ECE285: Semidefinite and sum-of-squares optimization **Winter 2024**

Lecture 2: Mathematical Background

Lecturer: Yang Zheng *Scribe: Yang Zheng*

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. Any typos should be sent to* `zhengy@eng.ucsd.edu`.

Learning goals:

1. Inner products and norms
2. Dual norms
3. Positive semidefinite matrices

In this lecture, we review some basic concepts from linear algebra, analysis, and positive semidefinite matrices.

1 Inner products, and norms

We work with n-dimensional Euclidean space $\mathbb{R}^n$. The elements in $\mathbb{R}^n$ are called points or vectors of dimension n. We also work with the space of real $m \times n$ matrices $\mathbb{R}^{m\times n}$. Each element (matrix) in $\mathbb{R}^{m\times n}$ can be viewed as a collection of n vectors in $\mathbb{R}^m$. Meanwhile, a matrix in $\mathbb{R}^{m\times n}$ can be seen as a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^m$.

1.1 Inner products

Definition 2.1 (Inner product)**.** *A function $\langle\cdot,\cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is called an inner product if*

1. $\langle x, x\rangle \geq 0$ *and* $\langle x, x\rangle = 0 \Rightarrow x = 0$ *(positivity);*
2. $\langle x, y\rangle = \langle y, x\rangle, \forall x, y \in \mathbb{R}^n$ *(symmetry);*
3. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle, \forall x, y, z \in \mathbb{R}^n$ *(additivity);*
4. $\langle cx, y\rangle = c\langle x, y\rangle, \forall x, y \in \mathbb{R}^n, c \in \mathbb{R}$ *(homogeneity).*

The properties of additivity and homogeneity in the second argument follows from the symmetry property, i.e., $\langle x, y + z\rangle = \langle y + z, x\rangle = \langle y, x\rangle + \langle z, x\rangle = \langle x, y\rangle + \langle x, z\rangle, \forall x, y, z \in \mathbb{R}^n$, and $\langle x, cy\rangle = \langle cy, x\rangle = c\langle y, x\rangle = c\langle x, y\rangle, \forall x, y \in \mathbb{R}^n, c \in \mathbb{R}$. Property (4) indicates that $\langle 0, x\rangle = 0 \times \langle 0, x\rangle = 0, \forall x \in \mathbb{R}^n$.

The standard inner product on $\mathbb{R}^n$ is

$$\langle x, y\rangle = x^{\mathsf{T}} y = \sum_{i=1}^{n} x_i y_i, \qquad \forall x, y \in \mathbb{R}^n.$$

The standard inner product on $\mathbb{R}^{m\times n}$ is

$$\langle X, Y\rangle = \operatorname{trace}(X^{\mathsf{T}} Y) = \sum_{j=1}^{n}\sum_{i=1}^{m} X_{ij} Y_{ij}, \qquad \forall X, Y \in \mathbb{R}^{m\times n},$$

where trace(Z) denotes the trace of a real matrix $Z \in \mathbb{R}^{n\times n}$, i.e., trace$(Z) = \sum_{i=1}^n Z_{ii}$. Note that the inner product of two matrices in $\mathbb{R}^{m\times n}$ is the same as the standard inner product between two vectors of length mn obtained by stacking the columns of the two matrices: denoting these two vectors as vec$(X) \in \mathbb{R}^{mn}$ and vec$(Y) \in \mathbb{R}^{mn}$, we have $\langle X, Y\rangle = \text{vec}(X)^\mathsf{T}\text{vec}(Y)$.

Example 2.1 (Weighted inner products)**.** *There are many other inner products on $\mathbb{R}^n$. For example, any symmetric matrix Q with positive eigenvalues (which is a positive definite matrix; see Section 2) can be used to define an inner product $\langle x, y\rangle = x^\mathsf{T}Qy$. This is known as a weighted inner product. A concrete example on $\mathbb{R}^2$ is*

$$Q = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \Rightarrow \langle x, y\rangle = 2x_1y_1 - x_1y_2 - x_2y_1 + 2x_2y_2.$$

Properties (2), (3), (4) in Definition 2.1 are obvious. We verify the positivity as follows

$$\langle x, x\rangle = 2x_1^2 - 2x_1x_2 + 2x_2^2 = (x_1 - x_2)^2 + x_1^2 + x_2^2 \geq 0, \forall x \in \mathbb{R}^2,$$
$$\langle x, x\rangle = 0 \Leftrightarrow (x_1 - x_2)^2 = 0, x_1^2 = 0, x_2^2 = 0 \Leftrightarrow x = 0.$$

The usual inner product is same as choosing $Q = I$. □

Two vectors x and y are called orthogonal if $\langle x, y\rangle = 0$. Given any inner product, we define the length of $x \in \mathbb{R}^n$ as $\|x\| = \sqrt{\langle x, x\rangle}$ (which is a norm; see Lemma 2.2). For the usual inner product, this is also called the Euclidean norm, or l_2 norm, denoted as $\|x\|_2 = \sqrt{x^\mathsf{T}x} = (x_1^2 + \ldots + x_n^2)^{1/2}$.

Lemma 2.1 (Cauchy–Schwarz inequality)**.** *For any $x, y \in \mathbb{R}^n$, we have*

$$|\langle x, y\rangle| \leq \|x\|\|y\|. \tag{1}$$

The equality is achieved if and only if x and y are linearly dependent, i.e., there exist $\alpha, \beta \in \mathbb{R}$ not all zero such that $\alpha x + \beta y = 0$.

Proof. Given any $x, y \in \mathbb{R}^n$, define a real function $p(t) = \langle tx+y, tx+y\rangle$. By properties (2)-(4) in Definition 2.1, it is easy to verify

$$p(t) = t^2\|x\|^2 + 2t\langle x, y\rangle + \|y\|^2,$$

which is a quadratic function in t. By property (1), we know $p(t) \geq 0, \forall t \in \mathbb{R}$. Thus, the discriminant satisfies $\Delta := 4\langle x, y\rangle^2 - 4\|x\|^2\|y\|^2 \leq 0$. This is the same as (1).

It remains to prove that $\Delta = 0$ is equivalent to the condition that x and y are linearly dependent. If x or y or both are zero, we obviously have $|\langle x, y\rangle| = 0 = \|x\|\|y\|$. Suppose they are both nonzero, we can write $y = t_0x$, then $|\langle x, y\rangle| = |t_0|\|x\|^2 = \|x\|\|y\|$. Conversely, if $\Delta = 0$, we have $\langle x, y\rangle = \|x\|\|y\|$ or $\langle x, y\rangle = -\|x\|\|y\|$. We focus on the former case (the latter case follows the same argument), in which we have $p(t) = (t\|x\| + \|y\|)^2$. If $x \neq 0$, we let $t_0 = \|y\|/\|x\|$, which leads to

$$0 = p(-t_0) = \langle -t_0x + y, -t_0x + y\rangle \;\Rightarrow\; y = t_0x.$$

If $x = 0$, then x and y are obviously linearly dependent. We now complete the proof. □

We note that Cauchy–Schwarz inequality holds for any inner product $\langle x, y\rangle$ and its induced length $\|x\| = \sqrt{\langle x, x\rangle}$. As in Example 2.1, given any symmetric matrix Q with positive eigenvalues, we have

$$|x^\mathsf{T}Qy| \leq \sqrt{x^\mathsf{T}Qx}\sqrt{y^\mathsf{T}Qy}, \;\; \forall x, y \in \mathbb{R}^n.$$

1.2 Vector norms

We define a generalized length of a vector $x \in \mathbb{R}^n$ below.

Definition 2.2 (Vector Norm)**.** *A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a norm if*

1. $f(x) \geq 0, \forall x \in \mathbb{R}^n$ *and* $f(x) = 0 \Leftrightarrow x = 0$ *(positivity);*
2. $f(\lambda x) = |\lambda| f(x), \forall x \in \mathbb{R}^n, \lambda \in \mathbb{R}$ *(homogeneity);*
3. $f(x + y) \leq f(x) + f(y)$ *(triangle inequality).*

We use the notation $f(x) = \|x\|$ to denote a norm. When specifying a particular norm, we use $\|x\|_{\text{symb}}$ in which the subscript indicates which norm is meant. Note that property (2) indicates $\|-x\| = \|x\|$, and property (3) also implies that $\|x - y\| \leq \|x\| + \|y\|$.

A simple example of a norm is the Euclidean or l_2-norm, defined above in Section 1.1. Another two frequently used norms on $\mathbb{R}^n$ are

- the *sum-absolute-value*, or l_1-norm, defined as $\|x\|_1 = |x_1| + \cdots + |x_n|$,
- the *Chebyshev* or l_∞-norm, defined as $\|x\|_\infty = \max\{|x_1|, \ldots, |x_n|\}$.

These three norms are part of a family of so-called l_p-norm (with $p \geq 1$), defined as

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}. \tag{2}$$

When $p = 1$, we get the l_1 norm, and when $p = 2$, we get the Euclidean or l_2-norm. It is easy to see when $p \to \infty$, we get the l_∞ norm. Note that if $0 < p < 1$, then (2) does not define a norm. In particular, the triangle property does not hold. For example, let $x = [1, 0]^\mathsf{T}$ and $y = [0, 1]^\mathsf{T}$, we have

$$\|x + y\|_p = 2^{1/p} > 2 = \|x\|_p + \|y\|_p, \ \text{ if } 0 < p < 1.$$

In many applications, we are interested in the number of non-zero elements in $x \in \mathbb{R}^n$: defining $0^0 = 0$, the "zero or l_0 norm" of x is $\|x\|_0 = |x_1|^0 + \cdots + |x_n|^0$. Again, this "$l_0$ norm" is not a norm because it is not homogeneous (we have $\|\lambda x\|_0 = \|x\|_0, \forall \lambda \neq 0$).

In addition to l_p-norm, any inner product on $\mathbb{R}^n$ can be used to define a norm.

Lemma 2.2. *Given any inner product $\langle \cdot, \cdot \rangle$, let $f(x) = \sqrt{\langle x, x \rangle}$. Then, $f(x)$ is a norm.*

Proof. Positivity follows from Definition 2.1(1). Homogeneity can be seen from

$$f(\lambda x) = \sqrt{\langle \lambda x, \lambda x \rangle} = \sqrt{\lambda^2 \langle x, x \rangle} = |\lambda| f(x).$$

The triangle inequality is from Cauchy-Schwarz inequality. Let $\|x\| = \sqrt{\langle x, x \rangle}$. We have

$$\begin{aligned} \langle x, y \rangle \leq \|x\| \|y\| \ &\Rightarrow 2\langle x, y \rangle + \langle x, x \rangle + \langle y, y \rangle \leq 2\|x\|\|y\| + \langle x, x \rangle + \langle y, y \rangle \\ &\Rightarrow \langle x + y, x + y \rangle \leq \left(\sqrt{\langle x, x \rangle} + \sqrt{\langle y, y \rangle} \right)^2 \\ &\Rightarrow f(x + y) \leq f(x) + f(y). \end{aligned}$$

This completes the proof. □

Example 2.2 (Quadratic norms)**.** *Similar to Example 2.1, an important family of norms is the quadratic norms. Given any symmetric matrix Q with positive eigenvalues, we define the Q-quadratic norm as*

$$\|x\|_Q = \sqrt{x^\mathsf{T} Q x}.$$

By the definition of symmetric square root $Q^{1/2}$ (see Section 2.1), it is easy to see that the Q-quadratic norm is related to the l_2 norm as $\|x\|_Q = \|Q^{1/2} x\|_2$. □

If a norm $\|x\|$ is induced by an inner product $\langle\cdot,\cdot\rangle$, then we have the *parallelogram identity*

$$2\|x\|^2 + 2\|y\|^2 = \|x+y\|^2 + \|x-y\|^2, \quad \forall x, y \in \mathbb{R}^n. \tag{3}$$

This can be easily verified by observing that

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2,$$
$$\|x-y\|^2 = \langle x-y, x-y\rangle = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2.$$

A remarkable fact is that any norm satisfying the parallelogram law (3) arises from an inner product [1].

Theorem 2.1. *A norm $\|x\|$ is induced by an inner product $\|x\| = \sqrt{\langle x, x\rangle}$ if and only if the parallelogram identity* (3) *holds.*

When (3) holds, the inner product can be constructed as $\langle x, y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$. We note that not every norm comes from an inner product. For example, for the l_p norm (2), only when $p = 2$, it arises from the usual inner product $\langle x, y\rangle = x^\mathsf{T} y$. The l_1 and l_∞ norms do not come from any inner product. It is easy to find examples for which (3) fails for l_1 and l_∞ norms. The following norm is a mixture between l_1 and l_∞ norms: letting $k \in \{1, \ldots, n\}$, we define $\|x\|_{1,k} = \sum_{i=1}^k |x|_{[i]}$, where $|x|_{[i]}$ denote the i-th largest absolute value of elements of $x \in \mathbb{R}^n$. This norm $\|x\|_{1,k}$ also makes (3) fail.

1.3 Matrix norms

Matrix norms are functions $f : \mathbb{R}^{m\times n} \to \mathbb{R}$ that satisfy the same properties (positivity, homogeneity, triangle inequality) of the vector norms in Definition 2.2. Three common norms on $\mathbb{R}^{m\times n}$ are element-wise defined as follows

- the Frobenius norm $\|A\|_\mathrm{F} = \sqrt{\mathrm{trace}(A^\mathsf{T} A)} = \big(\sum_{j=1}^n \sum_{i=1}^m A_{ij}^2\big)^{1/2}$;
- the sum-absolute-value norm $\|A\|_\mathrm{sav} = \sum_{j=1}^n \sum_{i=1}^m |A_{ij}|$;
- the max-absolute-value norm $\|A\|_\mathrm{mav} = \max_{i=1,\ldots,m;j=1,\ldots,n} |A_{ij}|$.

These three norms coincide with the Euclidean, l_1, l_∞ norms of $\mathrm{vec}(A) \in \mathbb{R}^{mn}$ respectively. However, it should be careful that the l_p (with $p = 1, 2, \infty$) norm of a matrix is different from its element-wise defined version $\|\mathrm{vec}(A)\|_p$, as we define it below.

Another frequently used class of matrix norms are the operator (or induced) norms.

Definition 2.3 (Operator norms)**.** *An operator or induced norm $\|\cdot\|_\mathrm{a,b} : \mathbb{R}^{m\times n} \to \mathbb{R}$ is defined as*

$$\|A\|_\mathrm{a,b} = \max_{\|x\|_\mathrm{b}\le 1} \|Ax\|_\mathrm{a}, \tag{4}$$

where $\|\cdot\|_\mathrm{a}$ is a vector norm on $\mathbb{R}^m$ and $\|\cdot\|_\mathrm{b}$ is a vector norm on $\mathbb{R}^n$.

It can be verified that (4) defines a norm on $\mathbb{R}^{m\times n}$ (positivity and homogeneity are obvious, and triangle inequality comes from $\|(A+B)x\|_\mathrm{a} \le \|Ax\|_\mathrm{a} + \|Bx\|_\mathrm{a}$).

When the same vector norm is used on both spaces $\mathbb{R}^m$ and $\mathbb{R}^n$, we write

$$\|A\|_\mathrm{c} = \max_{\|x\|_\mathrm{c}\le 1} \|Ax\|_\mathrm{c}.$$

Accordingly, the l_p norm on vectors lead to the l_p operator norm on matrices $\mathbb{R}^{m\times n}$:

[1] https://math.stackexchange.com/questions/21792/norms-induced-by-inner-products-and-the-parallelogram-law

- If both $\|\cdot\|_{\rm a}$ and $\|\cdot\|_{\rm b}$ are the l_2 norm, the operator norm of A is its maximum singular value:
$$\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)},$$
where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of a symmetric matrix. This norm is also called the *spectral norm* or l_2 norm of $A \in \mathbb{R}^{m\times n}$.

- If both $\|\cdot\|_{\rm a}$ and $\|\cdot\|_{\rm b}$ are the l_1 norm, (4) leads to the *max-column-sum* norm
$$\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^{m} |A_{ij}|,$$

- If both $\|\cdot\|_{\rm a}$ and $\|\cdot\|_{\rm b}$ are the l_∞ norm, we get the *max-row-sum* norm
$$\|A\|_\infty = \max_{i=1,\dots,m} \sum_{j=1}^{n} |A_{ij}|.$$

Lemma 2.3. *Every induced norm is submultiplicative, i.e.,*
$$\|AB\|_{\rm c} \le \|A\|_{\rm c}\|B\|_{\rm c}, \;\; \forall A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$$

Proof. By the definition $\|A\|_{\rm c} = \max_{\|x\|_{\rm c}\le 1} \|Ax\|_{\rm c}$, we have $\|Ax\|_{\rm c} \le \|A\|_{\rm c}\|x\|_{\rm c}, \forall x \in \mathbb{R}^n$. Therefore, the following inequality holds
$$\begin{aligned}\|AB\|_{\rm c} = \max_{\|x\|_{\rm c}\le 1} \|ABx\|_{\rm c} &\le \max_{\|x\|_{\rm c}\le 1} \|A\|_{\rm c}\|Bx\|_{\rm c}\\ &= \|A\|_{\rm c} \max_{\|x\|_{\rm c}\le 1} \|Bx\|_{\rm c} = \|A\|_{\rm c}\|B\|_{\rm c}.\end{aligned}$$

This completes the proof. □

Lemma 2.3 is true for operator norms induced by the same vector norm in all three spaces $\mathbb{R}^m, \mathbb{R}^n$ and $\mathbb{R}^p$. Otherwise, the submultiplicative property may fail.

Example 2.3. *Not every matrix norm is an induced norm. For example, the Frobenius norm is not an induced norm. This is because for any induced norm, we have $\|I\|_{\rm c} = 1$, while $\|I\|_{\rm F} = \sqrt{n}$.*

Not all matrix norms are submultiplicative. For example, the max-absolute-value norm $\|A\|_{mav}$ does not satisfy the submultiplicative property. Consider
$$A = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}, \;\; A^2 = \begin{bmatrix}2 & 2\\ 2 & 2\end{bmatrix}.$$

Then, $\|A^2\|_{mav} = 2 > 1 = \|A\|_{mav}^2$.

Finally, not all submultiplicative norms are induced norms. An example is the Frobenius norm for which we have
$$\|AB\|_{\rm F} \le \|A\|_{\rm F}\|B\|_{\rm F}, \;\; \forall A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}. \tag{5}$$
This inequality comes from the Cauchy-Schwarz inequality (Lemma 2.1). □

Note that $\lambda_{\max}(A^\top A) \le \mathrm{trace}(A^\top A)$ (see Lemma 2.5 in Section 2.2), thus we always have $\|A\|_2 \le \|A\|_{\rm F}$. A better bound than (5) is
$$\|AB\|_{\rm F} \le \|A\|_2\|B\|_{\rm F}, \;\; \forall A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p},$$
which can be proved as follows: let $b_i, i = 1, \dots, p$ be the columns of B, and we have
$$\|AB\|_{\rm F}^2 = \sum_{i=1}^{p} \|Ab_i\|_2^2 \le \sum_{i=1}^{p} \|A\|_2^2\|b_i\|_2^2 = \|A\|_2^2 \sum_{i=1}^{p} \|b_i\|_2^2 = \|A\|_2^2\|B\|_{\rm F}^2.$$

1.4 Dual norms

Dual norms are also very frequently used.

Definition 2.4 (Dual norms)**.** *Let* $\|\cdot\|$ *be any norm on* $\mathbb{R}^n$*. Its dual norm is defined as*

$$\|x\|_* = \max_{\|y\|\leq 1} x^\top y. \tag{6}$$

The dual norm is indeed a norm. Positivity and homogeneity are straightforward, and the triangle inequality follows from the fact

$$\|x+z\|_* = \max_{\|y\|\leq 1} (x+z)^\top y \leq \max_{\|y\|\leq 1} x^\top y + \max_{\|y\|\leq 1} z^\top y = \|x\|_* + \|z\|_*.$$

The dual norm (6) can be considered as the operator norm (4) of $x^\top$ that is a matrix of dimension $1 \times n$, with $\|\cdot\|$ on $\mathbb{R}^n$ and the absolute value $|\cdot|$ on $\mathbb{R}$. Some common examples are

- The dual of the Euclidean norm is the Euclidean norm:

  $$\max_{\|y\|_2\leq 1} x^\top y = \|x\|_2,$$

  since $x^\top y \leq \|x\|_2\|y\|_2$ by the Cauchy-Schwarz inequality (Lemma 2.1).

- The dual of the l_1 norm is the l_∞ norm:

  $$\max_{\|y\|_1\leq 1} x^\top y = \max\{|x_1|, \ldots, |x_n|\} = \|x\|_\infty.$$

- The dual of the l_∞ norm is the l_1 norm:

  $$\max_{\|y\|_\infty\leq 1} x^\top y = |x_1| + \cdots + |x_n| = \|x\|_1.$$

Note that these three dual norms are consistent with the l_p operator norms of $x^\top \in \mathbb{R}^{1\times n}$, discussed before Lemma 2.3. More generally, the dual of the l_p norm (where $p \geq 1$) is the the l_q norm, with q satisfying $1/p + 1/q = 1$. By construction, we have the following inequality, which can be considered as a generalized Cauchy-Schwartz inequality.

Lemma 2.4. *Consider any norm* $\|\cdot\|$ *and its associated dual norm* $\|\cdot\|_*$ *on* $\mathbb{R}^n$*. We have*

$$|y^\top x| \leq \|y\|\|x\|_*, \quad \forall\, x, y \in \mathbb{R}^n. \tag{7}$$

This result is directly from the definition (6). For example, we have $|y^\top x| \leq \|y\|_1\|x\|_\infty$. In (7), the norm $\|\cdot\|$ and its dual norm $\|\cdot\|_*$ may not be associated with an inner product (see Theorem 2.1).

We conclude this section by discussing the dual norm of the l_2 or spectral norm on $\mathbb{R}^{m\times n}$

$$\|X\|_{2*} = \max_{\|Y\|_2\leq 1} \operatorname{trace}(Y^\top X),$$

which can be shown to be the sum of the singular values

$$\|X\|_{2*} = \sigma_1(X) + \ldots + \sigma_r(X),$$

where $r = \operatorname{rank}(X)$. This is also known as the *nuclear norm*. It can be verified that $\|X\|_{\mathrm{F}} = (\sigma_1^2(X) + \ldots + \sigma_r^2(X))^{1/2}$, thus we have $\|X\|_2 \leq \|X\|_{\mathrm{F}} \leq \|X\|_{2*}$.

Remark 2.1 (Intuition behind dual norms)**.** *We can consider the dual norm $\|x\|_* = \max_{\|y\|\leq 1} x^\top y$ as the operator norm of $x^\top$ which is a matrix of dimension $1 \times n$. Then, the dual norm $\|x\|_*$ can be viewed as the maximal stretch of the linear map $x^\top$ when it is applied to any vector in $\mathbb{R}^n$ (again we need to define a norm in $\mathbb{R}^n$ first).*

- *When we consider 2 norm, we have $\|A\|_2 = \sigma_{\max}(A)$ (the maximal singular value). For any vector $x \in \mathbb{R}^n$, the maximal singular value is its 2 norm $\|x\|_2 = \sqrt{x_1^2 + \ldots + x_n^2}$.*
- *When we consider 1 norm, $\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^m |A_{ij}|$, which is the maximum column sum. The maximum column sum of $x^\top$ is just its inf norm. So $\|x\|_{1,*} = \|x\|_\infty$.*
- *When we consider inf norm, $\|A\|_\infty = \max_{i=1,\ldots,m} \sum_{j=1}^n |A_{ij}|$, which is the maximum row sum. The maximal row sum of $x^\top$ is just its 1 norm. So $\|x\|_{\infty,*} = \|x\|_1$.*

Another way to see dual norms is that it gives a measure of the size of the linear function $f(x) = x^\top y$ for any $y \in \mathbb{R}^n$. In particular, how big is the number $f(x) = x^\top y$ relative to the size (norm) of y? We can see it is

$$\frac{x^\top y}{\|y\|}.$$

Now, we want to see the maximum size of this, which is

$$\max_{\|y\|\neq 0} \frac{x^\top y}{\|y\|},$$

This is the same as the definition of dual norms in (6) [2]. □

2 Positive semidefinite matrices

One central object in semidefinite optimization is the set of positive semidefinite matrices. In this section, we collect some basic facts and properties.

2.1 Basic facts

We let $\mathbb{S}^n$ denote the set of real symmetric $n \times n$ matrices, and I_n denote the identity matrix of dimension $n \times n$. A fundamental property of any real symmetric matrix $A \in \mathbb{S}^n$ is that it has all real eigenvalues and a set of real eigenvectors $v_1, \ldots, v_n$ forming an orthonormal basis of $\mathbb{R}^n$. This is known as the *spectral decomposition theorem*, one of the most important properties about real symmetric matrices.

Theorem 2.2 (Spectral decomposition theorem)**.** *Let $A \in \mathbb{S}^n$. Then, we can write A as*

$$A = \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}^\top, \tag{8}$$

where $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ are the eigenvalues of A, and $v_1, \ldots, v_n$ are the corresponding eigenvectors that form an orthonormal basis of $\mathbb{R}^n$.

Another convenient way to look at (8) is $A = \sum_{i=1}^n \lambda_i v_i v_i^\top$ which is in the form of rank-one decomposition. Let $V \in \mathbb{R}^{n\times n}$ be an orthogonal matrix with v_i's as its columns, and D be the diagonal matrix with λ_i's on the diagonal. We can also read (8) as $A = VDV^\top$ with $V^{-1} = V^\top$ (or $VV^\top = V^\top V = I_n$).

[2] https://math.stackexchange.com/questions/903484/dual-norm-intuition

Another fundamental result in linear algebra states that if a set of symmetric matrices $A_1, \ldots, A_m$ commute with each other (i.e., $A_iA_j = A_jA_i, \forall i, j = 1, \ldots, m$), then they have a common set of eigenvectors, which implies that they can be "simultaneously diagonalized",

$$A_i = V\Lambda_i V^\top, i = 1, \ldots, m,$$

where V is a common orthogonal matrix and Λ_i is a diagonal matrix with the eigenvalues of A_i being the diagonal entries.

Let $\mathbb{S}^n_+$ (resp. $\mathbb{S}^n_{++}$) denote the set of positive semidefinite (resp. definite) matrices, i.e., the set of real symmetric matrices with nonnegative (resp. strictly positive) eigenvalues. Throughout this course, we shall use the following notations

$$A \succeq 0 \quad \Leftrightarrow \quad A \in \mathbb{S}^n_+ \quad \Leftrightarrow \quad A \text{ is positive semidefinite},$$

and

$$A \succ 0 \quad \Leftrightarrow \quad A \in \mathbb{S}^n_{++} \quad \Leftrightarrow \quad A \text{ is positive definite}.$$

We have the following characterizations for $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$.

Theorem 2.3 (Positive semidefinite matrices)**.** *Let $A \in \mathbb{S}^n$. The following statements are equivalent.*

1. *$A \in \mathbb{S}^n_+$.*
2. *Its spectral decomposition has the form $A = \sum_{i=1}^n \lambda_i v_i v_i^\top$ with all $\lambda_i \geq 0$.*
3. *$x^\top Ax \geq 0, \forall x \in \mathbb{R}^n$.*
4. *There exists a lower triangular matrix $L \in \mathbb{R}^{n\times n}$ such that $A = LL^\top$ (Cholesky factorization).*
5. *All principle minors of A are nonnegative, i.e., $\det(A[S,S]) \geq 0$ for any nonempty $S \subset \{1, 2, \ldots, n\}$ where $A[S,S]$ is the submatrix of A consisting of the rows and columns indexed by S (Sylvester criterion).*

Proof. The equivalence between (1) and (2) is by definition. The direction (1) $\Rightarrow$ (3) is observed from the fact

$$x^\top Ax = x^\top \left(\sum_{i=1}^n \lambda_i v_i v_i^\top\right) x = \sum_{i=1}^n \lambda_i (v_i^\top x)^2 \geq 0, \ \forall x \in \mathbb{R}^n.$$

The converse holds by observing that $v_i^\top A v_i = \lambda_i \|v_i\|^2 \geq 0$ implies $\lambda_i \geq 0$ for all i. The equivalence (1) $\Leftrightarrow$ (5) can be found in any standard textbook on linear algebra.

We discuss the equivalence (1) $\Leftrightarrow$ (4). The Cholesky factorization of a positive semidefinite matrix $A = LL^\top$ requires a lower triangular matrix L. If one only seeks for a matrix $Q \in \mathbb{R}^{n\times n}$ such that $A = QQ^\top$, then we can choose the ith column of Q as $\sqrt{\lambda_i} v_i$ from the spectral decomposition of A in (8) where $\lambda_i \geq 0$ when $A \succeq 0$. Let q_i be the i-th column of $Q^\top$. Applying the Gram-Schmidt orthogonalization process, we can find another orthonormal basis $u_1, \ldots, u_n$ and a lower triangular matrix $L \in \mathbb{R}^{n\times n}$ such that $q_i = \sum_{j=1}^i L_{ij} u_j$, i.e., $Q^\top = UL^\top$ where $U = [u_1, \ldots, u_n]$. Then, we have

$$A = QQ^\top = LU^\top UL^\top = LL^\top,$$

which proves (1) $\Rightarrow$ (4). The converse holds true since $x^\top Ax = x^\top LL^\top x = (L^\top x)^\top (L^\top x) \geq 0, \forall x \in \mathbb{R}^n$. ☐

Note that in the discussion for (1) $\Leftrightarrow$ (4), we can further make $Q = Q^\top$ by choosing

$$Q = V \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}) V^\top,$$

which leads to $A = Q^2$ since $V^\top V = VV^\top = I_n$. In this case, we denote the *symmetric square root* of A as $A^{1/2} = Q \succeq 0$. Upon letting $Q^\top = [q_1, \ldots, q_n]$ that satisfies $A = QQ^\top$, we have $A_{ij} = q_i^\top q_j$ for all $i, j = 1, \ldots, n$. These vectors $q_1, \ldots, q_n$ are also called a Gram representation of A.

The interior of $\mathbb{S}^n_+$ are the set of positive definite matrices.

Theorem 2.4 (Positive definite matrices)**.** *Let $A \in \mathbb{S}^n$. The following statements are equivalent.*

1. *$A \in \mathbb{S}^n_{++}$.*
2. *Its spectral decomposition has the form $A = \sum_{i=1}^n \lambda_i v_i v_i^\mathsf{T}$ with all $\lambda_i > 0$.*
3. *$x^\mathsf{T} A x > 0, \forall x \in \mathbb{R}^n \backslash \{0\}$.*
4. *There exists a lower triangular matrix $L \in \mathbb{R}^{n \times n}$ with $L_{ii} > 0, i = 1, \ldots, n$ such that $A = LL^\mathsf{T}$ (Cholesky factorization).*
5. *All leading principle minors of A are strictly positive, i.e., $\det(A[S,S]) > 0$ for $S = \{1, \ldots, k\}, k = 1, \ldots, n$ (Sylvester criterion).*

The Sylvester criterion in Theorems 2.3 and 2.4 indicates that $A \in \mathbb{S}^n_+$ and $A \in \mathbb{S}^n_{++}$ are defined by a set of polynomial inequalities on its element A_{ij}. For example, consider

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

We have $A \succ 0 \Leftrightarrow a > 0, ac - b^2 > 0$, and $A \succeq 0 \Leftrightarrow a \geq 0, c \geq 0, ac - b^2 \geq 0$. Consider

$$A = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.$$

We have $A \succ 0 \Leftrightarrow a > 0, ad - b^2 > 0, \det(A) > 0$, and

$$A \succeq 0 \Leftrightarrow \begin{cases} a \geq 0, d \geq 0, f \geq 0, \\ ad - b^2 \geq 0, af - c^2 \geq 0, df - e^2 \geq 0, \\ \det(A) \geq 0. \end{cases}$$

2.2 Basic operations

The trace of a square matrix is a linear mapping, i.e.,

$$\text{trace}(\alpha A + \beta B) = \alpha \text{trace}(A) + \beta \text{trace}(B), \forall \alpha, \beta \in \mathbb{R}, A, B \in \mathbb{R}^{n \times n}.$$

The trace also satisfies the following properties $\text{trace}(A) = \text{trace}(A^\mathsf{T}), \forall A \in \mathbb{R}^{n \times n}$ and

$$\text{trace}(AB) = \text{trace}(BA), \quad \forall A \in \mathbb{R}^{n \times m}, B \in \mathbb{R}^{m \times n}. \tag{9}$$

The last property (9) indicates that $\text{trace}(uu^\mathsf{T}) = u^\mathsf{T} u = \|u\|^2, \forall u \in \mathbb{R}^n$. Together with the spectral decomposition theorem (Theorem 2.2), it is easy to see that the trace of $A \in \mathbb{S}^n$ is equal to the sum of its eigenvalues.

Lemma 2.5. *Let $A \in \mathbb{S}^n$ with eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. We have trace$(A) = \lambda_1 + \ldots, \lambda_n$.*

Proof. By the spectral decomposition theorem $A = \sum_{i=1}^n \lambda_i v_i v_i^\mathsf{T}$, we have

$$\text{trace}(A) = \sum_{i=1}^n \lambda_i \text{trace}(v_i v_i^\mathsf{T}) = \sum_{i=1}^n \lambda_i \|v_i\|^2 = \sum_{i=1}^n \lambda_i,$$

which completes the proof. □

We note that the trace property in Lemma 2.5 indeed holds for any square matrices $\mathbb{R}^{n\times n}$ (not necessarily symmetric)[3]. Further, the determinant of a square matrix is the product of its eigenvalues.

Lemma 2.6. *Let $A \in \mathbb{R}^{n\times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. we have*

$$\mathit{trace}(A) = \sum_{i=1}^{n} \lambda_i, \quad \det(A) = \prod_{i=1}^{n} \lambda_i.$$

For symmetric matrices $A, B \in \mathbb{S}^n$, their inner product is given by

$$\langle A, B\rangle = \mathrm{trace}(A^\mathsf{T} B) = \mathrm{trace}(AB) = \sum_{i,j=1}^{n} A_{ij}B_{ij}.$$

We further review two useful facts about positive semidfinite matrices.

Lemma 2.7. *Let $P \in \mathbb{R}^{n\times n}$ be an invertible matrix. Then,*

$$A \in \mathbb{S}^n_+ \;\Leftrightarrow\; PAP^\mathsf{T} \in \mathbb{S}^n_+, \;\text{ and }\; A \in \mathbb{S}^n_{++} \;\Leftrightarrow\; PAP^\mathsf{T} \in \mathbb{S}^n_{++},$$

Proof. If $A \in \mathbb{S}^n_+$, then $y^\mathsf{T} PAP^\mathsf{T} y = x^\mathsf{T} Ax \geq 0, \forall y \in \mathbb{R}^n$, where $x = P^\mathsf{T} y$, which implies $PAP^\mathsf{T} \in \mathbb{S}^n_+$. Conversely, if $PAP^\mathsf{T} \in \mathbb{S}^n_+$, then $x^\mathsf{T} Ax = y^\mathsf{T} PAP^\mathsf{T} y \geq 0, \forall x \in \mathbb{R}^n$, where $y = (P^{-1})^\mathsf{T} x$ since P is invertible. Thus, $A \in \mathbb{S}^n_+$.

The equivalence $A \in \mathbb{S}^n_{++} \;\Leftrightarrow\; PAP^\mathsf{T} \in \mathbb{S}^n_{++}$ can be proved similarly. □

Definition 2.5 (Schur complement)**.** *Consider a symmetric matrix $X \in \mathbb{S}^{m+n}$ partitioned as*

$$X = \begin{bmatrix} A & B \\ B^\mathsf{T} & C \end{bmatrix}, \tag{10}$$

with $A \in \mathbb{S}^n, B \in \mathbb{R}^{n\times m}, C \in \mathbb{S}^m$. If A is non-singular (i.e., $\det(A) \neq 0$), the matrix $C - B^\mathsf{T} A^{-1}B$ is called the Schur complement of A in X.

Lemma 2.8. *Consider a block-partitioned matrix X in (10). Suppose A is non-singular. Then, we have*

1. $X \succeq 0 \;\Leftrightarrow A \succeq 0$ and $C - B^\mathsf{T} A^{-1}B \succeq 0$;
2. $X \succ 0 \;\Leftrightarrow A \succ 0$ and $C - B^\mathsf{T} A^{-1}B \succ 0$.

We end this lecture with a useful property of the kernel of a positive semidefinite matrix. Recall that the *kernel* (or null space) of a matrix $A \in \mathbb{R}^{m\times n}$ is the subspace

$$\ker A = \{x \in \mathbb{R}^n \mid Ax = 0\}.$$

Lemma 2.9. *Let A be a positive semidefinite matrix. Then, we have*

$$Ax = 0 \;\Leftrightarrow\; x^\mathsf{T} Ax = 0.$$

Proof. The direction $\Rightarrow$ is obvious. Consider the converse direction. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A and $v_1, \ldots, v_n \in \mathbb{R}^n$ be the corresponding orthonormal eigenvectors. Then, we can write the vector x in terms of $v_1, \ldots, v_n \in \mathbb{R}^n$ as

$$x = \alpha_1 v_1 + \cdots + \alpha_n v_n = V\mathrm{diag}(\alpha_1, \ldots, \alpha_n),$$

[3]See proofs `https://www.adelaide.edu.au/mathslearning/ua/media/120/evalue-magic-tricks-handout.pdf`.

where $V = [v_1, v_2, \ldots, v_n]$. Now, it is straightforward to verify that

$$0 = x^\top Ax = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)V^\top \underbrace{V\mathrm{diag}(\lambda_1, \ldots, \lambda_n)V^\top}_{A} V\mathrm{diag}(\alpha_1, \ldots, \alpha_n)$$
$$= \sum_{i=1}^{n} \lambda_i \alpha_i^2.$$

Then, we have $\alpha_i = 0$ whenever $\lambda_i > 0$. This indicates that x is a linear combination of the eigenvectors corresponding to $\lambda_i = 0$. Therefore, we have $Ax = 0$. □

A direct application of Lemma 2.9 is the following fact.

Lemma 2.10. *Let* $A = L^\top L$ *where* $L \in \mathbb{R}^{k\times n}$. *Then we have* $\ker A = \ker L$, *and thus*

$$\mathrm{rank}(A) = \mathrm{rank}(L) \le \min\{k, n\}.$$

Notes

For more mathematical background materials, see [1, Appendix A] and [2, Chapter 1].

References

[1] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004.

[2] Monique Laurent and Frank Vallentin. *Semidefinite optimization.* Lecture notes, 2016.