1 Positive semidefinite matrices

1.1 Some basics

Let $\mathbb{M}_n = \mathbb{M}_n(\mathbb{R})$ be the set of $n \times n$ matrices with real entries, and let $\langle \cdot, \cdot \rangle$ denote the standard inner product on $\mathbb{R}^n$. We say that a matrix $A \in \mathbb{M}_n$ is *positive definite* if it symmetric ($A = A^T$), and it holds that

$$\langle x, Ax \rangle > 0 \qquad \forall x \in \mathbb{R}^n. \tag{1.1}$$

The matrix A is said to be *positive semidefinite (PSD)* if we replace '$>$' by '$\geqslant$'. The definition makes it clear that the set of PSD matrices forms a closed, convex cone: If A_1 and A_2 are PSD, then so is $c_1 A_1 + c_2 A_2$ for all $c_1, c_2 \geqslant 0$.

The spectral theorem asserts that every real symmetric matrix A is diagonalizable: We can write $A = UDU^T$ where U is an orthogonal matrix and the diagonal matrix D contains the eigenvalues of A on its diagonal. For a symmetric matrix $A \in \mathbb{M}_n$ with eigenvalues $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$, the variational characterization yields

$$\lambda_k = \min_{\substack{S \subseteq \mathbb{R}^n: \\ \dim(S) = k}} \max \left\{ \langle x, Ax \rangle : x \in S \right\},$$

where the minimum is over all k-dimensional subspaces. Thus the definition (1.1) makes it clear that a symmetric matrix A is positive definite (resp., PSD) if and only if all its eigenvalues are positive (resp., nonnegative).

If we let $u_1, u_2, \ldots, u_n \in \mathbb{R}^n$ denote the rows of U, then each u_i is an eigenvector of A with corresponding eigenvalue λ_i, and

$$A = \lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T + \cdots + \lambda_n u_n u_n^T.$$

Thus the PSD cone is spanned by the collection $\{uu^T : u \in \mathbb{R}^n \setminus \{0\}\}$ of rank-1 PSD matrices. Every PSD matrix A has a unique PSD square root $A^{1/2} = UD^{1/2}U^T = \sum_{i=1}^n \lambda_i^{1/2} u_i u_i^T$.

Remark 1.1 (Hermitian matrices and quantum information). Essentially everything we discuss will apply more generally to matrices with complex entries (which we denote $\mathbb{M}_n(\mathbb{C})$). Recall that a Hermitian matrix A is one that satisfies $A = A^*$ where A^* denotes the conjugate transpose. If $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{C}^n$, then $A \in \mathbb{M}_n(\mathbb{C})$ is said to be positive definite if it is Hermitian and

$$\langle x, Ax \rangle > 0 \qquad \forall x \in \mathbb{C}^n.$$

In quantum information theory, the analog of a probability density function is a density matrix: A PSD matrix $A \in \mathbb{M}_n(\mathbb{C})$ with $\mathrm{Tr}(A) = 1$. A *positive operator-valued measurement (POVM)* is a decomposition of the identity:

$$\rho_1 + \cdots + \rho_m = I,$$

where $\rho_j \in \mathbb{M}_n(\mathbb{C})$ is PSD for each $j \in \{1, 2, \ldots, m\}$. The outcome of the measurement is j with probability $\mathrm{Tr}(\rho_j A)$.

The Loewner order. Let us write $A \succeq 0$ to denote that $A \in \mathbb{M}_n$ is PSD, and $A \succ 0$ to denote that A is positive definite. We write $A \succeq B$ (resp., $A \succ B$) for $A - B \succeq 0$. Note that $A \succeq B \iff \langle x, Ax\rangle \geqslant \langle x, Bx\rangle$ for all $x \in \mathbb{R}^n$.

1.2 Gram matrices

If X is an $m \times n$ real matrix, then $A = X^T X \in \mathbb{M}_n$, and we have $A_{ij} = \langle x_i, x_j\rangle$, where $x_i \in \mathbb{R}^m$ is the ith column of X. One can see that A is PSD:

$$\langle u, Au\rangle = \langle u, X^T X u\rangle = \langle Xu, Xu\rangle \geqslant 0.$$

A is called the *Gram matrix* of the vectors $x_1, x_2, \ldots, x_n \in \mathbb{R}^m$. Moreover, it is not hard to see that every PSD matrix is a Gram matrix since $A = A^{1/2}A^{1/2}$ and $A^{1/2}$ is symmetric.

Suppose that $f_1, \ldots, f_n : \mathbb{R} \to \mathbb{R}$ and $A \in \mathbb{M}_n$ is defined by

$$A_{ij} = \langle f_i, f_j\rangle_{L^2(\mathbb{R})} = \int_{-\infty}^{\infty} f_i(t) f_j(t)\, dt.$$

Then A is a Gram matrix, hence A is PSD. Let's examine some examples. See Remark 1.11 below regarding Hilbert spaces of random variables.

Example 1.2 (Cauchy matrix)**.** Consider numbers $\alpha_1, \ldots, \alpha_n > 0$ and the matrix $A \in \mathbb{M}_n$ defined by $A_{ij} = 1/(\alpha_i + \alpha_j)$. Then A is PSD because

$$\frac{1}{\alpha_i + \alpha_j} = \int_0^{\infty} e^{-t(\alpha_i+\alpha_j)}\, dt = \langle f_i, f_j\rangle_{L^2([0,\infty))},$$

where $f_i(t) := e^{-t\alpha_i}$.

Example 1.3. Consider the $2^n \times 2^n$ matrices $A_{S,T} := c^{|S\triangle T|}$ and $B_{S,T} = |c|^{|S\cup T|}$ for some constant $c \in [-1, 1]$, where $S, T \subseteq \{1, \ldots, n\}$, and $S\triangle T := (S \cup T) \setminus (S \cap T)$ denotes the symmetric difference. Let us see that A and B are PSD.

Let $X_1, \ldots, X_n$ be random variables, and define random variables $Y_S := \prod_{i\in S} X_i$ for $S \subseteq \{1, \ldots, n\}$, and the matrix

$$M_{S,T} := \mathbb{E}[Y_S Y_T].$$

Then M is a Gram matrix, hence PSD.

Take $X_1, \ldots, X_n$ to be i.i.d. with $\mathbb{E}[X_1^2] = 1$ and $\mathbb{E}[X_1] = c$ (this can be done for $|c| \leqslant 1$). Then

$$M_{S,T} = \left(\mathbb{E}[X_1^2]\right)^{|S\cap T|} (\mathbb{E}[X_1])^{|S\triangle T|} = c^{|S\triangle T|},$$

hence A is PSD.

Take $X_1, \ldots, X_n \in \{0, 1\}$ to be i.i.d. with $p = \mathbb{P}[X_1 = 1]$. Then,

$$M_{S,T} = p^{|S\cup T|},$$

hence B is PSD (taking $p = |c|$).

Example 1.4 (Clique counts)**.** Consider an undirected graph $G = (V, E)$ on $n = |V|$ vertices and a number $\ell \in \mathbb{N}$. Define the matrix $J^G \in \mathbb{M}_{2^n}$ so that $J^G_{S,T}$ is the number of ℓ-cliques in the induced graph $G[S \cap T]$. We can see that J^G is PSD by writing:

$$J^G = \sum_U \mathbf{1}_{\{U\subseteq S\}} \mathbf{1}_{\{U\subseteq T\}},$$

where the sum is over all ℓ-cliques U in G.

Example 1.5 (Grigoriev)**.** Let us see a related example where it's not so clear that the matrix is PSD. Define:

$$M_{I,J} := \frac{\binom{m/2}{|I \cup J|}}{\binom{m}{|I \cup J|}}$$

where m is a positive integer and $I, J \subseteq [m]$ range over all subsets with $|I|, |J| \leqslant m/4$. When m is odd, $m/2$ is not an integer, and we use the generalized binomial coefficient

$$\binom{r}{k} := \frac{r \cdot (r-1) \cdots (r-k+1)}{k \cdot (k-1) \cdots 1}.$$

Consider the case when $m = 2k$ and k is an integer. Then for $I, J \subseteq [2k]$, we can interpret $\binom{k}{|I \cup J|}$ as the number of ways to map $I \cup J$ injectively into $[k]$. Hence $M_{I,J}$ is the probability that $I \cup J \subseteq S$, when $S \subseteq [2k]$ with $|S| = k$ is chosen uniformly at random. In other words,

$$M_{I,J} = \frac{1}{\binom{2k}{k}} \sum_{S \subseteq [2k] : |S| = k} \mathbf{1}_{\{I \subseteq S\}} \mathbf{1}_{\{J \subseteq S\}}.$$

Hence for m an even integer, we see that M is PSD. It is a remarkable fact that M remains PSD even when m is an odd integer.

Theorem 1.6 (Grigoriev)**.** *For any integer $m \geqslant 1$, the matrix M is PSD.*

Low-degree sos certificates. Consider m odd and the function $f : \{0,1\}^m \to \mathbb{R}$ given by

$$f(z) = \left(z_1 + \cdots + z_m - \frac{m}{2}\right)^2 - \frac{1}{4}.$$

It holds that $f(z) \geqslant 0$ for $z \in \{0,1\}^m$. It is possible to certify that a function on $\{0,1\}^m$ is nonnegative by writing it as a nonnegative sum of squares:

$$f(z) = \sum_{y \in \{0,1\}^m} \left[\prod_{i=1}^{m} \left(y_i + z_i - 1\right) \right]^2 f(y).$$

But the degree of this "certificate" is very large. Is there a certificate of smaller degree? Nope. In fact, any representation $f(z) = \sum_i p_i(z)^2$ must have $\max\{\deg(p_i)\} \geqslant m/2$. This is the basis for many such lower bounds (integrality gaps for the "SOS hierarchy"). One can use the matrix M to demonstrate this.

For $S \subseteq [m]$, define $z_S := \prod_{i \in S} z_i$. Then every degree-$d$ multilinear polynomial $p : \{0,1\}^m \to \mathbb{R}$ can be written uniquely as $p(z) = \sum_{|S| \leqslant d} p_S z_S$. Define a linear map Φ_M on multilinear polynomials of degree at most m by

$$\Phi_M(z_S) := M_{S,S}, \quad S \subseteq [m],$$

and extending Φ_M linearly. The next claim is a calculation.

Claim 1.7. *If $q(z) = \left(|z| - \frac{m}{2}\right)^2$, then*

$$\Phi_M(q) = 0.$$

This implies that $\Phi_M(f) = -1/4$. On the other hand, for any multilinear polynomial p with $\deg(p) \leqslant m/2$, let us see that $\Phi_M(p^2)$ is nonnegative.

Claim 1.8. *For any multilinear polynomial p with $\deg(p) \leqslant m/4$,*

$$\Phi_M(p^2) \geqslant 0.$$

Proof. Write $p(z) = \sum_{|S| \leqslant m/2} c_S z_S$ so that

$$\Phi_M(p^2) = \sum_{|I|,|J| \leqslant m/4} c_I c_J \Phi_M(z_{I \cup J}) = \sum_{|I|,|J| \leqslant m/4} c_I c_J M_{I,J} = \mathrm{Tr}(CM),$$

where $C_{I,J} = c_I c_J$ defines a rank-1 PSD matrix. By Theorem 1.6, we have $M \succeq 0$ as well. Hence, the next fact yields the claim.

Fact 1.9. *For any matrices $A, B \in \mathbb{M}_d$ with $A, B \succeq 0$, it holds that* $\mathrm{Tr}(AB) \geqslant 0$.

An easy way to see this is to write $A = \sum_i \lambda_i u_i u_i^T$ with $\lambda_i \geqslant 0$ so that

$$\mathrm{Tr}(AB) = \sum_i \lambda_i \mathrm{Tr}(u_i u_i^T B) = \sum_i \lambda_i \langle u_i, B u_i \rangle \geqslant 0. \qquad \square$$

Example 1.10 (Truncated Euclidean distances)**.** Here is a simple application that arises in metric embedding theory. Consider vectors $x_1, x_2, \ldots, x_n \in \mathbb{R}^n$. Then there are vectors $y_1, y_2, \ldots, y_n \in \mathbb{R}^n$ with $\|y_1\|_2 = \cdots = \|y_n\|_2 = 1$,

$$C^{-1} \min(\|x_i - x_j\|_2, 1) \leqslant \|y_i - y_j\|_2 \leqslant C \min(\|x_i - x_j\|_2, 1), \tag{1.2}$$

where $C \geqslant 1$ is some universal constant.

Let $g = (g_1, \ldots, g_n)$ be an n-dimensional standard Gaussian. Define the random complex numbers

$$Y_j := \exp\left(i\langle x_j, g\rangle\right), j = 1, \ldots, n.$$

Then $\left(\mathbb{E}\,|Y_j|^2\right)^{1/2} = 1$. Indeed, each Y_j is uniformly distributed on the unit circle in the complex plane.

Moreover, we have

$$\mathbb{E}\,|Y_i - Y_j|^2 = \mathbb{E}\left|\exp\left(i\langle x_i - x_j, g\rangle\right) - 1\right|^2 = 2\left(1 - \cos(\langle x_i - x_j, g\rangle)\right),$$

where the first equality uses $|e^{i\langle x_j, g\rangle}| = 1$ and the second uses

$$|e^{i\theta} - 1|^2 = |\cos(\theta) - 1 + i\sin(\theta)|^2 = (\cos(\theta) - 1)^2 + \sin(\theta)^2 = 2(1 - \cos(\theta)).$$

Now note that by rotational invariance, $\langle x_i - x_j, g\rangle$ has the same law as $g_1\|x_i - x_j\|_2$, hence

$$\mathbb{E}\,|Y_i - Y_j|^2 = 2\,\mathbb{E}[1 - \cos(g_1\|x_i - x_j\|^2)]$$

One can calculate this integral explicitly, but to achieve (1.2) simply with some constant, one only needs to note that $\cos(\varepsilon) = 1 - \varepsilon + O(\varepsilon^2)$ as $\varepsilon \to 0$. Finally, note that since all n-dimensional Hilbert spaces are isomorphic, there are $y_1, \ldots, y_n \in \mathbb{R}^n$ with $\|y_i - y_j\|_2^2 = \mathbb{E}\,|Y_i - Y_j|^2$. (See Remark 1.11.)

Remark 1.11 (Hilbert spaces of random variables). Consider a probability space (Ω, μ) and a Euclidean space $\mathbb{R}^m$. An $\mathbb{R}^m$-valued random variable X on (Ω, μ) is a measurable function $X : \Omega \to \mathbb{R}^m$, with $\mathbb{E}[X] := \int_\Omega X(\omega)\, d\mu(\omega)$. Define

$$L^2(\Omega, \mu; \mathbb{R}^m) = \left\{X : \Omega \to \mathbb{R}^m : \mathbb{E}\left[\|X\|_2^2\right] < \infty\right\}.$$

Then $\mathcal{H} := L^2(\Omega, \mu; \mathbb{R}^m)$ is a Hilbert space when equipped with the inner product

$$\langle X, Y \rangle_{\mathcal{H}} := \mathbb{E}\left[\langle X, Y \rangle\right],$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product on $\mathbb{R}^m$. Thus we can think of random variables $X \in L^2(\Omega, \mu; \mathbb{R}^m)$ as vectors. If $X_1, \ldots, X_n \in L^2(\Omega, \mu; \mathbb{R}^m)$, then the matrix $A \in \mathbb{M}_n$ given by $A_{ij} := \langle X_i, X_j \rangle_{\mathcal{H}} = \mathbb{E}[X_i X_j]$ is a Gram matrix. This construction occurs in Example 1.3 and Example 1.10.

Indeed, let us find vectors $x_1, \ldots, x_n \in \mathbb{R}^n$ with $\langle x_i, x_j \rangle = \mathbb{E}[X_i X_j]$ for all i, j. Take $u \in \mathbb{R}^n$ and write

$$\langle u, Au \rangle = \sum_{i,j=1}^n u_i u_j \, \mathbb{E}[X_i X_j] = \mathbb{E}\left[\left(\sum_{i=1}^n u_i X_i\right)^2\right] \geqslant 0.$$

This shows that A is PSD, hence once take $x_1, \ldots, x_n$ to be the rows of $\sqrt{A}$.

1.3 A first look at subtleties

Note that if $A, B \succeq 0$, then AB is usually not PSD beacuse AB is usually not symmetric. Indeed, if A, B are symmetric, then AB is symmetric if and only if A and B commute.

The Hadamard product. For $A, B \in \mathbb{M}_n$, define the Hadamard (aka Schur) product $A \circ B \in \mathbb{M}_n$ by $(A \circ B)_{ij} := A_{ij} B_{ij}$. It is a basic fact that if A, B are PSD, then so is $A \circ B$. Using the representation of A, B as conic combinations of rank-1 matrices, it suffices to prove that $aa^T \circ bb^T$ is PSD for all $a, b \in \mathbb{R}^n$, and this follows from the calculation

$$\left\langle u, \left(aa^T \circ bb^T\right) u \right\rangle = \sum_i u_i \sum_j a_i a_j b_i b_j u_j = \left(\sum_i a_i b_i u_i\right)^2 \geqslant 0.$$

Remark 1.12. This is certainly the "wrong" way to prove that $A \circ B$ is PSD. A more principled way is based on the three facts:

1. If $A \in \mathbb{M}_n$ and $B \in \mathbb{M}_m$ are PSD, then so is $A \otimes B \in \mathbb{M}_{mn}$. One can see this by writing $A = \sum_i a_i u_i u_i^T$ and $B = \sum_j b_j v_j v_j^T$ with $\{a_i, b_j \geqslant 0\}$ so that

$$A \otimes B = \sum_{i,j} a_i b_j (u_i \otimes v_j)(u_i \otimes v_j)^T.$$

2. If $A \in \mathbb{M}_n$ is PSD, then so is every principal submatrix of A. (Principal submatrices are those matrices resulting from removing the ith row and column from A for $i \in S \subseteq \{1, 2, \ldots, n\}$.)

3. For $A, B \in \mathbb{M}_n$, it holds that $A \circ B$ is a principal submatrix of $A \otimes B$, specifically,

$$(A \circ B)_{ij} = (A \otimes B)_{ii,jj}.$$

In other words, the Hadamard is just one (basis-dependent) PSD slice of the larger PSD matrix $A \otimes B$.

The symmetrized product $\frac{1}{2}(AB + BA)$ is always symmetric, but still not necessarily PSD when $A, B \succeq 0$. On the other hand, it holds that if A is positive definite, then the symmetrized product with B can only be PSD if B is PSD.

Lemma 1.13. *If $A > 0$ and $\frac{1}{2}(AB + BA) \geq 0$, then $B \geq 0$.*

[Scalar analog: If a, b are real numbers with $a > 0$ and $ab \geqslant 0$, then $b \geqslant 0$.]

Proof. Since the statement of the lemma is not about matrices, but about linear operators, we can assume that A is diagonal (by writing our matrices in the correct choice of basis).

If $A = \mathrm{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n)$, then $\frac{1}{2}(AB + BA) = B \circ [\![(\alpha_i + \alpha_j)/2]\!]$, where we use the notation $M = [\![m_{ij}]\!]$ to denote that $M_{ij} = m_{ij}$. Since $A > 0$, we have $\alpha_i + \alpha_j > 0$, hence we can invert the Hadamard product and obtain

$$B = \tfrac{1}{2}(AB + BA) \circ [\![2/(\alpha_i + \alpha_j)]\!].$$

Since the Hadamard product of PSD matrices is PSD, we are done as long as the latter matrix is PSD. But this is precisely the content of Example 1.2. □

We can use this fact to see that the matrix square root is *operator monotone:* If $A, B > 0$, then

$$A \geq B \implies A^{1/2} \geq B^{1/2}.$$

For numbers, we have the elementary identity $a^2 - b^2 = (a - b)(a + b)$. In the noncommutative setting, things are more delicate: For $X, Y \in \mathbb{M}_n$,

$$X^2 - Y^2 = \tfrac{1}{2}\left[(X - Y)(X + Y) + (X + Y)(X - Y)\right].$$

Now if $X^2 \geq Y^2$, then the LHS is PSD, hence if $X + Y > 0$ as well, then Lemma 1.13 asserts that $X - Y \geq 0$, completing the proof.

But things are subtle: The *matrix square* is *not* operator monotone. In other words, $A \geq B$ does *not* imply $A^2 \geq B^2$. For example:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

We will return to the study of operator monotonicity in Lecture 5.