I. Linear Algebra

I.7. Positive Semi-Definite Matrices

Lecture based on

https://github.com/gwthomas/math4ml (Garrett Thomas, 2018)

Prof. Dr. Christoph Lippert

Digital Health & Machine Learning

A symmetric matrix $\mathbf{A}$ is **positive semi-definite** if for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$.

Sometimes people write $\mathbf{A} \succeq 0$ to indicate that $\mathbf{A}$ is positive semi-definite.

A symmetric matrix $\mathbf{A}$ is **positive definite** if for all nonzero $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^\top \mathbf{A}\mathbf{x} > 0$. Sometimes people write $\mathbf{A} \succ 0$ to indicate that $\mathbf{A}$ is positive definite.

Note that positive definiteness is a strictly stronger property than positive semi-definiteness, in the sense that every positive definite matrix is positive semi-definite but not vice-versa.

These properties are related to eigenvalues in the following way.

Proposition

A symmetric matrix is positive semi-definite if and only if all of its eigenvalues are nonnegative, and positive definite if and only if all of its eigenvalues are positive.

Proof.

Suppose A is positive semi-definite, and let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ with eigenvalue λ.
Then

$$0 \leq \mathbf{x}^\top \mathbf{A} \mathbf{x} = \mathbf{x}^\top (\lambda \mathbf{x}) = \lambda \mathbf{x}^\top \mathbf{x} = \lambda \|\mathbf{x}\|_2^2$$

Since $\mathbf{x} \neq \mathbf{0}$ (by the assumption that it is an eigenvector), we have $\|\mathbf{x}\|_2^2 > 0$.
Therefore, we can divide both sides by $\|\mathbf{x}\|_2^2$ to arrive at $\lambda \geq 0$.

If $\mathbf{A}$ is positive definite, the inequality above holds strictly, so $\lambda > 0$.

This proves one direction.

$\square$

These properties are related to eigenvalues in the following way.

Proposition

A symmetric matrix is positive semi-definite if and only if all of its eigenvalues are nonnegative, and positive definite if and only if all of its eigenvalues are positive.

Proof.

To simplify the proof of the other direction, we will use the machinery of Rayleigh quotients.

Suppose that $\mathbf{A}$ is symmetric and all its eigenvalues are nonnegative.

Then for all $\mathbf{x} \neq \mathbf{0}$,

$$0 \leq \lambda_{\min}(\mathbf{A}) \leq R_{\mathbf{A}}(\mathbf{x})$$

Since $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ matches $R_{\mathbf{A}}(\mathbf{x})$ in sign, we conclude that $\mathbf{A}$ is positive semi-definite.

If the eigenvalues of $\mathbf{A}$ are all strictly positive, then $0 < \lambda_{\min}(\mathbf{A})$. Hence $\mathbf{A}$ is positive definite. $\square$

As an example of how these matrices arise, consider

Proposition

Suppose $\mathbf{A} \in \mathbb{R}^{m \times n}$.
Then $\mathbf{A}^\top\mathbf{A}$ *is positive semi-definite.*
If $\mathrm{null}(\mathbf{A}) = \{\mathbf{0}\}$, *then* $\mathbf{A}^\top\mathbf{A}$ *is positive definite.*

Proof.

For any $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^\top(\mathbf{A}^\top\mathbf{A})\mathbf{x} = (\mathbf{A}\mathbf{x})^\top(\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|_2^2 \geq 0$$

so $\mathbf{A}^\top\mathbf{A}$ is positive semi-definite.
Note that $\|\mathbf{A}\mathbf{x}\|_2^2 = 0$ implies $\|\mathbf{A}\mathbf{x}\|_2 = 0$, which in turn implies $\mathbf{A}\mathbf{x} = \mathbf{0}$ (recall that this is a property of norms).
If $\mathrm{null}(\mathbf{A}) = \{\mathbf{0}\}$, $\mathbf{A}\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$, so $\mathbf{x}^\top(\mathbf{A}^\top\mathbf{A})\mathbf{x} = 0$ if and only if $\mathbf{x} = \mathbf{0}$, and thus $\mathbf{A}^\top\mathbf{A}$ is positive definite. $\square$

Positive definite matrices are invertible (since their eigenvalues are nonzero), whereas positive semi-definite matrices might not be.
However, if you already have a positive semi-definite matrix, it is possible to perturb its diagonal slightly to produce a positive definite matrix.

Proposition

If $\mathbf{A}$ is positive semi-definite and $\epsilon > 0$, then $\mathbf{A} + \epsilon\mathbf{I}$ is positive definite.

Proof.

Let $\mathbf{A}$ a be positive semi-definite and $\epsilon > 0$.
Then, we have for any $\mathbf{x} \neq \mathbf{0}$ that

$$\begin{aligned} \mathbf{x}^\top(\mathbf{A} + \epsilon\mathbf{I})\mathbf{x} &= \mathbf{x}^\top\mathbf{A}\mathbf{x} + \epsilon\mathbf{x}^\top\mathbf{I}\mathbf{x} \\ &= \underbrace{\mathbf{x}^\top\mathbf{A}\mathbf{x}}_{\geq 0} + \underbrace{\epsilon\|\mathbf{x}\|_2^2}_{>0} \\ &> 0 \quad \text{as claimed.} \end{aligned}$$

□

Corollary

An obvious but frequently useful consequence of the two propositions we have just shown is that $\mathbf{A}^\top \mathbf{A} + \epsilon \mathbf{I}$ *is positive definite (and in particular, invertible) for any (potentially non-square) matrix* $\mathbf{A}$ *and any* $\epsilon > 0$.

Proof.

Let $\epsilon > 0$ and $\mathbf{A}$ be an $M \times N$ dimensional matrix.
For any N-dimensional vector $\mathbf{x} \neq \mathbf{0}$

$$\begin{aligned}
\mathbf{x}^\top \left(\mathbf{A}^\top \mathbf{A} + \epsilon \mathbf{I} \right) \mathbf{x} &= \mathbf{x}^\top \mathbf{A}^\top \mathbf{A} \mathbf{x} + \epsilon \mathbf{x}^\top \mathbf{x} \\
&= (\mathbf{A}\mathbf{x})^\top (\mathbf{A}\mathbf{x}) + \epsilon \mathbf{x}^\top \mathbf{x} \\
&= \underbrace{\|\mathbf{A}\mathbf{x}\|_2^2}_{\geq 0} + \epsilon \underbrace{\|\mathbf{x}\|_2^2}_{> 0} \\
&> 0
\end{aligned}$$

as claimed. $\square$

A useful way to understand quadratic forms is by the geometry of their level sets.

A **level set** or **isocontour** of a function is the set of all inputs such that the function applied to those inputs yields a given output.

Mathematically, the c-isocontour of f is $\{\mathbf{x} \in \operatorname{dom} f : f(\mathbf{x}) = c\}$.

Let us consider the special case $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ where $\mathbf{A}$ is a positive definite matrix.

Since $\mathbf{A}$ is positive definite, it has a unique matrix square root

$$\mathbf{A}^{\frac{1}{2}} = \mathbf{Q}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{Q}^\top,$$

where $\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top$ is the eigendecomposition of $\mathbf{A}$ and $\mathbf{\Lambda}^{\frac{1}{2}} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots \sqrt{\lambda_n})$.

It is easy to see that this matrix $\mathbf{A}^{\frac{1}{2}}$ is positive definite (consider its eigenvalues) and satisfies $\mathbf{A}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}} = \mathbf{A}$.

Fixing a value $c \geq 0$, the c-isocontour of f is the set of $\mathbf{x} \in \mathbb{R}^n$ such that

$$c = \mathbf{x}^\top \mathbf{A}\mathbf{x} = \mathbf{x}^\top \mathbf{A}^{\frac{1}{2}}\mathbf{A}^{\frac{1}{2}}\mathbf{x} = \left(\mathbf{A}^{\frac{1}{2}}\mathbf{x}\right)^\top \left(\mathbf{A}^{\frac{1}{2}}\mathbf{x}\right) = \|\mathbf{A}^{\frac{1}{2}}\mathbf{x}\|_2^2$$

where we have used the symmetry of $\mathbf{A}^{\frac{1}{2}}$.

Making the change of variable $\mathbf{z} = \mathbf{A}^{\frac{1}{2}}\mathbf{x}$, we have the condition $\|\mathbf{z}\|_2 = \sqrt{c}$.

That is, the values $\mathbf{z}$ lie on a sphere of radius $\sqrt{c}$.

These can be parameterized as $\mathbf{z} = \sqrt{c}\hat{\mathbf{z}}$ where $\hat{\mathbf{z}}$ has $\|\hat{\mathbf{z}}\|_2 = 1$.

Then since $\mathbf{A}^{-\frac{1}{2}} = \mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^\top$, we have

$$\mathbf{x} = \mathbf{A}^{-\frac{1}{2}}\mathbf{z} = \mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{Q}^\top\sqrt{c}\hat{\mathbf{z}} = \sqrt{c}\mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}}\tilde{\mathbf{z}}$$

where $\tilde{\mathbf{z}} = \mathbf{Q}^\top\hat{\mathbf{z}}$ also satisfies $\|\tilde{\mathbf{z}}\|_2 = 1$ since $\mathbf{Q}$ is orthogonal.

Using this parameterization, we see that the solution set

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = c\}$$

is the image of the unit sphere

$$\{\tilde{\mathbf{z}} \in \mathbb{R}^n : \|\tilde{\mathbf{z}}\|_2 = 1\}$$

under the invertible linear map $\mathbf{x} = \sqrt{c}\mathbf{Q}\mathbf{\Lambda}^{-\frac{1}{2}}\tilde{\mathbf{z}}$.

What we have gained with all these manipulations is a clear algebraic understanding of the c-isocontour of f in terms of a sequence of linear transformations applied to a well-understood set.

We begin with the unit sphere, then scale every axis i by $\lambda_i^{-\frac{1}{2}}$, resulting in an axis-aligned ellipsoid.

Observe that the axis lengths of the ellipsoid are proportional to the inverse square roots of the eigenvalues of $\mathbf{A}$.

Hence larger eigenvalues correspond to shorter axis lengths, and vice-versa.

Then this axis-aligned ellipsoid undergoes a rigid transformation (i.e. one that preserves length and angles, such as a rotation/reflection) given by $\mathbf{Q}$.

The result of this transformation is that the axes of the ellipse are no longer along the coordinate axes in general, but rather along the directions given by the corresponding eigenvectors.

To see this, consider the unit vector $\mathbf{e}_i \in \mathbb{R}^n$ that has $[\mathbf{e}_i]_j = \delta_{ij}$.

In the pre-transformed space, this vector points along the axis with length proportional to $\lambda_i^{-\frac{1}{2}}$.

But after applying the rigid transformation $\mathbf{Q}$, the resulting vector points in the direction of the corresponding eigenvector $\mathbf{q}_i$, since

$$\mathbf{Q}\mathbf{e}_i = \sum_{j=1}^{n} [\mathbf{e}_i]_j \mathbf{q}_j = \mathbf{q}_i$$

where we have used the matrix-vector product identity from earlier.

In summary:

- The isocontours of $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x}$ are ellipsoids such that the axes point in the directions of the eigenvectors $\mathbf{q}_i$ of $\mathbf{A}$.
- The radii of these axes are proportional to the inverse square roots of the corresponding eigenvalues λ_i.