1   Symmetric Matrices

We review some basic results concerning symmetric matrices. All matrices that we discuss are over the real numbers.

Let A be a real, symmetric matrix of size $d \times d$ and let I denote the $d \times d$ identity matrix. Perhaps the most important and useful property of symmetric matrices is that their eigenvalues behave very nicely.

Definition 1 *Let U be a $d \times d$ matrix. The matrix U is called an **orthogonal matrix** if $U^{\mathsf{T}} U = I$. This implies that $U U^{\mathsf{T}} = I$, by uniqueness of inverses.*

Fact 2 *(Spectral Theorem). Let A be any $d \times d$ symmetric matrix. There exists an orthogonal matrix U and a (real) diagonal matrix D such that*

$$A \;=\; U D U^{\mathsf{T}}.$$

*This is called a **spectral decomposition** of A. Let u_i be the ith column of U and let λ_i denote the ith diagonal entry of D. Then $\{u_1, \dots, u_d\}$ is an orthonormal basis consisting of eigenvectors of A, and λ_i is the eigenvalue corresponding to u_i. We can also write*

$$A \;=\; \sum_{i=1}^{d} \lambda_i u_i u_i^{\mathsf{T}}. \tag{1}$$

The eigenvalues λ are uniquely determined by A, up to reordering.

Caution. The product of two symmetric matrices is usually not symmetric.

1.1   Positive semi-definite matrices

Definition 3 *Let A be any $d \times d$ symmetric matrix. The matrix A is called **positive semi-definite** if all of its eigenvalues are non-negative. This is denoted $A \succeq 0$, where here 0 denotes the zero matrix. The matrix A is called **positive definite** if all of its eigenvalues are strictly positive. This is denoted $A \succ 0$.*

There are many equivalent characterizations of positive semi-definiteness. One useful condition is

$$A \succeq 0 \qquad \Longleftrightarrow \qquad x^{\mathsf{T}} A x \geq 0 \;\; \forall x \in \mathbb{R}^d. \tag{2}$$

Similarly,

$$A \succ 0 \qquad \Longleftrightarrow \qquad x^{\mathsf{T}} A x > 0 \;\; \forall x \in \mathbb{R}^d$$

Another condition is: $A \succeq 0$ if and only if there exists a matrix V such that $A = V V^{\mathsf{T}}$.

The positive semi-definite condition can be used to define a partial ordering on all symmetric matrices. This is called the **Löwner ordering** or the **positive semi-definite ordering**. For any two symmetric matrices A and B, we write $A \succeq B$ if $A - B \succeq 0$.

Fact 4 $A \succeq B$ *if and only if* $v^{\mathsf{T}} A v \geq v^{\mathsf{T}} B v$ *for all vectors* v.

PROOF: Note that $v^{\mathsf{T}} A v \geq v^{\mathsf{T}} B v$ holds if and only if $v^{\mathsf{T}} (A - B) v \geq 0$ holds. By (2), this is equivalent to $A - B \succeq 0$ which is the definition of $A \succeq B$. □

The set of positive semi-definite matrices is closed under addition and non-negative scaling. (Such a set is called a **convex cone**.)

Fact 5 *Let A and B be positive semi-definite matrices of size $d \times d$. Let α, β be non-negative scalars. Then $\alpha A + \beta B \succeq 0$.*

PROOF: This follows easily from (2). □

Caution. The Löwner ordering does not have all of the nice properties that the usual ordering of real numbers has. For example, if $A \succeq B \succeq 0$ then it is **not** necessarily true that $A^2 \succeq B^2$.

Fact 6 *Let A and B be symmetric, $d \times d$ matrices and let C be any $d \times d$ matrix. If $A \succeq B$ then $CAC^{\mathsf{T}} \succeq CBC^{\mathsf{T}}$. Moreover, if C is non-singular then the "if" is actually "if and only if".*

PROOF: Since $A \succeq B$, then $A - B \succeq 0$, so there exists a matrix V such that VV^{T}. Then

$$C(A - B)C^{\mathsf{T}} \ = \ CVV^{\mathsf{T}}C^{\mathsf{T}} \ = \ CV(CV)^{\mathsf{T}},$$

which shows that $C(A - B)C^{\mathsf{T}}$ is positive semi-definite.

Suppose that C is non-singular and $CAC^{\mathsf{T}} \succeq CBC^{\mathsf{T}}$. Multiply on the left by C^{-1} and on the right by $(C^{-1})^{\mathsf{T}} = (C^{\mathsf{T}})^{-1}$ to get $A \succeq B$. □

1.2   Spectral Norm

Just as there are many norms of vectors, there are many norms of matrices too. We will only consider **spectral norm** of A. (This is also called the ℓ_2 operator norm and the Schatten ∞-norm.)

Definition 7 *Let A be a symmetric $d \times d$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_d$. The spectral norm of A is denoted $\|A\|$ and is defined by $\|A\| = \max_i |\lambda_i|$.*

Recall that an eigenvalue of A is any real number z such that there exists a vector u for which $Au = zu$. Then we also have $(A + tI)u = (z + t)u$, which implies that $z + t$ is an eigenvalue of $A + tI$. In fact, the entire set of eigenvalues of $A + tI$ is $\{\lambda_1 + t, \ldots, \lambda_d + t\}$.

Fact 8 *Let A be a symmetric matrix. Let $\lambda_{\min}$ and $\lambda_{\max}$ respectively denote the smallest and largest eigenvalues of A. Then $\lambda_{\min} \cdot I \preceq A \preceq \lambda_{\max} \cdot I$.*

PROOF: Let $\lambda_1, \ldots, \lambda_d$ be all the eigenvalues of A. We will show only the second inequality, which is equivalent to $\lambda_{\max} \cdot I - A \succeq 0$. As observed above, the eigenvalues of $\lambda_{\max} \cdot I - A$ are

$$\{\lambda_{\max} - \lambda_1, \ldots, \lambda_{\max} - \lambda_d\}.$$

The minimum of these eigenvalues is

$$\min_j \left(\lambda_{\max} - \lambda_j \right) \;=\; \lambda_{\max} - \max_j \lambda_j \;=\; 0.$$

So $\lambda_{\max} \cdot I - A \succeq 0$. $\square$

In particular, for any symmetric matrix A we have $A \preceq \|A\| \cdot I$.

1.3 Trace

Definition 9 *Let A be an arbitrary $d \times d$ matrix (not necessarily symmetric). The **trace** of A, denoted $\mathrm{tr}(A)$, is the sum of the diagonal entries of A.*

Fact 10 *(Linearity of Trace) Let A and B be arbitrary $d \times d$ matrices and let α, β be scalars. Then $\mathrm{tr}(\alpha A + \beta B) = \alpha \, \mathrm{tr}(A) + \beta \, \mathrm{tr}(B)$.*

Fact 11 *(Cyclic Property of Trace) Let A be an arbitrary $n \times m$ matrix and let B be an arbitrary $m \times n$ matrix. Then $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.*

PROOF: This easily follows from the definitions. The ith diagonal entry of AB is $\sum_{j=1}^m A_{i,j} B_{j,i}$, so

$$\mathrm{tr}(AB) \;=\; \sum_{i=1}^n \sum_{j=1}^m A_{i,j} B_{j,i} \;=\; \sum_{j=1}^m \sum_{i=1}^n B_{j,i} A_{i,j},$$

by swapping the order of summation. This equals $\mathrm{tr}(BA)$ because the jth diagonal entry of BA is $\sum_{i=1}^n B_{j,i} A_{i,j}$. $\square$

Fact 12 *Let A be a symmetric, $d \times d$ matrix with eigenvalues $\{\lambda_1, \ldots, \lambda_d\}$. Then $\mathrm{tr}(A) = \sum_{i=1}^d \lambda_d$.*

PROOF: Let $A = UDU^{\mathsf{T}}$ as in Fact 2. Then

$$\mathrm{tr}(A) = \mathrm{tr}(UDU^{\mathsf{T}}) = \mathrm{tr}(U^{\mathsf{T}}UD) = \mathrm{tr}(D) = \sum_{i=1}^d \lambda_i,$$

as required. $\square$

Claim 13 *Let A, B and C be symmetric $d \times d$ matrices satisfying $A \succeq 0$ and $B \preceq C$. Then $\mathrm{tr}(AB) \leq \mathrm{tr}(AC)$.*

PROOF: Suppose additionally that A is rank-one, i.e., $A = vv^{\mathsf{T}}$ for some vector v. Then

$$
\begin{aligned}
\mathrm{tr}(AB) \;&=\; \mathrm{tr}(vv^{\mathsf{T}}B) \;=\; \mathrm{tr}(v^{\mathsf{T}}Bv) \;=\; v^{\mathsf{T}}Bv \\
&\leq\; v^{\mathsf{T}}Cv \;=\; \mathrm{tr}(v^{\mathsf{T}}Cv) \;=\; \mathrm{tr}(vv^{\mathsf{T}}C) \;=\; \mathrm{tr}(AC).
\end{aligned}
$$

Now we consider the case where A has arbitrary rank. Let $A = \sum_{i=1}^d \lambda_i u_i u_i^{\mathsf{T}}$. Since we assume that $A \succeq 0$ we can set $v_i = \sqrt{\lambda_i} u_i$ and write $A = \sum_{i=1}^d v_i v_i^{\mathsf{T}}$. Then

$$
\begin{aligned}
\mathrm{tr}(AB) \;&=\; \mathrm{tr}\Big(\sum_i v_i v_i^{\mathsf{T}} B \Big) \;=\; \sum_i \mathrm{tr}(v_i v_i^{\mathsf{T}} B) \\
&\leq\; \sum_i \mathrm{tr}(v_i v_i^{\mathsf{T}} C) \;=\; \mathrm{tr}\Big(\sum_i v_i v_i^{\mathsf{T}} C \Big) \;=\; \mathrm{tr}(AC).
\end{aligned}
$$

3

Here the second and third equalities are by linearity of trace, and the inequality follows from the rank-one case. □

Corollary 14 *If $A \succeq 0$ then $\operatorname{tr}(A \cdot B) \leq \|B\| \cdot \operatorname{tr}(A)$.*

PROOF: Apply Claim 13 with $C = \|B\| \cdot I$. We get
$$\operatorname{tr}(A \cdot B) \ \leq \ \operatorname{tr}(A \cdot \|B\| \cdot I) \ \leq \ \|B\| \cdot \operatorname{tr}(A \cdot I),$$
by linearity of trace since $\|B\|$ is a scalar. □

1.4 Functions applied to matrices

For any function $f : \mathbb{R} \to \mathbb{R}$, we can extend f to a function on symmetric matrices as follows. Let $A = UDU^\mathsf{T}$ be a spectral decomposition of A where λ_i is the ith diagonal entry of D. Then we define
$$f(A) \ = \ U \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_d) \end{pmatrix} U^\mathsf{T}.$$

In other words, $f(A)$ is defined simplying by applying the function f to the eigenvalues of A.

Caution. $f(A)$ is **not** obtained applying the function f to the entries of A!

One important example of applying a function to a matrix is **powering**. Let A be any positive semi-definite matrix with spectral decomposition UDU^T. For any $c \in \mathbb{R}$ we can define A^c as follows. Let S be the diagonal matrix with $S_{i,i} = (D_{i,i})^c$. Then we define $A^c = USU^\mathsf{T}$. For example, consider the case $c = 1/2$. Note that
$$A^{1/2} \cdot A^{1/2} \ = \ USU^\mathsf{T} \cdot USU^\mathsf{T} \ = \ USSU^\mathsf{T} \ = \ UDU^\mathsf{T} \ = \ A.$$

So the square of the square root is the matrix itself, as one would expect. If A is non-singular, the matrix A^{-1} obtained by taking $c = -1$ is the same as the usual matrix inverse (by uniqueness of inverses, since $A^{-1} \cdot A = I$). So we see that the inverse of a non-singular symmetric matrix is obtained by inverting its eigenvalues.

Inequalities on real-valued functions also give us inequalities on matrices.

Claim 15 *Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ satisfy $f(x) \leq g(x)$ for all $x \in [l, u] \subset \mathbb{R}$. Let A be a symmetric matrix for which all eigenvalues lie in $[l, u]$ (i.e., $lI \preceq A \preceq uI$). Then $f(A) \preceq g(A)$.*

PROOF: Let $A = UDU^\mathsf{T}$ be a spectral decomposition of A where λ_i is the ith diagonal entry of D. We wish to show $f(A) \preceq g(A)$, i.e., $g(A) - f(A) \succeq 0$. By (2), this is equivalent to
$$v^\mathsf{T} U \begin{pmatrix} g(\lambda_1) - f(\lambda_1) & & \\ & \ddots & \\ & & g(\lambda_d) - f(\lambda_d) \end{pmatrix} Uv \ \geq \ 0 \qquad \forall v \in \mathbb{R}^d.$$
Since U is non-singular this is equivalent to
$$w^\mathsf{T} \begin{pmatrix} g(\lambda_1) - f(\lambda_1) & & \\ & \ddots & \\ & & g(\lambda_d) - f(\lambda_d) \end{pmatrix} w \ \geq \ 0 \qquad \forall w \in \mathbb{R}^d,$$
where we have simply replaced $w = Uv$. The left-hand side is simply $\sum_i w_i^2 \big(g(\lambda_i) - f(\lambda_i)\big)$, which is non-negative by our assumptions on g, f and the λ_i's. □

4

1.5 The pseudoinverse

Sometimes we would like to invert a matrix which is not quite invertible. A useful alternative is to invert the part of the matrix that is invertible (its image) and to leave alone the part of the matrix that is not invertible (its kernel).

We can do this by applying the real-valued function:

$$f(x) = \begin{cases} 1/x & (x \neq 0) \\ 0 & (x = 0) \end{cases}.$$

The function f inverts all non-zero numbers and maps 0 to 0. So if we apply f to a symmetric matrix, all non-zero eigenvalues will be inverted, and the zero eigenvalues will remain unchanged. The resulting matrix is called the **pseudoinverse** and is denoted A^+. If A is indeed invertible then $A^+ = A^{-1}$.

Suppose A is not invertible. We cannot have $A \cdot A^+ = I$ since I has full rank and A does not. However, we have the next best thing: $A \cdot A^+$ is the identity operator *on the image of A*. More precisely $A \cdot A^+$ is the orthogonal projection onto the image of A, which we denote $I_{\text{im } A}$. In other words, for every v in the image of A, we have $A \cdot A^+ v = v$ and for every v in the kernel of A we have $A \cdot A^+ v = 0$.

Claim 16 *Let A and $B_1, \ldots, B_k$ be symmetric, positive semi-definite matices of the same size. Let $A^{+/2} = (A^+)^{1/2}$ denote the square root of the pseudoinverse of A. Let ℓ and u be arbitrary scalars. Suppose that $\text{im}(B_i)$ is a subspace of $\text{im}(A)$ for all i. Then*

$$\ell A \preceq \sum_{i=1}^{k} B_i \preceq uA \qquad \Longleftrightarrow \qquad \ell I_{\text{im } A} \preceq \sum_{i=1}^{k} A^{+/2} B_i A^{+/2} \preceq u I_{\text{im } A}$$

PROOF: $\Longrightarrow$: Since $A^{+/2} A A^{+/2} = I_{\text{im } A}$, this follows from Fact 6.

$\Longleftarrow$: This also follows from Fact 6, because

$$A^{1/2} I_{\text{im } A} A^{1/2} = A$$

and

$$A^{1/2} A^{+/2} B_i A^{+/2} A^{1/2} = I_{\text{im } A} B_i I_{\text{im } A} = I_{\text{im } B_i} B_i I_{\text{im } B_i} = B_i,$$

where the second equality follows from the assumption that $\text{im}(B_i) \subseteq \text{im}(A)$. $\square$

1.6 Matrix exponentials

We will mainly be interested in the case $f(x) = e^x$. For any symmetric matrix A, note that e^A is positive semi-definite. Whereas in the scalar case $e^{a+b} = e^a e^b$ holds, it is not necessarily true in the matrix case that $e^{A+B} = e^A \cdot e^B$. It turns out that there is a useful inequality along these lines. It is called the Golden-Thompson inequality.

Theorem 17 (Golden-Thompson Inequality) *Let A and B be symmetric $d \times d$ matrices. Then* $\text{tr}(e^{A+B}) \leq \text{tr}(e^A \cdot e^B)$.