Linear Algebra
Review

1 Linear vector spaces with a Euclidean structure

1.1 Definitions

- S is a real (resp. complex) *linear vector space* if for every x, y in S,

$$\begin{aligned} &x + y \in S \\ &\alpha x \in S \qquad \text{for every } \alpha \in \mathbf{R} \text{ (resp. } \mathbf{C}). \end{aligned}$$

- S is *finite dimensional* if it possesses a basis, i.e. if there exists a finite set of linearly independent vectors which span S. Here, we are only concerned with **finite dimensional** vector spaces.

- The *dimension* of a finite dimensional vector space S is the number of vectors in any basis of S.

1.2 Real or complex Euclidean structure

If S is a linear vector space over $\mathbf{R}$, an *inner product* $(x, y) \in S^2 \to < x, y > \in \mathbf{R}$ on S satisfies the following properties, for every $x,\ y,\ z \in S$ and every $\alpha \in \mathbf{R}$,

- It is *bilinear* and *symmetric*:

  1. $< x, y > = < y, x >$ (symmetric)
  2. $< x + y, z > = < x, z > + < y, z >$
  3. $< \alpha x, y > = \alpha < x, y >$

- It is *positive*:

  4. $< x, x > \geq 0$

- It is *definite*:

  5. $< x, x >= 0$ if and only if $x = 0$.

If S is a vector space over $\mathbf{C}$, the inner product $(x, y) \in S^2 \to < x, y > \in \mathbf{C}$ is *skew-symmetric*:

1. $< x, y >= \overline{< y, x >}$ (skew-symmetric)
2. $< x + y, z >=< x, z > + < y, z >$
3. $< \alpha x, y >= \alpha < x, y >$
4. $< x, x > \geq 0$
5. $< x, x >= 0$ if and only if $x = 0$,

for every $x,\ y,\ z \in S$ and every $\alpha \in \mathbf{C}$.

1.3 Normed vector spaces

A norm on S is a function $||.|| :\ S \to \mathbf{R}^+$, such that for every $x \in S$,

1. $||x|| > 0$ if $x \neq 0$
2. $||x|| = 0$ if and only if $x = 0$
3. $||\alpha x|| = |\alpha| ||x||$ for every α in $\mathbf{R}$ (resp. $\mathbf{C}$)
4. Triangle inequality: $||x + y|| \leq ||x|| + ||y||$

Induced norm: If S has a Euclidean structure, $||x|| = \sqrt{< x, x >}$ is a norm and is called the *induced norm*.

Exercise: Check that the induced norm is a norm.

Cauchy-Schwarz inequality:

$$| < x, y > |^2 \leq ||x||^2 ||y||^2$$

Proof: Compute

$$\begin{aligned} ||x-\alpha y||^2 &= <x-\alpha y, x-\alpha y> \\ &= <x,x> -\alpha <y,x> -\bar{\alpha} <x,y> +|\alpha|^2 <y,y> \\ &= ||x||^2 - 2\Re e\,(\bar{\alpha} <x,y>) + |\alpha|^2 ||y||^2. \end{aligned}$$

With $\alpha = \dfrac{<x,y>}{||y||^2}$, we get

$$\begin{aligned} ||x-\alpha y||^2 &= ||x||^2 - 2\Re e\left(\frac{|<x,y>|^2}{||y||^2}\right) + \frac{|<x,y>|^2}{||y||^4}||y||^2 \\ &= ||x||^2 - \frac{|<x,y>|^2}{||y||^2}. \end{aligned}$$

Since the left-hand side of this inequality is positive, we obtain the Cauchy-Schwarz inequality.

1.4 Orthogonality

Two vectors x and y are orthogonal if and only if $<x,y>=0$.

Gram-Schmidt orthogonalization procedure

This procedure allows us to generate a set of orthogonal vectors from any set of linearly independent vectors. Given a set of n linearly independent vectors $x_1,\ x_2,\ \ldots,\ x_n$, we proceed as follows:

- Take $\phi_1 = x_1$
- Take $\phi_2 = x_2 + \alpha x_1$ and impose $<\phi_1, \phi_2>=0$, i.e. take $\alpha = -\dfrac{<x_1,x_2>}{||x_1||^2} = -\dfrac{<\phi_1,x_2>}{||\phi_1||^2}$. Then,

  $$\phi_2 = x_2 - \frac{<\phi_1,x_2>}{||\phi_1||^2}\phi_1,$$

  which shows that ϕ_2 is obtained from x_2 by substracting its projection onto ϕ_1.

- More generally, take $\phi_k = x_k - \sum_{i=1}^{k-1} \frac{< \phi_i, x_k >}{||\phi_i||^2} \phi_i$.

The set $\{\phi_i\}_{i=1,\dots,n}$ is then a set of n linearly independent, orthogonal vectors.

Orthogonal complement

The *orthogonal complement* of a subspace X of S is the linear subspace of S such that

$$X^{\perp} = \{z \in S| < z, x >= 0, \text{ for every } x \in S\}.$$

Since S **is finite dimensional**, the orthogonal $X^{\perp}$ complement of every subspace X exists. Moreover, each vector y in S can be uniquely written as the sum of a vector x in X and a vector $x^{\perp}$ in $X^{\perp}$, i.e. $y = x + x^{\perp}$. This is the **orthogonal projection theorem**.

2 Linear mappings and matrices

2.1 Linear mappings

Let X and U be two vector spaces over $\mathbf{R}$ (or $\mathbf{C}$). The mapping $T : X \to U$ is linear if for every $x, y \in X$ and every $\alpha \in \mathbf{R}$ (or $\mathbf{C}$),

$$T(x + y) = T(x) + T(y)$$
$$T(\alpha x) = \alpha T(x)$$

Properties of linear mappings

- We call $\Re_T$ the *range* of T, i.e. the image of X under T. It is a linear subspace of U.

- We call *nullspace* or *kernel* of T the set $\mathcal{N}_T$ of vectors of X whose image under T is zero:
$$\mathcal{N}_T = \{x \in X|\ T(x) = 0\}.$$
It is a linear subspace of X.

- The dimensions of $\Re_T$ and $\mathcal{N}_T$ are such that $dim(\Re_T) + dim(\mathcal{N}_T) = dim(X)$.

- **Adjoint of a linear mapping:** If A is a linear mapping of a Euclidean space S into itself, the *adjoint* A^* of A is the linear map of S into itself such that for every x and y in S

$$< Ax, y >=< x, A^* y > .$$

  Exercise: Show that $(A^*)^* = A$.

2.2 Matrices

Every linear map

$$\begin{aligned} T \quad : \quad & \mathbf{R}^n \to \mathbf{R}^m \\ & x \to u \end{aligned}$$

can be written as

$$u_i = \sum_{j=1}^{n} t_{ij} x_j \quad i = 1, \ldots, m; \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix}; \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

where (t_{ij}) is a $m \times n$ (m rows, n columns) matrix.

Note that once a basis has been chosen, every linear vector space of dimension n is isomorphic to $\mathbf{R}^n$. We can then represent any linear mapping between two finite dimensional vector spaces by a matrix.

Definitions

1. The *transpose* of T is T^T such that $(T_{ij}^T) = (T_{ji})$.

2. The *row space* of T is the column space of T^T; the *left nullspace* of T is the nullspace of T^T.

3. The *rank* of T is the dimension of $\Re_T$, which is equal to the dimension of $\Re_{T^T}$.

Properties

1. If A is a $m \times n$ matrix, the system $Ax = b$ has **at least** one solution for every b if and only if the columns of A span $\mathbf{R}^m$. Then the rank of A, r, is such that $r = m$, which implies $m \leq n$.

2. The system $Ax = b$ has **at most** one solution for every b if and only if the columns of A are linearly independent, i.e. if and only if the nullspace of A is trivial. Then, $r = n$, which implies $n \leq m$.

3. In $\mathbf{R}^n$, the *standard Euclidean structure* is defined by

$$< x, y >= x^T y.$$

   Then, A^T is such that $A^* = A^T$ since

$$< Ax, y >= (Ax)^T y = x^T A^T y =< x, A^T y > .$$

   In $\mathbf{C}^n$, $< x, y >= \bar{x}^T y$ and $A^* = \bar{A}^T$, i.e. the adjoint of A is the *Hermitian transpose* of A.

4. The rowspace of A is orthogonal to the nullspace of A (in $\mathbf{R}^n$). Indeed, $Ax = 0$ and $v = A^T z$ implies

$$v^T x = z^T (A^T)^T x = z^T A x = 0.$$

5. The column space of A is orthogonal to the left nullspace of A (in $\mathbf{R}^m$). Indeed, $v = Az$ and $A^T x = 0$ give

$$v^T x = z^T A^T x = 0.$$

6. **Fredholm alternative:** The equation $Ax = b$ is solvable if and only if $< b, y >= 0$ whenever $A^* y = 0$.

   For a real matrix, this is a direct consequence of property 5 above and of the fact that $\dim(\mathcal{N}_{A^T}) + \dim(\Re_A) = m$. In other words, the range of A is the orthogonal complement of the nullspace of A^T in $\mathbf{R}^m$. The same result holds for a complex matrix, where the adjoint is defined as the Hermitian transpose.

2.3 Eigenvalues and eigenvectors

Let A be a real or complex $n \times n$ matrix.

- $h \in S$ is an eigenvector of A with eigenvalue $a \in \mathbf{C}$ if

  $$Ah = ah, \qquad h \neq 0.$$

  Exercise: Show that eigenvectors of a matrix A corresponding to different eigenvalues are linearly independent.

- $f \in S$ is a generalized eigenvector of A with eigenvalue a if

  $$(A - aI)f \neq 0, \qquad (A - aI)^m f = 0, \qquad f \neq 0,$$

  for some positive integer $m \neq 1$.

2.4 Spectral theorem for square matrices

Let A be a $n \times n$ matrix with complex entries. Every vector in $\mathbf{C}^n$ can be written as a sum of eigenvectors of A, genuine or generalized.

3 Self-adjoint matrices

A matrix A is *self-adjoint* if $A^* = A$. If A has real entries, then A is *symmetric*. If A has complex entries, them A is *Hermitian*.

3.1 Spectral theorem for self-adjoint mappings

Theorem: A self-adjoint mapping A of a complex Euclidean space S into itself has **real eigenvalues** and a set of **eigenvectors** that form an **orthogonal basis** of S.

Exercise: Prove the above theorem using the following steps:

1. Show that A has only real eigenvalues.
2. Show that A has no generalized eigenvectors.

3. Show that eigenvectors of A corresponding to different eigenvalues are orthogonal.

4. Use the spectral theorem for square matrices to finish the proof.

A linear operator A mapping a Euclidean space into itself is called *anti-selfadjoint or skew-adjoint* if $A^* = -A$. Then, iA is self-adjoint and the spectral theorem becomes

Theorem: Let A be a skew-adjoint mapping of a complex Euclidean space S into itself. Then

1. The eigenvalues of A are pure imaginary

2. There is an orthonormal basis of S which consists of eigenvectors of A.

3.2 Proof of the spectral theorem

Assume that A is a self-adjoint mapping from S into itself. The **Rayleigh quotient** of A is defined as

$$R_A(x) = \frac{< x, Ax >}{< x, x >}$$

for every non-zero vector x of S.

Properties of the Rayleigh quotient:

1. R_A is real since A is self-adjoint.

2. R_A is a homogeneous function of degree zero, i.e. $R_A(hx) = R_A(x)$ where h is a scalar and x a non-zero vector of S.

3. If A is self-adjoint and if one assumes that A has an orthogonal basis of eigenvectors $\{f_i\}$ with real eigenvalues a_i, $a_1 \leq a_2 \leq \ldots \leq a_n$, then

$$\begin{aligned} \frac{< x, Ax >}{< x, x >} &= \frac{\sum_{i=1}^n |x_i|^2 \bar{a}_i}{\sum_{i=1}^n |x_i|^2} = \frac{\sum_{i=1}^n |x_i|^2 a_i}{\sum_{i=1}^n |x_i|^2} \\ &\geq \frac{\sum_{i=1}^n |x_i|^2 a_1}{\sum_{i=1}^n |x_i|^2} = a_1. \end{aligned}$$

Moreover, $R_A(f_1) = a_1$. In other words, assuming the spectral theorem is true, the smallest eigenvalue a_1 of A is equal to the minimum of the Rayleigh quotient over $S^* = \{x \in S |\ x \neq 0\}$.

On the basis of this last remark, we will prove the spectral theorem by using the Rayleigh quotient to find the eigenvalues of A and construct its eigenvectors. The proof is in two parts.

Part 1. We look for the minimum of R_A. Since R_A is a homogeneous function of degree 0, it suffices to search its minimum on the sphere of radius one. Since the latter is a compact set and since R_A is a continuous function, R_A reaches its minimum on this sphere at a vector which we call f_1. We now show that f_1 is an eigenvector of A. Let g be any vector, t a real variable, and compute

$$\begin{aligned} R_A(f_1 + tg) &= \frac{< f_1 + tg, A(f_1 + tg) >}{< f_1 + tg, f_1 + tg >} \\ &= \frac{< f_1, Af_1 > + 2t\Re e(< g, Af_1 >) + t^2 < g, Ag >}{< f_1, f_1 > + 2t\Re e(< g, f_1 >) + t^2 < g, g >}. \end{aligned}$$

Since R_A is minimum at f_1, we have

$$\begin{aligned} 0 = \left.\frac{dR_A}{dt}\right|_{t=0} &= \frac{2\Re e(< g, Af_1 >)}{< f_1, f_1 >} - \frac{2 < f_1, Af_1 > \Re e(< g, f_1 >)}{< f_1, f_1 >^2} \\ &= 2\Re e\left(< g, Af_1 - < f_1, Af_1 > f_1 >\right) \qquad (\text{since } < f_1, f_1 >= 1) \\ &= < g, Af_1 - < f_1, Af_1 > f_1 > + < Af_1 - < f_1, Af_1 > f_1, g > \quad (1) \end{aligned}$$

If one replaces g by ig in Eqn.(1), one gets

$$0 =< g, Af_1 - < f_1, Af_1 > f_1 > - < Af_1 - < f_1, Af_1 > f_1, g > . \qquad (2)$$

By taking the sum (1)+(2), we find

$$< g, Af_1 - < f_1, Af_1 > f_1 >= 0,$$

which is true for every vector g. Therefore, $Af_1 =< f_1, Af_1 > f_1$, i.e. f_1 is an eigenvector of A with eigenvalue

$$< f_1, Af_1 >= \min_{||f||=1} R_A(f).$$

We call a_1 this eigenvalue, which, by construction, is real.

Part 2. We now prove that A has a set of orthonormal eigenvectors which make a basis of S. Let $S_{f_1} = \{x \in S | < x, f_1 >= 0\}$ be the orthogonal complement of f_1. We have dim(S_{f_1})= dim(S)-1. Moreover, A maps S_{f_1} into itself. Indeed, for every x in S_{f_1},

$$< Ax, f_1 >=< x, Af_1 >=< x, a_1 f_1 >= a_1 < x, f_1 >= 0.$$

Since A restricted to S_{f_1} is self-adjoint, we can repeat the procedure of Part 1 above and find f_2 such that $||f_2|| = 1$ and $< f_2, Af_2 >$ is the minimum of R_A on the sphere of radius one in S_{f_1}. As before, f_2 is an eigenvector of A with eigenvalue $a_2 =< f_2, Af_2 > \geq a_1$. By construction, f_2 is perpendicular to f_1.

By induction, we can then construct an orthogonal basis of S made of eigenvectors of A with (real) eigenvalues $a_1, a_2, \ldots, a_n$ such that $a_1 \leq a_2 \leq \ldots \leq a_n$. The proof of the spectral theorem for self-adjoint matrices is therefore complete.

3.3 The minimax principle

Minimax principle: Let A be a real symmetric map of a real Euclidean space S of finite dimension into itself. Denote the eigenvalues of S, arranged in increasing order, by $a_1, a_2, \ldots, a_n$. Then

$$a_j = \min_{\dim(S_j)=j} \max_{\substack{x \in S_j \\ x \neq 0}} \frac{< x, Ax >}{< x, x >},$$

where S_j is a linear subspace of S.

Proof: The proof is in two steps. First, we show that for any linear subspace S_j of S with dim(S_j)= j, we have $\max_{x \neq 0, x \in S_j} \frac{< x, Ax >}{< x, x >} \geq a_j$. To show this, it is enough to find one vector x in S_j such that $R_A(x) \geq a_j$. Let $\{f_i\}$ be the eigenvectors of A with eigenvalues $a_1, \ldots, a_n$. Take $x \neq 0 \in S_j$ such that the $j-1$ linear conditions $< x, f_i >= 0,\ i = 1, \ldots, j-1$ are satisfied. This is always possible since S_j has dimension j. Then,

$$\begin{aligned} \frac{< x, Ax >}{< x, x >} &= \frac{< \sum_{i=1}^n x_i f_i, A\left(\sum_{i=1}^n x_i f_i\right) >}{< \sum_{i=1}^n x_i f_i, \sum_{i=1}^n x_i f_i >} = \frac{\sum_{i=1}^n x_i^2 a_i}{\sum_{i=1}^n x_i^2} \\ &= \frac{\sum_{i=j}^n x_i^2 a_i}{\sum_{i=j}^n x_i^2} \geq a_j, \end{aligned}$$

where we have written $x = \sum_{i=1}^{n} x_i f_i$.

Now, we have to find one j-dimensional subspace S_j of S such that for every non-zero x in S_j, $R_A(x) \leq a_j$. If we take S_j as the vector space spanned by $f_1, f_2, \ldots, f_j$, then every x in S_j can be written as $x = \sum_{i=1}^{j} x_i f_i$, and

$$R_A(x) = \frac{\sum_{i=1}^{j} x_i^2 a_i}{\sum_{i=1}^{j} x_i^2} \leq a_j.$$

Therefore, $\max_{x \neq 0, x \in S_j} \frac{< x, Ax >}{< x, x >}$ is always larger than a_j and can be equal to a_j for a given vector space. In other words,

$$a_j = \min_{\dim(S_j)=j} \max_{\substack{x \in S_j \\ x \neq 0}} \frac{< x, Ax >}{< x, x >}.$$

The proof is complete.

3.4 Application

Suppose A is a real symmetric map of a real Euclidean space S into itself. Denote $a_1, \ldots, a_n$ the eigenvalues of A arranged in increasing order, and consider the restriction of A to the linear subspace $\mathcal{N}_B$ defined by $\{x \in S | Bx = 0\}$, where B is a matrix of rank k. Although A does not map $\mathcal{N}_B$ into itself, one can still look at the relative extrema of $R_A(x)$ when x is constrained to be in $\mathcal{N}_B$. Denote by $\tilde{a}_1, \ldots, \tilde{a}_{n-k}$ these relative extrema, arranged in increasing order. Then,

$$a_i \leq \tilde{a}_i \leq a_{i+k}, \qquad i = 1, \ldots, n-k.$$

Proof: By definition,

$$\begin{aligned}
\tilde{a}_i &= \min_{\substack{\dim(S_i) = i \\ S_i \subset \mathcal{N}_B}} \max_{\substack{x \neq 0 \\ x \in S_i}} \frac{< x, Ax >}{< x, x >} \\
&= \max_{\substack{x \neq 0 \\ x \in \mathrm{span}(\{\tilde{f}_1, \ldots \tilde{f}_i\})}} \frac{< x, Ax >}{< x, x >} \geq a_i,
\end{aligned}$$

since span$(\{\tilde{f}_1, \dots \tilde{f}_i\})$ is a vector space of dimension i. Here, $\{\tilde{f}_1, \dots \tilde{f}_i\}$ denotes the set of vectors for which the relative maxima of $R_A(x)$ restricted to $\mathcal{N}_B$ are reached.

On the other hand,

$$\begin{aligned}
\tilde{a}_i &= \min_{\substack{\dim(S_i) = i \\ S_i \subset \mathcal{N}_B}} \max_{\substack{x \neq 0 \\ x \in S_i}} \frac{< x, Ax >}{< x, x >} \leq \max_{\substack{x \neq 0 \\ Bx = 0 \\ < x, f_j >= 0 \\ j = i+k+1, \dots, n}} \frac{< x, Ax >}{< x, x >} \\
&\leq \max_{\substack{x \neq 0 \\ < x, f_j >= 0 \\ j = i+k+1, \dots, n}} \frac{< x, Ax >}{< x, x >} \leq a_{i+k}.
\end{aligned}$$

The first inequality comes from the fact that $< x, f_j >= 0,\ j = i+k+1, \dots, n$ and $Bx = 0$ correspond to k constraints in a space of dimension $i+k$, i.e. these two conditions define a space of dimension at least equal to i, whose elements satisfy $Bx = 0$.

The proof is then complete, since by combining the two inequalities above, we get

$$a_i \leq \tilde{a}_i \leq a_{i+k}.$$

Example: See Keener's book, page 23.

References

- Peter D. Lax, *Linear Algebra*, John Wiley & Sons, New York, 1996.
- Gilbert Strang, *Linear Algebra and its Applications*, Harcourt Brace Jovanovich Publishers, San Diego, 1986.
- James P. Keener, *Principles of Applied Mathematics*, Addison Wesley, 1988.