, 1)$? If yes, what does the Arzelà-Ascoli Theorem say? Does it contradict the non-uniform convergence of this sequence (see Example 1.4.18A)? ∎

- **3.0.14 Exercise.** Let (M, d) be a compact metric space, and let (M', d') be complete. Show that $C(M, M')$ is complete in the metric (3.0.3). ∎

• **3.0.15 Exercise.** Give an example of a sequence of functions that satisfies the hypotheses of the Arzelà-Ascoli Theorem and has at least *two* subsequences that converge uniformly on compact sets. (The more "natural" the example, the better: $f_n(x) \equiv +1$ for n even and $f_n(x) \equiv -1$ for n odd will work — try to be more inventive.) ∎

Subsection 3.1 Solutions of ordinary differential equations

As the first of two applications of the Arzelà - Ascoli theorem, we present Peano's proof[22] of the existence of solutions to the initial value problem for a system of ordinary differential equations.

The setup is as follows. G is a connected open subset of $\mathbf{R}^n$, and $\bar{G}$ is its closure, assumed to be compact. Just think of $\bar{G}$ as a closed sphere or cube. Let f be a continuous function from $\bar{G}$ to $\mathbf{R}^n$; f is a vector $(f_1, \ldots, f_n)$ of real-valued functions. Suppose that $\|f(x)\|_2 < M$ for all $x \in \bar{G}$; in particular, $\max_i \max_{x \in \bar{G}} |f_i(x)| < M$ by (I.1.2.10). Fix a point $x^0 \in G$, and let D be the distance from x^0 to the boundary of G.

Consider the initial-value problem

$$\frac{dx_i(t)}{dt} = f_i(x_1(t), \ldots x_n(t)), \quad x_i(0) = x_i^0, \quad i = 1, \ldots, n,$$

or, more compactly,

$$\frac{dx(t)}{dt} = f(x(t)), \qquad x(0) = x^0. \tag{3.1.1}$$

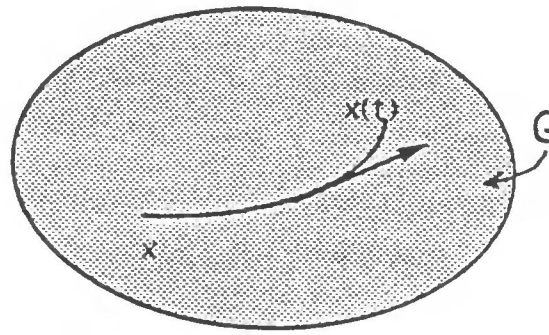


Figure 14. *Solution of the initial value problem (3.1.1)*

3.1.2 Theorem. *There is at least one solution x of (3.1.1). It is a differentiable function of t, and it exists at least for $|t| < D/M$.*

[22] Giuseppe Peano, 1858-1932.

3.1.3 Examples

A. The initial value problem $\dot{x} = \sqrt{x}$, $x(0) = 0$, has infinitely many solutions: $x(t) \equiv 0$, $x(t) = t^2/4$, and more generally

$$x(t) = \begin{cases} 0, & \text{if } t < c \\ (t-c)^2/4, & \text{if } t \geq c \end{cases}$$

for any $c > 0$. So the existence theorem certainly does not imply uniqueness of solutions. To get uniqueness, one must ask that f be smoother than just continuous. Notice in the proof of the theorem that, at a crucial point, the Arzelà - Ascoli theorem is invoked to guarantee the existence of "a convergent subsequence" whose limit is a solution of the initial value problem. There may be many different such subsequences, producing many different solutions.

B. The initial value problem $\dot{x} = x^2, x(0) = 1$ has the (unique) solution $x(t) = 1/(1-t)$, which exists only for $t < 1$. If we let G be the set (-1.1), then the numbers M and D in the theorem are, respectively, 1 and 1, and the solution actually fails to exist for $t \geq D/M = 1$. This shows that the estimate $|t| < D/M$ of the existence interval is the best possible. ∎

Proof. First, recall the Euler approximation method which is presented in most introductory o.d.e. courses.

The solution $x(t)$ of (3.1.1) is approximated, near $t = 0$, by its tangent line: $x(t) \sim x(0) + t\dot{x}(0) = x^0 + tf(x^0)$. The approximation is (we hope) good up to $t = h$ for some (small) h. Set $x^1 = x^0 + hf(x^0)$, and use the differential equation to determine a new linear approximation starting at x^1: $x(t) \sim x^1 + tf(x^1)$, for $h \leq t \leq 2h$. This is probably the worst possible method for getting a numerical solution, but a small modification will suffice to produce an existence proof.

Fix $\epsilon > 0$. We will find an approximate solution with error less than ϵ, in a sense to be made precise later on. Because $\bar{G}$ is compact, the continuous function $f : \bar{G} \longrightarrow \mathbf{R}^n$ is uniformly continuous. Hence there is a $\delta > 0$ such that $||f(x) - f(y)||_2 < \epsilon$ whenever $||x - y||_2 < \delta$. Divide $\bar{G}$ into cubes with the property that any two points in a cube are within distance δ of each other. The approximate solution $\tilde{x}(t)$ is now constructed as follows.

Start at $t = 0$ and $x(0) = x^0$. Let $\tilde{x}(t) = x^0 + tf(x^0)$ as in Euler's method, but, rather than continuing t for a pre-determined interval h, go only until that line hits the boundary of the cube containing x^0. Call the point where it hits $\tilde{x}^1$, and the time at which it hits t_1. Now change direction, and for $t > t_1$, let $\tilde{x}(t) = \tilde{x}^1 + tf(\tilde{x}^1)$. Continue with this straight line until it hits the boundary of the second cube, at time $t = t_2$ at the point $\tilde{x}^2$.

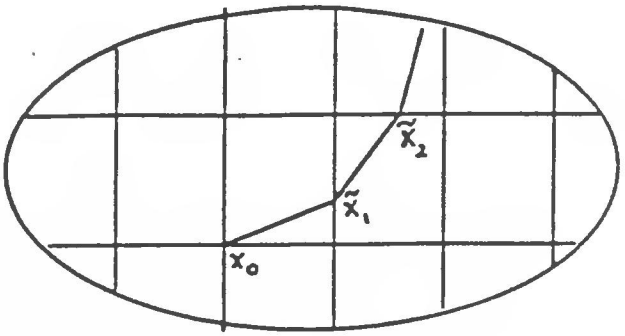


Figure 15. *Euler-Peano approximation*

In this way, we get a polygonal curve $\tilde{x}(t)$. It is convenient also to define a "step function" $\tilde{x}(t)$ by

$$\tilde{x}(t) = \begin{cases} x^0, & \text{if } 0 \le t \le t_1, \\ \tilde{x}^1, & \text{if } t_1 \le t \le t_2, \\ \tilde{x}^i, & \text{for } t_i \le t \le t_{i+1}. \end{cases}$$

Similarly, we build a polygonal curve $\tilde{X}$ and a step function $\tilde{x}$ in the direction of negative t.

We now ask: for what time interval will this curve lie in the domain G? It is possible that it remains in G for all t; perhaps it spirals into a particular point. The most pessimistic estimate, however, is that it remains in G only until the total length of the straight segments equals the distance D of x^0 from the boundary of G. That length, by the formula for arclength, is given by

$$L = \int_0^T \sqrt{\dot{\tilde{x}}_1(t)^2 + \ldots + \dot{\tilde{x}}_n(t)^2}\, dt. \tag{3.1.4}$$

We are to determine T so that $L < D$. The function $\tilde{x}_i$ are only piecewise differentiable, but the arclength formula is still correct; we should interpret it as giving the sum of the lengths of the smooth line segments. Since $\dot{\tilde{x}}_i(t) = \tilde{x}_i(t) + t f_i(\tilde{x}(t))$, and since the step function $\tilde{x}$ has zero derivative on the intervals of constancy, the expression under the radical in (3.1.4) is

$$f_1(\tilde{x}(t))^2 + \ldots + f_n(\tilde{x}(t))^2,$$

which is less than M^2 by the boundedness assumption on f. Hence (3.1.4) gives $L \le MT$; we want this to be less than D, so it suffices to take $T < D/M$. The

polygonal curve $\bar{x}(t)$ can therefore be continued within G for $0 \le t < D/M$ in the positive time direction, and by a similar argument, for $-D/M < t \le 0$ in the negative time direction. According to Example 3.1.3, this pessimistic estimate cannot be improved. It will be important in the sequel that the time interval is independent of ϵ.

The construction so far was based on a fixed ϵ. We now choose a sequence of ϵ's tending to zero, e.g. $\epsilon_j = 1/j$. For each ϵ_j, there is a continuous polygonal function $\bar{x}^{(j)}(t)$ and a step function $\tilde{x}^{(j)}(t)$. The Arzelà - Ascoli theorem will be used to extract a convergent subsequence whose limit solves (3.1.1).

We show that the sequence $\{\tilde{x}^{(j)}\}$ satisfies the conditions of uniform boundedness and equicontinuity stated in Corollary 3.0.11. Since $\bar{G}$ is compact, it is bounded, and there is a $K > 0$ such that $||x||_2 < K$ for all $x \in \bar{G}$. Then

$$\begin{aligned} ||\tilde{x}^{(j)}(t)||_2 &= ||\bar{x}^{(j)}(t) + tf(\tilde{x}^{(j)}(t))||_2 \\ &\le ||\bar{x}^{(j)}(t)||_2 + |t|\,||f(\tilde{x}^{(j)}(t))||_2 \\ &\le K + (D/M) \times M. \end{aligned}$$

A uniform bound, therefore, is $K + D$, which is of course independent of the index j.

To prove that the sequence $\{\tilde{x}^{(j)}\}$ is equicontinuous, we derive a differential equation satisfied by these functions (away from the discontinuity points). Since

$$\dot{\bar{x}}^{(j)}(t) = f(\tilde{x}^{(j)}(t)),$$

we have

$$\dot{\bar{x}}^{(j)}(t) = f(\bar{x}^{(j)}(t)) + \theta^{(j)}(t), \tag{3.1.5}$$

where

$$\theta^{(j)}(t) \stackrel{\text{def}}{=} f(\tilde{x}^{(j)}(t)) - f(\bar{x}^{(j)}(t)).$$

Because the points $\tilde{x}^{(j)}(t)$ and $\bar{x}^{(j)}(t)$ always lie in the same cube, they are within δ of each other, the values of f differ by less than ϵ, and we have $||\theta^{(j)}(t)||_2 < \epsilon_j$.

Integrate (3.1.5) from t to $t + h$:

$$\bar{x}^{(j)}(t+h) - \bar{x}^{(j)}(t) = \int_t^{t+h} f(\bar{x}^{(j)}(s))\,ds + \int_t^{t+h} \theta^{(j)}(s)\,ds,$$

and estimate:

$$||\bar{x}^{(j)}(t+h) - \bar{x}^{(j)}(t)||_2 \le hM + \epsilon_j h \le (M+1)h.$$

This establishes equicontinuity of the sequence $\{\bar{x}^{(j)}\}$: if $\delta = \epsilon/(M+1)$, then $|h| < \delta$ implies that $||\bar{x}^{(j)}(t+h) - \bar{x}^{(j)}(t)||_2 < \epsilon$ *independently* of the index j.

By the Arzelà - Ascoli theorem, there is a convergent subsequence. In order to keep the notation from becoming too baroque, we denote this subsequence by the same symbols, $\{x^{(j)}\}$. It converges uniformly on compact sets and the limit is therefore continuous on any compact subset of $|t| < D/M$. So, the limit function, called $x(t)$, is continuous on $|t| < D/M$.

Finally, we must show that the limit solves the initial value problem (3.1.1). This is not automatic, since *a priori* the limit function might not be differentiable anywhere, and so it might not be possible to plug it into the differential equation.

To accomplish this last step, we integrate the approximate differential equation (3.1.5), to obtain

$$\bar{x}^{(j)}(t) = x^0 + \int_0^t f(\bar{x}^{(j)}(s))\,ds + \int_0^t \theta^{(j)}(s)\,ds.$$

Let $j \longrightarrow \infty$ in this relation. The sequence $\{\bar{x}^{(j)}\}$ is uniformly bounded, so $f(x^{(j)}(s))$ is uniformly bounded by the assumption on f, and then according to the Lebesgue Dominated Convergence Theorem, the limit of the first integral on the right is the integral of the limit. The second integral on the right is estimated by $\epsilon_j |t| < \epsilon_j \times D/M$ which tends to zero. Hence

$$x(t) = x^0 + \int_0^t f(x(s))\,ds,$$

which shows that x is a differentiable function, that $x(0) = x^0$, and that $\dot{x}(t) = f(x(t))$. ∎

• **3.1.6 Exercise.** Divide the cube $\{0 \le x, y \le 3\} \subset \mathbf{R}^2$ into nine smaller cubes of equal size, corresponding to the choice $\delta = \sqrt{2}$ at the beginning of the proof of Theorem 3.1.2. Consider the system

$$\dot{x}_1(t) = x_1(t),$$
$$\dot{x}_2(t) = 1,$$

with initial conditions $x_1(0) = 1$, $x_2(0) = 0$. Compute and graph the polygonal curve $\bar{x}(t)$ and the step function $\bar{\bar{x}}(t)$. ∎

Subsection 3.2 The length of curves

As a second application of the Arzelà - Ascoli theorem, we will prove the existence of curves of minimal length (*geodesics*) in a very general setting. Fix two points x_1 and x_2 in a metric space, and consider all the continuous curves that connect x_1 and x_2. *There exists at least one (and possibly more than one) curve of smallest length.*

The idea of the proof is not complicated. We would like to show that curve length is a continuous function of the curve, and we would like to find a compact set of curves on which "length" achieves its minimum value.

We would like to, but unfortunately, we can't. Curve length does not depend continuously on the curve, so a new idea is required. The compact set of curves is provided by the Arzelà-Ascoli Theorem, whose hypotheses must be checked. Additional technical niceties must be observed in the mere definition of "curve" and "length". We begin with these introductory definitions.[23]

In what follows, (M, d) will be a fixed *complete* metric space.

3.2.1 Definition. *A* ***parametrized curve*** *in M is the image of $[0,1]$ under a continuous function $f : [0,1] \longrightarrow M$. A curve can be parametrized in many different ways. We say that two continuous functions $f_1, f_2 : [0,1] \longrightarrow M$ are* ***equivalent*** *if there is a change of independent variable that converts one to the other. Precisely, $f_1 \equiv f_2$ if there are continuous, nondecreasing functions $\phi_1, \phi_2 : [0,1] \longrightarrow [0,1]$ satisfying*

$$\phi_i(0) = 0, \ \phi_i(1) = 1 \quad \text{for } i = 1, 2,$$

such that

$$f_1(\phi_1(t)) = f_2(\phi_2(t)) \quad \text{for all } t \in [0,1].$$

This concept of ***equivalence*** *is indeed an equivalence relation (1.2.4) on the set of continuous functions from $[0,1]$ to M or, what is the same, on the set of parametrized curves in M. An equivalence class is simply called a* ***curve*** *in M. Curves will be denoted by Γ; the parametrized curve defined by a function f will be written Γ_f.*

We now introduce the space $C([0,1], M)$ with its metric ρ as defined in (3.0.2).

3.2.2 Definition. *A sequence $\{\Gamma_n\}$ of curves is said* ***to approach a curve*** *Γ if there exist parametrizations Γ_{f_n} of Γ_n and Γ_f of Γ such that $\rho(f_n, f) \longrightarrow 0$ as $n \longrightarrow \infty$.*

3.2.3 Definition. *Let Γ_f be a parametrized curve. Let Δ be a partition $0 = t_0 < t_1 < \ldots < t_n = 1$ of the interval $[0,1]$. Define*

$$l(\Gamma_f, \Delta) = \sum_{i=1}^{n} d(f(t_{i-1}), f(t_i)),$$

---

[23] The discussion is taken from the book *Elements of the Theory of Functions and Functional Analysis* by A.N. Kolmogorov and S.V. Fomin. I prove the main theorems, which according to them are "easily verified".

and define the ***length*** *of* Γ_f *to be*

$$l(\Gamma_f) \stackrel{\text{def}}{=} \sup_{\Delta} l(\Gamma_f, \Delta).$$

This is only a glorified version of the usual definition of arclength. The sum $l(\Gamma_f, \Delta)$ is the sum of the lengths of finitely many chords on the curve, and the length is found by taking shorter and shorter chords.

3.2.4 Example. The length may be infinite. For example, consider the graph of the continuous function

$$g(t) = \begin{cases} t\sin(\pi/t), & \text{if } 0 < t \le 1, \\ 0, & \text{if } t = 0. \end{cases}$$

To see this as a parametrized curve in the metric space $\mathbf{R}^2$, you should view the graph as the image of the function $f(t) = (t, g(t))$ from $[0,1]$ to $\mathbf{R}^2$.

Consider the chords that connect the points

$$(2/3, -2/3), (2/5, 2/5), (2/7, -2/7), \ldots.$$

in sequence. They are longer than $2/3, 2/5, 2/7, \ldots$. (Draw the graph to convince yourself of this). Thus, for the partition

$$0 < 2/(2N+1) < \ldots < 2/7 < 2/5 < 2/3 < 1,$$

we have $\sum_1^N 2/(2j+1)$ as *lower* bound for $l(\Gamma_f, \Delta)$. The harmonic series diverges, and the length, which is the supremum over all approximate sums, is infinite. ∎

3.2.5 Exercise. Prove: if $g : [0,1] \to \mathbf{R}$ has a continuous derivative, then its graph has finite length. ∎

• **3.2.6 Exercise.** Suppose that the functions f_1 and f_2 from $[0,1]$ to M are equivalent. Show that $l(\Gamma_{f_1}) = l(\Gamma_{f_2})$. (Do this by setting up a one-to-one correspondence between approximating sums for the two parametrized curves). ∎

In view of Exercise 3.2.6, we speak of the *length of a curve* and not just of the length of a particular parametrized version.

3.2.7 Example. Length of a curve is not a particularly well-behaved function of the curve. Consider once again the curve in Example 3.2.4, and define curves Γ_j by using the function $g(t)/j$. The effect is to scale the curve in (3.2.4) by a factor $1/j$. Clearly, the functions g/j tend to the zero function uniformly, so the curves Γ_j tend to the limiting curve Γ = the interval $[0,1]$. We then have $\Gamma_j \longrightarrow \Gamma$, but $l(\Gamma_j) = \infty$ while $l(\Gamma) = 1$. *The length of the limit need not equal the limit of the lengths.* In other words, *length* is not a continuous function of the curve.

This example shows that the parametrizing function g may be arbitrarily close (in the sup norm) to a given parametrizing function f, and still the length of Γ_g can be *much greater* than the length of Γ_f. It cannot, however, be *much less* than the length of Γ_f (proved in Proposition 3.2.13 below). This is a weaker version of continuity, called *semicontinuity*, which will suffice to let us minimize the length of a curve. ∎

3.2.8 Definition. *Let $(X, \mathcal{T})$ be a topological space, and let $f : X \longrightarrow \mathbf{R}$. The function f is said to be* **lower semicontinuous** *if $f^{-1}((\alpha, \infty)) \in \mathcal{T}$ for every $\alpha \in \mathbf{R}$. It is said to be* **upper semicontinuous** *if $f^{-1}((-\infty, \beta)) \in \mathcal{T}$ for every $\beta \in \mathbf{R}$.*

3.2.9 Examples

A. The function

$$f(x) = \begin{cases} x + 1, & \text{if } x > 1, \\ x, & \text{if } x \leq 1, \end{cases}$$

is lower semicontinuous on R. This is most easily seen from a picture. To prove it, look at Figure 17 and consider the set

(3.2.10) $$\{x \mid f(x) > \alpha\}.$$

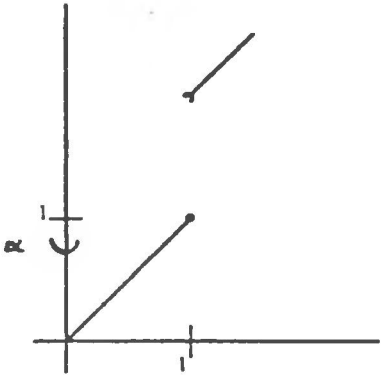


Figure 16. *A lower semicontinuous function*

If $\alpha > 2$, then (3.2.10) is $(\alpha - 1, \infty)$, which is open. If $1 \leq \alpha \leq 2$, then (3.2.10) is $(2, \infty)$ (open). If $\alpha < 1$, then (3.2.10) is (α, ∞) (again open).

If the value of $f(1)$ were changed to 1.5, the resulting function $\tilde{f}$ would no longer be lower semicontinuous. Indeed, $\{x \mid \tilde{f}(x) > 1.2\} = [1, \infty)$, which is not an open set.

If the value of $f(1)$ were changed to anything less than 1, say 0, the resulting function $\hat{f}$ would still be lower semicontinuous. Check this!

B. A function on R that is continuous from below,

$$\lim_{x \uparrow x_0} = f(x_0) \quad \text{for all } x,$$

is lower semicontinuous. The function $\hat{f}$ in the last sentence of part A shows that lower semicontinuous functions need not be continuous from below. $\hat{f}$ satisfies $\lim_{x \uparrow x_0} \hat{f}(x) \geq \hat{f}(x_0)$ for all x, and the next proposition explains why this is sufficient for lower semicontinuity.

C. A real-valued function f on a metric space (M, d) is continuous if and only if $\lim_{y_n \to y} f(y_n) = f(y)$. There is a similar criterion for lower semicontinuity.

3.2.11 Proposition. *Let f be a real-valued function on a metric space (M, d). Then f is lower semicontinuous if, and only if,*

$$\liminf f(y_n) \geq f(y)$$

whenever $y_n \to y$.

Proof. Suppose that f is lower semicontinuous. Abbreviate $\ell = \liminf f(y_n)$ and $c = f(y)$. We are to show that $c \leq \ell$.

It follows from the definition (2.5.2) of lim inf that there is a subsequence $\{y_{n_j}\}$ such that $f(y_{n_j}) \to \ell$. Of course, $\{y_{n_j}\}$ also tends to y. By lower semicontinuity, the set

$$U = \{x \mid f(x) > c - \epsilon\}$$

is open for every $\epsilon > 0$. It contains y (because $f(y) = c > c - \epsilon$), and so in fact U is a neighborhood of y. Because $y_{n_j} \to y$, we have $y_{n_j} \in U$ for all large j. Then

$$c - \epsilon < f(y_{n_j}) \quad \text{for all large } j,$$

and by letting $j \to \infty$,

$$c - \epsilon \leq \ell.$$

Since ϵ is arbitrary, it follows that $c \leq \ell$.

Suppose now that $\liminf f(y_n) \geq f(y)$ whenever $y_n \to y$. Pick any $\alpha \in \mathbf{R}$. To show that $A = \{x \mid f(x) > \alpha\}$ is open, we must show that every $y \in A$ in this set has a neighborhood contained in A. Put simply,

(3.2.12) if $f(y) > \alpha$, then $f(x) > \alpha$ for $d(x, y)$ sufficiently small.

Let us prove (3.2.12).

Because $f(y) > \alpha$, there is an η such that $f(y) > \alpha + \eta$. If A is not open, then for every n there is a y_n within $1/n$ of y, $d(y, y_n) < 1/n$, such that $y_n \notin A$, which means

$$y_n \to y \quad \text{but} \quad f(y_n) \leq \alpha.$$

Then

$$\liminf f(y_n) \leq \alpha < \alpha + \eta < f(y),$$

which contradicts the hypothesis in the proposition. Hence A must be open, and f is lower semicontinuous according to Definition 3.2.8. ∎

This proposition provides a large supply of bizarre lower semicontinuous functions for those who collect such things. For instance:

$$f(x) = \begin{cases} \sin(1/x), & \text{if } x \neq 0, \\ -2, & \text{if } x = 0 \end{cases}$$

is lower semicontinuous on R. ∎

We now return to curves, and prove the fundamental result that justifies the digression into semicontinuity.

3.2.13 Proposition. *Length is a lower semi-continuous function from the metric space $C([0,1], M)$ (of parametrizations of curves in M) to the reals.*

Proof. According to Definition 3.2.8, this means that $\{f \mid l(\Gamma_f) > \alpha\}$ is open in the metric space $C([0,1], M)$ for every $\alpha \in \mathbb{R}$. We need to show that whenever $l(\Gamma_f) > \alpha$, there exists a $\delta > 0$ such that $\rho(f,g) < \delta$ implies $l(\Gamma_g) > \alpha$.[24]

Let α be given, and suppose that $l(\Gamma_f) > \alpha$. Then there is a partition Δ whose approximating sum is so close to $l(\Gamma_f)$ that it is also larger than α. In fact, for some ϵ,

$$\sum_{j=1}^{n} d(f(t_{i-1}), f(t_i)) > \alpha + \epsilon.$$

Now suppose that $\rho(f,g) < \epsilon/4n$. We have, by the triangle inequality,

$$d(g(t_{i-1}), g(t_i)) \leq d(g(t_{i-1}), f(t_{i-1})) + d(f(t_{i-1}), f(t_i)) + d(f(t_i), g(t_i)).$$

Write the same inequality with f and g interchanged, and subtract one from the other. Rearrange to get

$$\begin{aligned} \left|d(g(t_{i-1}), g(t_i)) - d(f(t_{i-1}), f(t_i))\right| &\leq d(g(t_{i-1}), f(t_{i-1})) + d(g(t_i), f(t_i)) \\ &< 2 \times \epsilon/4n = \epsilon/2n. \end{aligned}$$

Now sum this inequality from $j = 1$ to n, to arrive at the estimate

$$|l(\Gamma_g, \Delta) - l(\Gamma_f, \Delta)| < \epsilon/2.$$

Since $l(\Gamma_f, \Delta) > \alpha + \epsilon$, we conclude that $l(\Gamma_g, \Delta) > \alpha + \epsilon/2$. The length of Γ_g can only be larger (it is the supremum of all approximating sums), and the proposition is proved. ∎

We are now ready to tackle the main result.

3.2.14 Theorem. *Let x_1, x_2 be two points in M, let C be a compact subset of M containing x_1 and x_2, and suppose that C contains a parametrized curve Γ_f of finite length that connects x_1 to x_2 ($f(0) = x_1, f(1) = x_2$). Among all curves lying in C and connectiong x_1 and x_2, there is at least one of minimal length.*

Proof. By assumption, there is a curve of finite length K, which connects x_1 and x_2 and lies entirely in the compact set $C \subset M$. Let $\mathcal{G}$ be the collection of all curves of length $\leq K$, connecting x_1 and x_2, and lying in C. We show that it is possible to

[24] In words, again: if the parametrizing function g is sufficiently close to a given f, the length of Γ_g cannot be much smaller than the length of Γ_f.

choose parametrizations in such a way that the parametrizing functions in $\mathcal{G}$ form an equicontinuous family.

Let Γ be a curve of finite length and let f be one of its parametrizations. For $t \in [0,1]$, let s be the length of the curve $\{f(r) \mid 0 \le r \le t\}$. s is a non-decreasing function of t; write $s = \phi(t)$, and note that $\phi : [0,1] \longrightarrow [0,S]$ where S is the length of Γ_f. We can exhibit the same curve with a different parametrization, $g(s) = f(\phi^{-1}(s))$, but now the parameter s runs over $[0,S]$ rather than over $[0,1]$. To remedy this, introduce $\tau = s/S$ and $F(\tau) = g(S\tau)$.

The new parametrization of our curve by F has the following advantage. By the definition of length, the length of a chord does not exceed the arclength, and so

$$d(g(s_1), g(s_2)) \le |s_1 - s_2|,$$

whence

$$d(F(\tau_1), F(\tau_2)) \le S|\tau_1 - \tau_2|.$$

Now return to the collection $\mathcal{G}$ of curves defined above. Each $\Gamma \in \mathcal{G}$ can be parametrized by a function F as just described, and since they all have total length $\le K$, all the F's satisfy $d(F(\tau_1), F(\tau_2)) \le K|\tau_1 - \tau_2|$. The collection $\mathcal{F}$ of these functions F is therefore equicontinuous (if you choose $\delta = \epsilon/K$, then $|\tau_1 - \tau_2| < \delta$ implies $d(F(\tau_1), F(\tau_2)) < \epsilon$, and the δ is independent of F). Furthermore, the images of all $F \in \mathcal{F}$ are contained in the compact set C, and the hypotheses of the Arzelà - Ascoli theorem are satisfied.

Now let ℓ = greatest lower bound of $\{l(\Gamma) \mid \Gamma \in \mathcal{G}\}$. The set of lengths is bounded below by zero, so the greatest lower bound ℓ does exist. There must be a sequence $\{\Gamma_n\} \subset \mathcal{G}$ such that $l(\Gamma_n) \longrightarrow \ell$. These curves may be paramtrized by functions F_n as just described. The sequence $\{F_n\}$ has a subsequence that converges uniformly on the compact interval $[0,1]$ to a continuous function F. Denote the curve corresponding to F by Γ. It belongs to $\mathcal{G}$.

To finish the proof, we need to show that $l(\Gamma) = \ell$. Assume the contrary, that $l(\Gamma) > \ell$ (it cannot be smaller, since ℓ is a lower bound on all the lengths of curves in $\mathcal{G}$). For some $\epsilon > 0$, we will also have $l(\Gamma) > \ell + \epsilon$. Because length is lower semi-continuous, the set $O = \{f \mid l(\Gamma_f) > \ell + \epsilon\}$ is open, and in fact it is an open neighborhood of F. Since $F_n \longrightarrow F$, there is an N such that for $n > N$, $F_n \in O$. Then $l(\Gamma_n) > \ell + \epsilon$ for all $n > N$. This contradicts the property used to define the Γ_n, that their lengths tend to ℓ. Hence, $l(\Gamma) = \ell$ and the proof is concluded. ∎

• **3.2.15 Exercise.** Let $\{x_n\}$ be a sequence of real numbers, and suppose it is bounded from below. Show that $\ell = \liminf x_n$ has the following two properties: i) for every $\epsilon > 0$, there is an N_ϵ such that $x_n > \ell - \epsilon$ when $n > N_\epsilon$, and ii) for every $\epsilon > 0$ and $M > 0$ there is an $n > M$ such that $x_n < \ell + \epsilon$. Show conversely that if a number ℓ has properties i) and ii), then $\ell = \liminf x_n$. ∎

• **3.2.16 Exercise.** Characterize $\limsup x_n$ by two properties analogous to the ones in the preceding exercise. ∎

• **3.2.17 Exercise.** Formulate and prove a characterization of upper semicontinuous functions in terms of lim sup. Use Proposition 3.2.11 as a model. ∎

• **3.2.18 Exercise.** Show that the collection $\mathcal{S}$ of sets $\{(a, \infty) \mid a \in \mathbf{R}\}$ is a topology on $\mathbf{R}$. What functions are continuous with respect to this topology? Answer the same questions for $\mathcal{S} = \{(-\infty, a) \mid a \in \mathbf{R}\}$. ∎

• **3.2.19 Exercise.** Let f be a lower semicontinuous real-valued function on a compact topological space (X, T). Show that f is bounded from below, and that it assumes its minimum value. Show by example that there need not be an $x_* \in X$ such that $f(x) \leq f(x_*)$ for all $x \in X$. ∎

• **3.2.20 Exercise.** I restate Proposition 3.2.13 in words: if the parametrizing functions g and f are close, then the length $\ell(\Gamma_g)$ is not much less than $\ell(\Gamma_f)$. But then (just interchange the roles of f and g) the length $\ell(\Gamma_f)$ is not much smaller than $\ell(\Gamma_g)$. Since neither of $\ell(\Gamma_f)$ and $\ell(\Gamma_g)$ is much smaller than the other, they must be close to each other. Hence, length *is* a continuous function of the curve after all. What is wrong? ∎

Proof of Proposition II.1.4.29.

One should first prove Proposition 1.4.34.
Then, in Proposition 1.4.29, one considers (for $a, x \in J$)

$$\int_a^x \sum_{n=1}^{N} f_n'(t)\,dt = \sum_{n=1}^{N} f_n(x) - \sum_{n=1}^{N} f_n(a).$$

By Proposition 1.4.34, the left side has limit

$$\int_a^x g(t)\,dt$$

as $N \to \infty$. Because $\sum f_n$ converges to f, the right side has limit $f(x) - f(a)$ (as $N \to \infty$). Hence

$$\int_a^x g(t)\,dt = f(x) - f(a),$$

which establishes 1.4.29. ∎

Comment to II - 325

Note, however, that $1 + \cos(21^{n-m}\pi q) \geq 0$: this allows us in (1.4.31) to underestimate the sum by its first term. Since by definition (bottom of II - 324) $|q| \leq 1/2$, this first term is

$$\frac{1 + \cos\theta}{3^m}, \quad \text{where} \quad -\pi/2 \leq \theta \leq \pi/2.$$

The result is the first inequality in the second line of (1.4.31). The second inequality comes from the reciprocal of the first displayed equation on II - 325.

Comment to II - 328

Example 1.4.33: In 1.4.18B, f_n was $n(1-x)x^n$. It is true that $f_n(x) = n^2(1-x)x^n$ tends to zero pointwise on $[0, 1]$ but not uniformly—but the earlier proof must be modified.