Positive Semi-Definite Matrices

Danica Sutherland

January 2024

The following concept about matrices is really important to some areas of machine learning, but is only sometimes covered in intro linear algebra courses and CPSC 340. It's a matrix generalization of the concept of a nonnegative number.

1 Positive semi-definite matrices

An $n \times n$ symmetric matrix A is positive semi-definite ("psd") if any of the following equivalent conditions hold:

1. All of the eigenvalues λ_i of A are nonnegative, $\lambda_i \geq 0$.
2. For all vectors $x \in \mathbb{R}^n$, it holds that $x^\mathsf{T} A x \geq 0$.
3. There exists some matrix $B \in R^{m \times n}$ such that $A = B^\mathsf{T} B$ (where m can be any number you like).

Proof of equivalence. If 1 holds, then we know by the definition of the eigendecomposition that $A = \sum_{i=1}^n \lambda_i v_i v_i^\mathsf{T}$, where v_i are the orthonormal eigenvectors of A. Orthonormality means that $\|v_i\|^2 = v_i^\mathsf{T} v_i = 1$, and for all $i \neq j$, $v_i^\mathsf{T} v_j = 0$. Then we have that

$$x^\mathsf{T} A x = x^\mathsf{T} \left(\sum_i \lambda_i v_i v_i^\mathsf{T} \right) x = \sum_i \lambda_i x^\mathsf{T} v_i v_i^\mathsf{T} x = \sum_i \underbrace{\lambda_i}_{\geq 0} \underbrace{\left(v_i^\mathsf{T} x \right)^2}_{\geq 0} \geq 0,$$

showing 2.

If 2 holds, we can similarly show 1: by orthonormality,

$$v_i^\mathsf{T} A v_i = \sum_j \lambda_j \underbrace{v_i^\mathsf{T} v_j}_{\mathbb{1}(i=j)} \underbrace{v_j^\mathsf{T} v_i}_{\mathbb{1}(i=j)} = \lambda_i,$$

implying by 2 that $\lambda_i \geq 0$ for all i. Thus 1 and 2 are equivalent.

We also have that 1 implies 3: use the matrix $B = \sum_i \sqrt{\lambda_i} v_i v_i^\mathsf{T}$ (the "matrix square root"), which exists since each $\lambda_i \geq 0$. Then

$$B^\mathsf{T} B = \left(\sum_i \sqrt{\lambda_i} v_i v_i^\mathsf{T} \right)^\mathsf{T} \left(\sum_i \sqrt{\lambda_i} v_i v_i^\mathsf{T} \right) = \left(\sum_i \sqrt{\lambda_i} v_i v_i^\mathsf{T} \right) \left(\sum_j \sqrt{\lambda_j} v_j v_j^\mathsf{T} \right)$$
$$= \sum_i \sum_j \sqrt{\lambda_i \lambda_j} v_i \underbrace{v_i^\mathsf{T} v_j}_{\mathbb{1}(i=j)} v_j^\mathsf{T} = \sum_i \lambda_i v_i v_i^\mathsf{T} = A.$$

There are also other ways to construct a B satisfying this condition; the Cholesky decomposition is frequently useful in practice for full-rank A. (For A of rank $r < n$, doing incomplete Cholesky for r steps and then truncating gives an $r \times n$ matrix B.)

Finally, if (3) holds, then we can directly show (2):

$$x^\mathsf{T} A x = x^\mathsf{T} (B^\mathsf{T} B) x = \|Bx\|^2 \geq 0. \qquad \square$$

2 Strictly positive definite matrices

An $n \times n$ symmetric matrix is called *strictly positive definite* ("strictly pd") if any of the following equivalent conditions hold:

1. All of the eigenvalues λ_i of A are positive, $\lambda_i > 0$.
2. For all nonzero vectors $x \in \mathbb{R}^n$, it holds that $x^\mathsf{T} A x > 0$.
3. There exists some full-rank matrix $B \in \mathbb{R}^{n \times n}$ such that $A = B^\mathsf{T} B$.

The proof that these conditions are equivalent is essentially the same as for positive semi-definiteness.

A warning about terminology Some authors use "positive semi-definite" to mean nonnegative eigenvalues (as here) and "positive definite" to mean positive eigenvalues. Other authors use "positive definite" to mean nonnegative eigenvalues and "strictly positive definite" to mean positive eigenvalues.

So, if you just hear someone say "positive definite," it's not clear which they mean! I try to only use the unambiguous terms. (This is kind of similar to how some authors use $\subseteq$ and $\subset$, while others use $\subset$ and $\subsetneq$.)

Abbreviations also vary: I like "psd," but don't use "spd" because sometimes people use that for "symmetric positive definite" (which could mean either of these two properties).

3 Applications

Positive semi-definite matrices are kind of the matrix analogue to nonnegative numbers, while strictly positive definite matrices are kind of the matrix analogue to positive numbers. In particular, notice that a 1×1 matrix $\begin{bmatrix} a \end{bmatrix}$ is positive semi-definite iff $a \geq 0$, and strictly positive definite iff $a > 0$.

For example, a continuously twice-differentiable $f : \mathbb{R}^n \to \mathbb{R}$ is convex iff its Hessian is positive semi-definite everywhere (similarly to $f''(x) \geq 0$), and is strictly convex if (but not only if!) its Hessian is strictly positive definite everywhere (similarly to $f''(x) > 0$).

Similarly, a symmetric matrix $\Sigma \in \mathbb{R}^{n \times n}$ is a valid covariance matrix for the variables $X_1, \ldots, X_n$ iff it's positive semi-definite: this is because

$$\operatorname{Var}\left(\sum_i w_i X_i \right) = \sum_{ij} w_i \operatorname{Cov}(X_i, X_j) w_j = w^\mathsf{T} \Sigma w$$

must be nonnegative for all vectors w.

This analogy justifies the common notation $A \succeq 0$ for A being psd, and $A \succ 0$ for A being strictly pd. This can be extended to the "Loewner ordering," where we say $A \succeq B$ iff $A - B \succeq 0$, similar to how $a \geq b$ iff $a - b \geq 0$; we define $A \succ B$ similarly, and $B \preceq A$ iff $A \succeq B$, $B \prec A$ iff $A \succ B$. In particular, the statement $cI \preceq A \preceq dI$ means that all the eigenvalues of A are in $[c, d]$.