Contents

1 Inner product

1.1 Inner product spaces

Basics: Inner product, or dot product, brings a metric for vector lengths. It takes two vectors and generates a number. In $\mathbb{R}^n$, we have

$$\langle a, b\rangle \triangleq a^T b = [a_1, a_2, \dots, a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Clearly, $\langle a, b\rangle \triangleq a^T b = \langle b, a\rangle$. Letting $b = a$ above, we get the square of the length of a:

$$||a|| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}$$

Formal definitions:

Definition 1. A real vector space $\mathbf{V}$ is called a real inner product space, if for any vectors a and b in $\mathbf{V}$ there is associated a real number $\langle a, b\rangle$, called the inner product of a and b, such that the following axioms hold:

- (linearity) For all scalars q_1 and q_2 and all vectors $a, b, c \in \mathbf{V}$

$$\langle q_1 a + q_2 b, c \rangle = q_1 \langle a, b \rangle + q_2 \langle b, c \rangle$$

- (symmetry) $\forall a, b \in \mathbf{V}$

$$\langle a, b \rangle = \langle b, a \rangle$$

- (positive definiteness) $\forall a \in \mathbf{V}$

$$\langle a, a \rangle \geq 0$$

where $\langle a, a \rangle = 0$ if and only if $a = 0$.

□

Definition 2 (2-norm of vectors)**.** The length of a vector in $\mathbf{V}$ is defined by

$$||a|| = \sqrt{\langle a, a \rangle} \geq 0$$

For $\mathbb{R}^n$,

$$||a|| = \sqrt{a^T a} = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}$$

This is the Euclidean norm or 2-norm of the vector. $\mathbb{R}^n$ equiped with the inner product $\langle a, b \rangle = \sqrt{a^T b}$ is called the n-dimensional Euclidean space.

Example 3 (Inner product for functions, function spaces)**.** The set of all real-valued continuous functions $f(x)$, $g(x)$, $\ldots$ $x \in [\alpha, \beta]$ is a real vector space under the usual addition of functions and multiplication by scalars. An inner product on this function space is

$$\langle f, g \rangle = \int_\alpha^\beta f(x)\, g(x)\, \mathrm{d}x$$

and the norm of f is

$$||f(x)|| = \sqrt{\int_\alpha^\beta f(x)^2\, \mathrm{d}x}$$

□

Inner products is also closely related to the geometric relationships between vectors. For the two-dimensional case, it is readily seen that

$$v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

is a basis of the vector space. The two vectors are additionally orthogonal, by direct observation.

More generally, we have:

Definition 4 (Orthogonal vectors)**.** Vectors whose inner product is zero are called orthogonal.

□

Definition 5 (Orthonormal vectors)**.** Orthogonal vectors with unity norm is called orthonormal.

□

Definition 6. The angle between two vectors is defined by

$$\cos\angle(a, b) = \frac{\langle a, b \rangle}{||a|| \cdot ||b||} = \frac{\langle a, b \rangle}{\sqrt{\langle a, a \rangle} \cdot \sqrt{\langle b, b \rangle}}$$

1.2 Trace (standard matrix inner product)

The trace of an $n \times n$ matrix $A = [a_{jk}]$ is given by

$$\mathrm{Tr}\,(A) = \sum_{i=1}^{n} a_{ii} \tag{1}$$

Trace defines the so-called **matrix inner product**:

$$\langle A, B \rangle = \mathrm{Tr}\left(A^T B\right) = \mathrm{Tr}\left(B^T A\right) = \langle B, A \rangle \tag{2}$$

which is closely related to vector inner products. Take an example in $\mathbb{R}^{3\times 3}$: write the matrices in the column-vector form $B = [\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3]$, $A = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$, then

$$A^T B = \begin{bmatrix} \mathbf{a}_1^T \mathbf{b}_1 & * & * \\ * & \mathbf{a}_2^T \mathbf{b}_2 & * \\ * & * & \mathbf{a}_3^T \mathbf{b}_3 \end{bmatrix} \tag{3}$$

So

$$\mathrm{Tr}\left(A^T B\right) = \mathbf{a}_1^T \mathbf{b}_1 + \mathbf{a}_2^T \mathbf{b}_2 + \mathbf{a}_3^T \mathbf{b}_3$$

which is nothing but the inner product of the two "stacked" vectors $\begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}$ and $\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix}$. Hence

$$\langle A, B \rangle = \mathrm{Tr}\left(A^T B\right) = \left\langle \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{bmatrix}, \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} \right\rangle$$

2 Norms

Previously we have used $||\cdot||$ to denote the Euclidean length function. At different times, it is useful to have more general notions of size and distance in vector spaces. This section is devoted to such generalizations.

2.1 Vector norm

Definition 7. A *norm* is a function that assigns a real-valued length to each vector in a vector space $\mathbb{C}^m$. To develop a reasonable notion of length, a norm must satisfy the following properties: for any vectors a, b and scalars $\alpha \in \mathbb{C}$,

- the norm of a nonzero vector is positive: $||a|| \geq 0$, and $||a|| = 0$ if and only if $a = 0$
- scaling a vector scales its norm by the same amount: $||\alpha a|| = |\alpha| \, ||a||$
- triangle inequality: $||a + b|| \leq ||a|| + ||b||$

Let w_1 be a $n \times 1$ vector. The most important class of vector norms, the p norms, of w are defined by

$$||w||_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \ 1 \leq p < \infty$$

Specifically, we have

$\|w\|_1 = \sum_{i=1}^{n} |w_j|$ (absolute column sum)

$\|w\|_\infty = \max_i |w_i|$

$\|w\|_2 = \sqrt{w^H w}$ (Euclidean norm)

Remark 8. When unspecified, $||\cdot||$ refers to 2 norm in this set of notes.

Intuitions for the infinity norm By definition

$$||w||_\infty = \lim_{p \to \infty} \left(\sum_{i=1}^{n} |w_i|^p \right)^{1/p}$$

Intuitively, as p increases, $\max_i |w_i|$ takes more and more weighting in $\sum_{i=1}^{n} |w_i|^p$. More rigorously, we have

$$\lim_{p \to \infty} \left((\max |w_i|)^p \right)^{1/p} \leq \lim_{p \to \infty} \left(\sum_{i=1}^{n} |w_i|^p \right)^{1/p} \leq \lim_{p \to \infty} \left(\sum_{i=1}^{n} (\max |w_i|)^p \right)^{1/p}$$

Both $\lim_{p \to \infty} \left((\max |w_i|)^p \right)^{1/p}$ and $\lim_{p \to \infty} \left(\sum_{i=1}^{n} (\max |w_i|)^p \right)^{1/p}$ equals $\max_i |w_i|$. Hence $||w||_\infty = \max |w_i|$

2.2 Induced matrix norm

As matrices define linear transformations between vector spaces, it makes sense to have a measure of the "size" of the transformation. Induced matrix norms are defined by

$$||M||_{p\leftarrow q} = \max_{x\neq 0} \frac{||Mx||_p}{||x||_q} \tag{4}$$

In other words, $||M||_{q\leftarrow q}$ is the maximum factor by which M can "stretch" a vector x.

In particular, the following matrix norms are common:

$\|M\|_{1\leftarrow 1} = \max_j \sum_{i=1}^n |M_{ij}|$ maximum absolute column sum

$\|M\|_{\infty\leftarrow\infty} = \max_i \sum_{j=1}^m |M_{ij}|$ maximum absolute row sum

$\|M\|_{2\leftarrow 2} = \sqrt{\lambda_{\max}(M^*M)}$ maximum singular value

The induced 2 norm can be understood as follows:

$$||M||_{2\leftarrow 2} = \max_{x\neq 0} \frac{||Mx||_2}{||x||_2} = \max_{x\neq 0} \sqrt{\frac{x^*M^*Mx}{\langle x, x\rangle^2}} = \sqrt{\lambda_{\max}(M^*M)}$$

Remark 9. When $p = q$ in (4), often the induced matrix norm is simply written as $||M||_p$.

2.3 Norm inequalities

1. Cauchy-Schwartz Inequality:

   $$|\langle x, y\rangle| \leq ||x||_2||y||_2$$

   which is the special case of the Holder inequality

   $$|\langle x, y\rangle| \leq ||x||_p||y||_q,\ \frac{1}{p} + \frac{1}{q} = 1,\ 1 \leq p, q \leq \infty \tag{5}$$

   Both bounds are tight: for certain choices of x and y, the inequalities become equalities.

2. Bounding induced matrix norms:

   $$||AB||_{l\leftarrow n} \leq ||A||_{l\leftarrow m}||B||_{m\leftarrow n} \tag{6}$$

   which comes from

   $$||ABx||_l \leq ||A||_{l\leftarrow m}||Bx||_m \leq ||A||_{l\leftarrow m}||B||_{m\leftarrow n}||x||_n$$

   In general, the bound is not tight. For instance, $||A^n|| = ||A||^n$ does not hold for $n \geq 2$ unless A has special structures.

3. (5) and (6) are useful for computing bounds of difficult-to-compute norms. As a special case of (6), we have

   $$||A||_2^2 \leq ||A||_1||A||_\infty$$

   Notice that $||A||_2^2$ is expensive to compute but $||A||_1$ and $||A||_\infty$ are not. We can obtain an upper bound of $||A||_2^2$ by computing $||A||_1||A||_\infty$.

4. Any matrix induced norms of A are larger than the absolute eigenvalues of A:

   $$|\lambda(A)| \leq ||A||_p$$

2.4 Frobenius norm and general matrix norms

Matrix norms do not have to be induced by vector norms. Similar to Definition 7, a general matrix norm is defined by satisfying the following three properties:

- $||A|| \geq 0$ and $||A|| = 0$ if and only if $A = 0$
- $||A + B|| \leq ||A|| + ||B||$
- $||\alpha A|| = |\alpha| \, ||A||$

The most important matrix norm which is not induced by a vector norm is the Frobenius norm, defined by

$$\|A\|_F \triangleq \sqrt{Tr\left(A^* A\right)} = \sqrt{< A, A >} = \sqrt{\sum_{i,j} |a_{i,j}|^2}$$

Frobenius norm is just the Euclidean norm of the matrix, written out as a long column vector:

$$||A||_F = \left(\mathrm{Tr}\left(A^* A\right)\right)^{\frac{1}{2}} = \left(\sum_{i=1}^{m} \sum_{j=1}^{m} |a_{i,j}|^2\right)^{\frac{1}{2}}$$

We also have bounds for Frobenius norms:

$$||AB||_F^2 \leq ||A||_F^2 ||B||_F^2$$

Fact 10. **Let x be a random vector with $\mathrm{E}\left[x\right] = 0$, $\mathrm{E}\left(xx^T\right) = I$, then*

$$\|A\|_F^2 = E\left[\|Ax\|_2^2\right]$$

2.5 Exercises

1. Let x be an m vector and A be an $m \times n$ matrix. Verify each of the following inequalities, and give an example when the equality is achieved.

   (a) $||x||_\infty \leq ||x||_2$

   (b) $||x||_2 \leq \sqrt{m}||x||_\infty$

   (c) $||A||_\infty \leq \sqrt{n}||A||_2$

   (d) $||A||_2 \leq \sqrt{m}||A||_\infty$

3 Symmetric, skew-symmetric, and orthogonal matrices

3.1 Definitions and basic properties

A real square matrix A is called **symmetric** if $A = A^T$; **skew-symmetric** if $A = -A^T$.

Fact 11. *Any real square matrix A may be written as the sum of a symmetric matrix R and a skew-symmetric matrix S, where*

$$R = \frac{1}{2}\left(A + A^T\right), \; S = \frac{1}{2}\left(A - A^T\right)$$

If $A = [a_{jk}]$, then the **complex conjugate** of A is denoted as$\overline{A} = [\overline{a}_{jk}]$, i.e., each element $a_{jk} = \alpha + i\beta$ is replaced with its complex conjugate $\overline{a}_{jk} = \alpha - i\beta$.

A square matrix A is called **Hermitian** if $A^T = \overline{A}$; **skew-Hermitian** if $A^T = -\overline{A}$.

Example 12. Find the symmetric, skew-symmetric, Hermitian, and skew-Hermitian matrices in the set:

$$\left\{ \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2i \\ 2i & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2i \\ -2i & 1 \end{bmatrix}, \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 2+2i \\ 2-2i & 0 \end{bmatrix} \right\}$$

We introduce one more class of important matrices: a real square matrix A is called **orthogonal**[1] if

$$A^T A = A A^T = I \tag{7}$$

Writing A in the column-vector notation

$$A = [a_1, a_2, \dots, a_n]$$

we get the equivalent form of (7):

$$A^T A = \begin{bmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} a_1, & a_2, & \dots, & a_n \end{bmatrix} = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \dots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \dots & a_2^T a_n \\ \vdots & \vdots & \vdots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \dots & a_n^T a_n \end{bmatrix} = I$$

Hence it must be that

$$a_j^T a_j = 1$$
$$a_j^T a_m = 0 \; \forall j \neq m$$

namely, a_j and a_m are orthonormal for any $j \neq m$.

The complex version of an orthogonal matrix is the **unitary matrix**. A square matrix A is called unitary if $A\overline{A}^T = \overline{A}^T A = I$, namely $A^{-1} = \overline{A}^T$.

Remark 13. Often the complex conjugate transpose $\overline{A}^T$ is written as A^*.

Theorem 14. $\forall : A \in \mathbb{R}^{n\times n}$ with $A^T = A$, *then eigenvalues of A are all real.*

[1] Some people also call use the notion of orthonormal matrix. But orthogonal matrix is more often used (we can say orthonormal basis).

Proof. $Au = \lambda u \Rightarrow \overline{u}^T Au = \lambda \overline{u}^T u$. $\overline{u}^T u$ is a real number (norm of u). $\overline{u}^T Au$ is also a real number, as $\overline{\overline{u}^T Au} = u^T \overline{A}\overline{u} = u^T A \overline{u} = u^T A^T \overline{u} = \lambda u^T \overline{u} = \lambda \overline{u}^T u = \overline{u}^T Au$. □

Theorem 15. $\forall : A \in \mathbb{R}^{n\times n}$ with $A^T = -A$, *then eigenvalues of A are all imaginary or zero.*

The proof is left as an exercise.

Fact 16. *An orthogonal transformation preserves the value of the inner product of vectors a and b in $\mathbb{R}^n$. That is, for any a and b in $\mathbb{R}^n$, orthogonal $n \times n$ matrix A, and $u = Aa$, $v = Ab$ we have $\langle u, v\rangle = \langle a, b\rangle$, as*

$$u^T v = a^T A^T Ab = a^T b$$

Hence the transformation also preserves the length or 2-norm of any vector a in $\mathbb{R}^n$ given by $||a||_2 = \sqrt{\langle a, a\rangle}$.

Theorem 17. *The determinant of an orthogonal matrix is either 1 or -1.*

Proof. $UU^T = I \Rightarrow \det U \det U^T = (\det U)^2 = 1$ □

Theorem 18. *The eigenvalues of an orthogonal matrix A are real or complex conjugates in pairs and have absolute value 1.*

Proof. $Au = \lambda u \Rightarrow A^T Au = \lambda A^T u \Rightarrow u = \lambda A^T u \Rightarrow \overline{u}^T u = \lambda \overline{u}^T A^T u \Rightarrow \overline{u}^T u = \lambda \overline{u}^T \overline{A}^T u = \lambda\overline{\lambda}\overline{u}^T u \Rightarrow \left(|\lambda|^2 - 1\right) \overline{u}^T u = 0$ □

Properties of the special matrices

real matrix	complex matrix	properties
symmetric ($A = A^T$)	Hermitian ($A^* = A$)	eigenvalues are all real
orthogonal ($A^T A = AA^T = I$)	unitary ($A^* A = AA^* = I$)	eigenvalues have unity magnitude; Ax maintains the 2-norm of x
skew-symmetric ($A^T = -A$)	skew-Hermitian ($A^* = -A$)	eigenvalues are all imaginary or zero

Based on the eigenvalue characteristics:

- symmetric and Hermitian matrices are like the real line in the complex domain
- skew-symmetric/Hermitian matrices are like the imaginary line
- orthogonal/unitary matrices are like the unit circle

Exercise 19 (Representation of matrices using special matrices)**.** Many unitary matrices can be mapped as functions of skew-Hermitian matrices as follows

$$U = (I - S)^{-1} (I + S)$$

where $S \neq I$. Show that if S is skew-Hermitian, then U is unitary.

3.2 Symmetric eigenvalue decomposition (SED)

When $A \in \mathbb{R}^{n\times n}$ has n distinct eigenvalues, we have seen the useful result of matrix diagonalization:

$$A = U\Lambda U^{-1} = [u_1, \ldots, u_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} [u_1, \ldots, u_n]^{-1} \tag{8}$$

where λ_i's are the distinct eigenvalues associated to the eigenvector u_i's.

The inverse matrix in (8) can be painful to compute though.

The spectral theorem, aka symmetric eigenvalue decomposition theorem,[2] significantly simplifies the result when A is symmetric.

Theorem 20. $\forall : A \in \mathbb{R}^{n\times n},\ A^T = A$, *there always exist* λ_i *and* u_i, *such that*

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^T = U\Lambda U^T \tag{9}$$

where:[3]

- λ_i*'s: eigenvalues of* A
- u_i*: eigenvector associated to* λ_i*, normalized to have unity norms*
- $U = [u_1, u_2, \cdots, u_n]^T$ *is an orthogonal matrix, i.e.,* $U^TU = UU^T = I$
- $\{u_1, u_2, \cdots, u_n\}$ *forms a orthonormal basis*
- $\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$

To understand the result, we show an important theorem first.

Theorem 21. $\forall : A \in \mathbb{R}^{n\times n}$ with $A^T = A$, *then eigenvectors of* A*, associated with different eigenvalues, are orthogonal.*

Proof. Let $Au_i = \lambda_i u_i$ and $Au_j = \lambda_j u_j$. Then $u_i^T A u_j = u_i^T \lambda_j u_j = \lambda_j u_i^T u_j$. In the meantime, $u_i^T A u_j = u_i^T A^T u_j = (Au_i)^T u_j = \lambda_i u_i^T u_j$. So $\lambda_i u_i^T u_j = \lambda_j u_i^T u_j$. But $\lambda_i \neq \lambda_j$. It must be that $u_i^T u_j = 0$. □

Theorem 20 now follows. If A has distinct eigenvalues, then $U = [u_1, u_2, \cdots, u_n]^T$ is orthogonal if we normalize all the eigenvectors to unity norm. If A has $r(< n)$ distinct eigenvalues, we can *choose* multiple orthogonal eigenvectors for the eigenvalues with none-unity multiplicities.

[2] Recall that the set of all the eigenvalues of A is called the spectrum of A. The largest of the absolute values of the eigenvalues of A is called the spectral radius of A.

[3] $u_i u_i^T \in \mathbb{R}^{n\times n}$ is a symmetric dyad, the so-called outerproduct of u_i and u_i. It has the following properties:

- $\forall\ v \in \mathbb{R}^{n\times 1}$, $\left(vv^T\right)_{ij} = v_i v_j$. (Proof: $\left(vv^T\right)_{ij} = e_i^T \left(vv^T\right) e_j = v_i v_j$, where e_i is the unit vector with all but the i_{th} elements being zero.)
- link with quadratic functions: $q(x) = x^T \left(vv^T\right) x = \left(v^T x\right)^2$

Observations:

- If we "walk along" u_j, then

$$Au_j = \left(\sum_i \lambda_i u_i u_i^T\right) u_j = \lambda_j u_j u_j^T u_j = \lambda_j u_j \tag{10}$$

where we used the orthonormal condition of $u_i^T u_j = 0$ if $i \neq j$. This confirms that u_j is an eigenvector.

- $\{u_i\}_{i=1}^n$ is a orthonormal basis $\Rightarrow \forall x \in \mathbb{R}^n$, $\exists\ x = \sum_i \alpha_i u_i$, where $\alpha_i =< x, u_i >$. And we have

$$Ax = A\sum_i \alpha_i u_i = \sum_i \alpha_i A u_i = \sum_i \alpha_i \lambda_i u_i = \sum_i (\alpha_i \lambda_i) u_i \tag{11}$$

which gives the (intuitive) picture of the geometric meaning of Ax: decompose first x to the space spanned by the eigenvectors of A, scale each components by the corresponding eigenvalue, sum the results up, then we will get the vector Ax.

With the spectral theorem, next time you see a symmetric matrix A, you should immediately know that

- λ_i is real for all i
- associated with λ_i, we can always find a real eigenvector
- $\exists$ an orthonormal basis $\{u_i\}_{i=1}^n$, which consists of the eigenvectors
- if $A \in \mathbb{R}^{2\times 2}$, then if you compute first λ_1, λ_2 and u_1, you won't need to go through the regular math to get u_2, but can simply solve for a u_2 that is orthogonal to u_1 with $\|u_2\| = 1$.

Example 22. Consider the matrix $A = \begin{bmatrix} 5 & \sqrt{3} \\ \sqrt{3} & 7 \end{bmatrix}$. Computing the eigenvalues gives

$$\det \begin{bmatrix} 5-\lambda & \sqrt{3} \\ \sqrt{3} & 7-\lambda \end{bmatrix} = 35 - 12\lambda + \lambda^2 - 3 = (\lambda - 4)(\lambda - 8) = 0$$
$$\Rightarrow \lambda_1 = 4,\ \lambda_2 = 8$$

And we can know one of the eigenvectors from

$$(A - \lambda_1 I)\, t_1 = 0 \Rightarrow \begin{bmatrix} 1 & \sqrt{3} \\ \sqrt{3} & 3 \end{bmatrix} t_1 = 0 \Rightarrow t_1 = \begin{bmatrix} -\frac{\sqrt{3}}{2} \\ \frac{1}{2} \end{bmatrix}$$

Note here we normalized t_1 such that $||t_1||_2 = 1$. With the above computation, we no more need to do $(A - \lambda_2 I)\, t_2 = 0$ for getting t_2. Keep in mind that A here is symmetric, so has eigenvectors orthogonal to each other. By direct observation, we can see that

$$x = \begin{bmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{bmatrix}$$

is orthogonal to t_1 and $||x||_2 = 1$. So $t_2 = x$.

Theorem 23 (Eigenvalues of symmetric matrices)**.** *If $A = A^T \in \mathbb{R}^{n\times n}$, then the maximum eigenvalue of A satisfies*

$$\lambda_{\max} = \max_{x\in\mathbb{R}^n,\ x\neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{12}$$

$$\lambda_{\min} = \min_{x\in\mathbb{R}^n,\ x\neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{13}$$

Proof. Perform SED to get

$$A = \sum_{i=1}^{n} \lambda_i u_i^T u_i$$

where $\{u_i\}_{i=1}^n$ form a basis of $\mathbb{R}^n$. Then any vector $x \in \mathbb{R}^n$ can be decomposed as

$$x = \sum_{i=1}^{n} \alpha_i u_i$$

Thus

$$\max_{x\neq 0} \frac{x^T A x}{\|x\|_2^2} = \max_{\alpha_i} \frac{\left(\sum_i \alpha_i u_i\right)^T \sum_i \lambda_i \alpha_i u_i}{\sum_i \alpha_i^2} = \max_{\alpha_i} \frac{\sum_i \lambda_i \alpha_i^2}{\sum_i \alpha_i^2} = \lambda_{\max}$$

The proof for (13) is analogous and omitted. □

3.3 Symmetric positive-definite matrices

Definition 24. A symmetric matrix $P \in \mathbb{R}^{n\times n}$ is called **positive-definite**, written $P \succ 0$, if $x^T P x > 0$ for all $x\,(\neq 0) \in \mathbb{R}^n$. P is called **positive-semidefinite,** written $P \succeq 0$, if $x^T P x \geq 0$ for all $x \in \mathbb{R}^n$

□

Definition 25. A symmetric matrix $P \in \mathbb{R}^{n\times n}$ is called **negative-definite**, written $P \prec 0$, if $-P \succ 0$, i.e., $x^T P x < 0$ for all $x\,(\neq 0) \in \mathbb{R}^n$. P is called **negative-semidefinite,** written $P \preceq 0$, if $x^T P x \leq 0$ for all $x \in \mathbb{R}^n$

When A and B have compatible dimensions, $A \succ B$ means $A - B \succ 0$.

Positive-definite matrices can have negative entries, as shown in the next example.

Example 26. The following matrix is positive-definite

$$P = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

as $P = P^T$ and take any $v = [x, y]^T$, we have

$$v^T P v = \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2x^2 + 2y^2 - 2xy = x^2 + y^2 + (x+y)^2 \geq 0$$

and the equality sign holds only when $x = y = 0$.

Conversely, matrices whose entries are all positive are not necessarily positive-definite.

Example 27. The following matrix is not positive-definite

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

as

$$\begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 < 0$$

Theorem 28. *For a symmetric matrix P, $P \succ 0$ if and only if all the eigenvalues of P are positive.*

Proof. Since P is symmetric, we have

$$\lambda_{\max}(P) = \max_{x \in \mathbb{R}^n,\ x \neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{14}$$

$$\lambda_{\min}(P) = \min_{x \in \mathbb{R}^n,\ x \neq 0} \frac{x^T A x}{\|x\|_2^2} \tag{15}$$

which gives

$$x^T A x \in \left[\lambda_{\min}||x||_2^2,\ \lambda_{\max}||x||_2^2\right]$$

For $x \neq 0$, $||x||_2^2$ is always positive. It can thus be seen that $x^T A x > 0,\ x \neq 0 \Leftrightarrow \lambda_{\min} > 0$. □

Lemma. *For a symmetric matrix P, $P \succeq 0$ if and only if all the eigenvalues of P are none-negative.*

Theorem. *If A is symmetric positive definite, X is full column rank, then $X^T A X$ is positive definite.*

Proof. Consider $y\left(X^T A X\right) y = x^T A x$ which is always positive unless $x = 0$. But X is full rank so $Xy = x = 0$ if and only if $y = 0$. This proves $X^T A X$ is positive definite. □

Fact. *All principle submatrices of A are positive definite.*

Proof. Use the last theorem. Take $X = e_1$, $X = [e_1, e_2]$, etc. Here e_i is a column vector whose ith-entry is 1 and all other entries are zero. □

Example 29. The following matrices are all not positive definite:

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 1 \\ 1 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

Positive-definite matrices are like positive real numbers. We can have the concept of *square root* of positive-definite matrices.

Definition 30. Let $P \succeq 0$. We can perform symmetric eigenvalue decomposition to obtain $P = UDU^T$ where U is orthogonal with $UU^T = I$ and D is diagonal with all diagonal elements being none negative

$$D = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} \succeq 0$$

Then the square root of P, written $P^{\frac{1}{2}}$, is defined as

$$P^{\frac{1}{2}} = UD^{\frac{1}{2}}U^T$$

where

$$D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\lambda_n} \end{bmatrix}$$

3.4 General positive-definite matrices

Definition 31. A general square matrix $Q \in \mathbb{R}^{n\times n}$ is called positive-definite, written as $Q \succ 0$, if $x^TQx > 0 \; \forall x \neq 0$.

We have discussed the case when Q is symmetric. If not, recall that any real square matrix can be decomposed as the sum of a symmetric matrix and a skew symmetric matrix:

$$Q = \frac{Q+Q^T}{2} + \frac{Q-Q^T}{2}$$

where $\frac{Q+Q^T}{2}$ is symmetric.

Notice that $x^T\frac{Q-Q^T}{2}x = x^TQx - \left(x^TQx\right)^T = 0$. So for a general square real matrix:

$$Q \succ 0 \Leftrightarrow Q + Q^T > 0$$

Example 32. The following matrices are positive definite but not symmetric

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

For complex matrices with $Q = Q^* = Q_R + jQ_I$, we have

$$\begin{aligned} Q \succ 0 &\Leftrightarrow x^*Qx > 0, \; \forall x \neq 0 \\ &\Leftrightarrow \left(x_R^T - jx_I^T\right)(Q_R + jQ_I)(x_R + jx_I) > 0 \\ &\Leftrightarrow \begin{pmatrix} x_R \\ x_I \end{pmatrix}^T \begin{pmatrix} 1 \\ j \end{pmatrix} \begin{pmatrix} Q_R & Q_I \end{pmatrix} \begin{pmatrix} 1 \\ j \end{pmatrix} \begin{pmatrix} 1 \\ j \end{pmatrix}^T \begin{pmatrix} x_R \\ x_I \end{pmatrix} \\ &\Leftrightarrow \begin{pmatrix} x_R \\ x_I \end{pmatrix}^T \begin{pmatrix} Q_R & Q_I \\ -Q_I & Q_R \end{pmatrix} \begin{pmatrix} x_R \\ x_I \end{pmatrix} > 0 \\ &\Leftrightarrow x_R^TQ_Rx_R - x_I^TQ_Ix_R + x_R^TQ_Ix_I + x_I^TQ_Rx_I > 0 \end{aligned}$$

4 Singular value and singular value decomposition (SVD)

4.1 Motivation

Symmetric eigenvalue decomposition is great but many matrices are not symmetric. A general matrix A may actually not even be square. Singular value decomposition is an important matrix decomposition technique that works for arbitrary matrices.[4]

For a general none-square matrix $A \in \mathbb{C}^{m\times n}$, eigenvalues and eigenvectors are generalized to

$$Av_j = \sigma_j u_j \tag{16}$$

Be careful about the dimensions: if $m > n$, we have

$$\begin{bmatrix} & \vdots & \\ \cdots & A & \cdots \\ & \vdots & \end{bmatrix} \underbrace{\begin{bmatrix} v_1 \,\big|\, v_2 \,\big|\, \cdots \,\big|\, v_n \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} u_1 \,\big|\, u_2 \,\big|\, \cdots \,\big|\, u_n \end{bmatrix}}_{\hat{U}} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}}_{\hat{\Sigma}}$$

It turns out that, if A has full column rank n, then we can find a V that is unitary ($VV^* = V^*V = I$) and a $\hat{U}$ that has orthonormal columns. Hence

$$A = \hat{U}\hat{\Sigma}V^* \tag{17}$$

4.2 SVD

(17) forms the so-called reduced singular value decomposition (SVD). The idea of a "full" SVD is as follows. The columns of $\hat{U}$ are n orthonormal vectors in the m-dimensional space $\mathbb{C}^m$. They do not form a basis for $\mathbb{C}^m$ unless $m = n$. We can add additional $m - n$ orthonormal columns to $\hat{U}$ and augment it to a unitary matrix U. Now the matrix dimension has changed, $\hat{\Sigma}$ needs to be augmented to compatible dimensions as well. To maintain the equality (17), the newly added elements to $\hat{\Sigma}$ are set to zero.

Theorem 33. *Let $A \in \mathbb{C}^{m\times n}$ with rank r. Then we can find orthogonal matrices $U \in \mathbb{C}^{m\times m}$ and $V \in \mathbb{C}^{n\times n}$ such that*

$$A = U\Sigma V^*$$

where

$$\begin{aligned} \Sigma &\in \mathbb{R}^{m\times n} \text{ is diagonal} \\ U &\in \mathbb{C}^{m\times m} \text{ is unitary} \\ V &\in \mathbb{C}^{n\times n} \text{ is unitary} \end{aligned}$$

In addition, the diagonal entries σ_j of Σ are nonnegative and in nonincreasing order; that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

[4]History of SVD: discovered between 1873 and 1889, independently by several pioneers; did not became widely known in applied mathematics until the late 1960s, when it was shown that SVD can be computed effectively and used as the basis for solving many problems.

Proof. Notice that A^*A is positive semi-definite. Hence, A^*A has a full set of orthonormal eigenvectors; its eigenvalues are real and nonnegative. Order these eigenvalues as

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$$

[5]Let $\{v_1, \ldots, v_n\}$ be an orthonormal choice of eigenvectors of A^*A corresponding to these eigenvalues:

$$A^*Av_i = \lambda_i v_i$$

Then,

$$||Av_i||^2 = v_i^* A^* A v_i = \lambda_i v_i^* v_i = \lambda_i$$

For $i > r$, it follows that $Av_i = 0$.

For $1 \leq i \leq r$, we have

$$A^*Av_i = \lambda_i v_i$$

Recall (16), we define $\sigma_i = \sqrt{\lambda_i}$ and get

$$\begin{aligned} Av_i &= \sigma_i u_i \\ A^* u_i &= \sigma_i v_i \end{aligned}$$

For $1 \leq i, j \leq r$, we have

$$\langle u_i, u_j \rangle = u_i^* u_j = \frac{1}{\sigma_i \sigma_j} v_i^* A^* A v_j = \frac{1}{\sigma_i \sigma_j} \lambda_j v_i^* v_j = \frac{\sigma_j}{\sigma_i} v_i^* v_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Hence $\{u_1, \ldots, u_r\}$ is an orthonormal set of eigenvectors. Extending this set to form an orthonormal basis for $\mathbb{C}^m$ gives

$$U = \left[\begin{array}{ccc|ccc} u_1, & \ldots, & u_r & u_{r+1}, & \ldots, & u_m \end{array}\right]$$

For $i \leq r$, we already have

$$Av_i = \sigma_i u_i$$

namely

$$A\left[v_1, \ldots v_r\right] = \left[u_1, \ldots, u_r\right] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} = \left[\begin{array}{ccc|ccc} u_1, & \ldots, & u_r & u_{r+1}, & \ldots, & u_m \end{array}\right] \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{bmatrix}$$

[5]Fact: $\text{rank}\,(A) = \text{rank}\,(A^*A)$. To see this, notice first, that $\text{rank}\,(A) \geq \text{rank}\,(A^*A)$ by definition of rank. Second, $A^*Ax = 0 \Rightarrow x^*A^*Ax = 0 \Rightarrow ||Ax|| = 0 \Rightarrow Ax = 0$, hence $\text{rank}\,(A) \leq \text{rank}\,(A^*A)$.

For $v_{r+1}, \ldots$, we have already seen that $Av_{r+1} = Av_{r+2} = \cdots = 0$, hence

$$A \underbrace{\left[v_1, \ldots v_r | v_{r+1}, \ldots, v_n\right]}_{n\times n} = \underbrace{\left[\begin{array}{ccc|ccc} u_1, & \ldots, & u_r & u_{r+1}, & \ldots, & u_m \end{array}\right]}_{m\times m} \underbrace{\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_r & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ & & & & & \vdots \\ & & & & & 0 \end{bmatrix}}_{m\times n}$$

$$\Rightarrow A = U\Sigma V^*$$

□

Theorem 34. *The range space of A is spanned by $\{u_1, \ldots, u_r\}$. The null space of A is spanned by $\{v_{r+1}, \ldots, v_n\}$.*

□

Theorem 35. *The nonzero singular values of A are the square roots of the nonzero eigenvalues of A^*A or AA^*.*

□

Theorem 36. $||A||_2 = \sigma_1$*, i.e., the induced two norm of A is the maximum singular value of A.*

The next important theorem can be easily proved via SVD.

Theorem (Fundermental theory of linear algebra)**.** *Let $A \in \mathbb{R}^{m\times n}$. Then*

$$\mathcal{R}(A) + \mathcal{N}\left(A^T\right) = \mathbb{R}^m$$

and

$$\mathcal{R}(A) \perp \mathcal{N}\left(A^T\right)$$

Proof. By singular value decomposition

$$A = U\Sigma V^T$$
$$A^T = V\Sigma U^T$$

Range of A is the first r columns of U, from the first equation; Null space of A^T is the last $m-r$ columns of U, from the second equation. □

New intuition of matrix vector operation With $A = U\Sigma V^*$, a new intuition for $Ax = U\Sigma V^*x$ is formed. Since V is unitary, it is norm-preserving, in the sense that V^*x does not change the 2-norm of the vector x. In other words, V^*x only rotates x in $\mathbb{C}^n$. The diagonal matrix Σ then functions to scale (by its diagonal values) the rotated vector. Finally, U is another rotation in $\mathbb{C}^m$.

4.3 Properties of singular values

Fact. *Let A and B be matrices with compatible dimensions. The following are true*
$\overline{\sigma}(A+B) \leq \overline{\sigma}(A) + \overline{\sigma}(B)$
$\overline{\sigma}(AB) \leq \overline{\sigma}(A)\overline{\sigma}(B)$

Proof. The first inequality comes from

$$\overline{\sigma}(A+B) = \max_{v\neq 0} \frac{||Av+Bv||_2}{||v||_2} \leq \max_{v\neq 0} \frac{||Av||_2 + ||Bv||_2}{||v||_2}$$

The second inequality uses

$$\overline{\sigma}(AB) = \max_{v\neq 0} \frac{||ABv||_2}{||v||_2} \leq \max_{v\neq 0} \frac{||A||_2||Bv||_2}{||v||_2}$$

□

4.4 Exercises

1. Compute the singular values of the following matrices

$$(a) \begin{bmatrix} 3 & \\ & -2 \end{bmatrix}, \ (b) \begin{bmatrix} 2 & \\ & 3 \end{bmatrix}, \ (c) \begin{bmatrix} 0 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \ (d) \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \ (e) \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

2. Show that if A is real, then it has a real SVD (i.e., U and V are both real).

3. For any matrix $A \in \mathbb{R}^{n\times m}$, construct

$$M = \begin{bmatrix} \overbrace{0}^{n\times n} & \overbrace{A}^{n\times m} \\ \underbrace{A^T}_{m\times n} & \underbrace{0}_{m\times m} \end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$$

which satisfies

$$M^T = M$$

M is Hermitian, and hence has real eigenvalues and eigenvectors:

$$\begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} u_j \\ v_j \end{bmatrix} = \sigma_j \begin{bmatrix} u_j \\ v_j \end{bmatrix} \tag{18}$$

(a) Show that

i. v_j is in the co-kernal (perpendicular to kernal/null space) of A and u_j is in the range of A.

ii. if σ_j and $\begin{bmatrix} u_j \\ v_j \end{bmatrix}$ form a eigen pair for M, then$-\sigma_j$ and $\left[u_j^T, -v_j^T\right]^T$ also form an eigen pair for M

iii. eigenvalues of M always appear in pairs that are symmetric to the imaginary axis.

(b) Use the results to show that, if

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 4 & 32 \end{bmatrix}$$

then M must have eigenvalues that are equal to 0.

4. Suppose $A \in \mathbb{C}^{m\times m}$ and has an SVD $A = U\Sigma V^*$. Find an eigenvalue decomposition of

$$\begin{bmatrix} 0 & A^* \\ A & 0 \end{bmatrix}$$

5. **Worst input direction** in matrix vector multiplications. Recall that any matrix defines a linear transformation:

$$Mw = z$$

Question: what is the worst input direction for the vector w? Here *worst* means: if we fix the input norm, say $\|w\| = 1$, $\|z\|$ will reach a maximum value (the worst case) for specific input direction defined by w.

(a) Show that the worst $||z||$ is $||M||$ when $||w|| = 1$.

(b) Provide procedures to obtain the w that gives the maximum $||z||$, for the cases of 1 norm, ∞ norm, and 2 norm.