Positive (Semi)Definiteness

This week we will learn about:

- Positive definite and positive semidefinite matrices,
- Gershgorin discs and diagonal dominance,
- The principal square root of a matrix, and
- The polar decomposition.

Extra reading and watching:

- Section 2.2 in the textbook
- Lecture videos 30, 31, 32, and 33 on YouTube
- Positive-definite matrix at Wikipedia
- Gershgorin circle theorem at Wikipedia
- Square root of a matrix at Wikipedia
- Polar decomposition at Wikipedia

Extra textbook problems:

⋆ 2.2.1, 2.2.2

⋆⋆ 2.2.3, 2.2.5–2.2.10, 2.2.12

⋆⋆⋆ 2.2.11, 2.2.14, 2.2.16, 2.2.19, 2.2.22

☠ 2.2.18, 2.2.27, 2.2.28

Recall that normal matrices play a particularly important role in linear algebra (they can be diagonalized by unitary matrices). There is one particularly important family of normal matrices that we now focus our attention on.

Definition 8.1 — Positive (Semi)Definite Matrices

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A = A^* \in \mathcal{M}_n(\mathbb{F})$. Then A is called

a) positive semidefinite (PSD) if $\mathbf{v}^* A\mathbf{v} \geq 0$ for all $\mathbf{v} \in \mathbb{F}^n$, and

b) positive definite (PD) if $\mathbf{v}^* A\mathbf{v} > 0$ for all $\mathbf{v} \neq \mathbf{0}$.

Positive (semi)definiteness is somewhat difficult to eyeball from the entries of a matrix, and we should emphasize that it does *not* mean that the entries of the matrix need to be positive. For example, if

$$A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix},$$

then...

The definition of positive semidefinite matrices perhaps looks a bit odd at first glance. The next theorem characterizes these matrices in several other equivalent ways, some of which are hopefully a bit more illuminating and easier to work with.

Theorem 8.1 — Characterization of PSD and PD Matrices

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A = A^* \in \mathcal{M}_n(\mathbb{F})$. The following are equivalent:

a) A is positive (semidefinite | definite),

b) All of the eigenvalues of A are (non-negative | strictly positive),

c) There exists a diagonal $D \in \mathcal{M}_n(\mathbb{R})$ with (non-negative | strictly positive) diagonal entries and a unitary matrix $U \in \mathcal{M}_n(\mathbb{F})$ such that $A = UDU^*$, and

d) There exists (a matrix | an invertible matrix) $B \in \mathcal{M}_n(\mathbb{F})$ such that $A = B^*B$.

Proof. We prove the theorem by showing that (a) $\implies$ (b) $\implies$ (c) $\implies$ (d) $\implies$ (a).

■

Example. *Show that* $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ *is PSD, but not PD, in several different ways.*

Example. *Show that* $A = \begin{bmatrix} 2 & -1 & i \\ -1 & 2 & 1 \\ -i & 1 & 2 \end{bmatrix}$ *is positive definite.*

OK, let's look at another way of determining whether or not a matrix is positive definite, which has the advantage of not requiring us to compute eigenvalues.

Theorem 8.2 — Sylvester's Criterion

Let $A = A^* \in \mathcal{M}_n$. Then A is positive definite if and only if the determinant of the top-left $k \times k$ block of A is strictly positive for all $1 \leq k \leq n$.

We won't prove Sylvester's Criterion (a proof is in the textbook if you're curious), but instead let's jump right to an example to illustrate how it works.

Example. *Use Sylvester's criterion to show that* $A = \begin{bmatrix} 2 & -1 & i \\ -1 & 2 & 1 \\ -i & 1 & 2 \end{bmatrix}$ *is positive definite.*

Let's wrap up this section by reminding ourselves of something that we already proved about positive definite matrices a few weeks ago:

Theorem 8.3 — Positive Definite Matrices Make Inner Products

A function $\langle \cdot, \cdot \rangle : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ is an inner product if and only if there exists a positive definite matrix $A \in \mathcal{M}_n(\mathbb{F})$ such that

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^* A \mathbf{w} \quad \text{for all} \quad \mathbf{v}, \mathbf{w} \in \mathbb{F}^n.$$

Diagonal Dominance and Gershgorin Discs

In order to motivate this next section, let's think a bit about what Sylvester's criterion says when the matrix A is 2×2.

Theorem 8.4 — Positive Definiteness for 2×2 Matrices

Let $a, d \in \mathbb{R}$, $b \in \mathbb{C}$, and suppose that $A = \begin{bmatrix} a & b \\ \overline{b} & d \end{bmatrix}$.

a) A is positive semidefinite if and only if $a, d \geq 0$ and $|b|^2 \leq ad$, and

b) A is positive definite if and only if $a > 0$ and $|b|^2 < ad$.

Indeed, case (b) is exactly Sylvester's criterion. For case (a)...

Example. *Show that* $A = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ *is positive semidefinite, but not positive definite.*

The previous theorem basically says that a 2×2 matrix is positive (semi)definite as long as its off-diagonal entries are "small enough" compared to its diagonal entries. This same intuition is well-founded even for larger matrices. However, to clarify exactly what we mean, we first need the following result that helps us bound the eigenvalues of a matrix based on simple information about its entries.

Theorem 8.5 — Gershgorin Disc Theorem

Let $A \in \mathcal{M}_n(\mathbb{C})$ and define the following objects:

- $r_i = \sum_{j \neq i} |a_{i,j}|$ (the sum of the off-diagonal entries of the i-th row of A),
- $D(a_{i,i}, r_i)$ is the closed disc in the complex plane centered at $a_{i,i}$ with radius r_i.

Then every eigenvalue of A is in at least one of the $D(a_{i,i}, r_i)$ (called **Gershgorin discs**).

Example. *Draw the Gershgorin discs for...*

Proof of Theorem 8.5. Let λ be an eigenvalue of A with associated eigenvector $\mathbf{v}$. Then...

■

The Gershgorin disc theorem is an approximation theorem. For diagonal matrices we have $r_i = 0$ for all i, so the Gershgorin discs have radius 0 and thus the eigenvalues are exactly the diagonal entries (which we already knew from the previous course). However, as the off-diagonal entries increase, the radii of the Gershgorin discs increase so the eigenvalues can wiggle around a bit.

In order to connect Gershgorin discs to positive semidefiniteness, we introduce one additional family of matrices:

Definition 8.2 — Diagonally Dominant Matrices

Suppose that $A \in \mathcal{M}_n(\mathbb{C})$. Then A is called

a) **diagonally dominant** if $|a_{i,i}| \geq \sum_{j \neq i} |a_{i,j}|$ for all $1 \leq i \leq n$, and

b) **strictly diagonally dominant** if $|a_{i,i}| > \sum_{j \neq i} |a_{i,j}|$ for all $1 \leq i \leq n$.

Example. *Show that the matrix*

$$A = \begin{bmatrix} 2 & 0 & i \\ 0 & 3 & 1 \\ -i & 1 & 5 \end{bmatrix}$$

is strictly diagonally dominant, and draw its Gershgorin discs.

In particular, since the eigenvalues of the previous matrix were positive, it was necessarily positive definite. This same type of argument works in general, and leads immediately to the following theorem:

Theorem 8.6 — Diagonal Dominance Implies PSD

Suppose that $A = A^* \in \mathcal{M}_n(\mathbb{C})$ has non-negative diagonal entries.

a) If A is diagonally dominant then it is positive semidefinite.

b) If A is strictly diagonally dominant then it is positive definite.

Be careful: this is a one-way theorem! DD implies PSD, but PSD does not imply DD. For example,

Unitary Freedom of PSD Decompositions

We saw earlier that for every positive semidefinite matrix A we can find a matrix B such that $A = B^*B$. However, this matrix B is not unique, since if U is a unitary matrix and we define $C = UB$ then

The following theorem says that we can find *all* decompositions of A using this same procedure:

Theorem 8.7 — Unitary Freedom of PSD Decompositions

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $B, C \in \mathcal{M}_{m,n}(\mathbb{F})$. Then $B^*B = C^*C$ if and only if there exists a unitary matrix $U \in \mathcal{M}_m(\mathbb{F})$ such that $C = UB$.

For the purpose of saving time, we do not show the "only if" direction of the proof here (it is in the textbook, in case you are interested).

The previous theorem raises the question of how simple we can make the matrix B in a positive semidefinite decomposition $A = B^*B$. The following theorem provides one possible answer: we can choose B so that it is also positive semidefinite.

Theorem 8.8 — Principal Square Root of a Matrix

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_n(\mathbb{F})$ is positive semidefinite. Then there exists a unique positive semidefinite matrix $P \in \mathcal{M}_n(\mathbb{F})$, called the **principal square root** of A, such that

$$A = P^2$$

Proof. To see that such a matrix P exists, we use our usual diagonalization arguments.

■

The principal square root P of a matrix A is typically denoted by $P = \sqrt{A}$, and is in analogy with the principal square root of a non-negative real number (indeed, for 1×1 matrices they are the exact same thing).

Example. *Find the principal square root of...*

By combining our previous two theorems, we also recover a new matrix decomposition, which answers the question of how simple we can make a matrix by multiplying it on the left by a unitary matrix—we can always make it positive semidefinite.

Theorem 8.9 — Polar Decomposition

Suppose $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, and $A \in \mathcal{M}_n(\mathbb{F})$. Then there exists a unitary matrix $U \in \mathcal{M}_n(\mathbb{F})$ and a positive semidefinite matrix $P \in \mathcal{M}_n(\mathbb{F})$ such that

$$A = UP.$$

Proof. Since A^*A is positive semidefinite, we know from the previous theorem that

■

The matrix $\sqrt{A^*A}$ in the polar decomposition can be thought of as the "matrix version" of the absolute value of a complex number $|z| = \sqrt{\overline{z}z}$. In fact, this matrix is sometimes even denoted by $|A| = \sqrt{A^*A}$. Similarly, the polar decomposition of a matrix generalizes the polar form of a complex number:

We don't know how to compute the polar decomposition yet (since we skipped a proof earlier this week), but we will learn a method soon.

Over the past couple of weeks, we learned about several new families of matrices. It is worth drawing a diagram illustrating their relationships with each other:

It is also worth noting that many of these families of matrices are analogous to important subsets of the complex plane: