Linear Algebra In Machine Learning

Chunpai Wang

1 Symmetric Matrices

Theorem 1.1. *Suppose $A \in \mathbb{R}^{n \times n}$ is symmetric, then*

1. *every eigenvalue λ of A is a real number, and there exists a real orthonormal eigenvector $\mu \in \mathbb{R}^n$ corresponding to λ such that $A\mu = \lambda\mu$. Note that, eigenvalues are the roots of characteristic polynomial*
$$det(A - \lambda \mathtt{I}) = 0$$

2. *eigenvectors corresponding to distinct eigenvalues are necessarily orthogonal, that is*
$$\left.\begin{array}{c} A\mu^{(1)} = \lambda_1 \mu^{(1)} \\ A\mu^{(2)} = \lambda_2 \mu^{(2)} \\ \lambda_1 \neq \lambda_2 \end{array}\right\} \Rightarrow \mu^{(1)} \cdot \mu^{(2)} = 0$$

3. *there exists a diagonal matrix $\Lambda \in \mathbb{R}^{n \times n}$ and an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ such that*
$$A = U \Lambda U^\top,$$
which is also called symmetric eigenvalue decomposition. The diagonal entries of Λ are the eigenvalues of A and the columns of U are the corresponding eigenvectors, where
$$A\mu^{(i)} = \lambda_i \mu^{(i)}, \quad i = 1, ..., n$$

4. $$A = U\Lambda U^\top = U \Lambda U^{-1} \quad (\textit{Spectrum Theorem})$$
because U is orthogonal, which means the columns of U are orthonormal (i.e. any two of them are orthogonal and each has norm 1), for $i \neq j$, $\mu^{(i)}\mu^{(j)} = 1$ and $\|\mu^{(i)}\|_2 = 1$, then $U^\top = U^{-1}$ because $U^\top U = \mathtt{I} = U^{-1}U$.

2 Positive Definite

A matrix $A \in S^n$ (symmetric) is called positive definite if for all vectors $x \neq 0$ and $x \in \mathbb{R}^n$,
$$x^\top A x > 0$$
which implies all eigenvalues of A are positive, and we denote $A \succ 0$ and $A \in S^n_{++}$. If $-A$ is positive definite, we say A is negative definite, and write it as $A \prec 0$. If A satisfies $x^\top Ax \succeq 0$ for all $x \in \mathbb{R}^n$, then A is positive semidefinite, and denote $A \succeq 0, A \in S^n_+$.

Theorem 2.1. *Suppose $A \in S^n_+$ (symmetric positive semidefinite), then*

1. *the eigenvalues of A are nonnegative, that is $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0$, where*
$$\lambda_1 = \lambda_{max}(A) = \sup_{x \neq 0} \frac{x^\top A x}{x^\top x}$$
$$\lambda_n = \lambda_{min}(A) = \inf_{x \neq 0} \frac{x^\top A x}{x^\top x}$$
$$\lambda_{min}(A) x^\top x \leq x^\top A x \leq \lambda_{max} x^\top x \qquad \textit{for any } x$$

2. *determinant of A is greater than and equal to 0, since* $det(A) = \prod_i^n \lambda_i$

3. *trace of A is greater than and equal to 0, since* $trace(A) = \sum_i^n \lambda_i$

4. $A = M^\top M$ *for some matrix* M*, since* $A = U\Lambda U^\top = U\Lambda^{1/2}\Lambda^{1/2}U^\top$ *and* $M = \Lambda^{1/2}U^\top$

5. $x^\top Ax = 0$ *implies* $Ax = 0$*, because* $x^\top Ax = x^\top M^\top Mx = \|Mx\|_2 = 0$ *and* $Ax = M^\top Mx = M^\top 0 = 0$

6. *A is nonsingular (invertible) if and only if* $A \succ 0$

7. *any principal submatrix of A is positive semidefinite*

3 Schur Complements and Additional Characteristic of SPD Matrices

The Schur complement comes up in solving linear equations, by eliminating one block of variables. Let M be an $n \times n$ matrix written a as 2×2 block matrix

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A is $p \times p$ matrix and D is a $q \times q$ matrix, with $n = p + q$. We can try to solve the linear system

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix}$$

that is

$$Ax + By = cCx + Dy = d$$

Now, we assume that D is invertible, we first solve for y getting

$$y = D^{-1}(d - Cx)$$

and substitute y back into first equation, we get

$$Ax + B(D^{-1}(d - Cx)) = c$$

that is

$$(A - BD^{-1}C)x = c - BD^{-1}d$$

Now, if the matrix $A - BD^{-1}C$ is invertible, then we obtain the solution to our system

$$x = (A - BD^{-1}C)^{-1}(c - BD^{-1}d)$$

$$y = D^{-1}(d - Cx) = D^{-1}(d - C(A - BD^{-1}C)^{-1}(c - BD^{-1}d))$$

The matrix $A - BD^{-1}C$ is called the Schur complement of D in M. If A is invertible, then by eliminating x first using the first equation, we find that the Schur complement of A in M is $D - CA^{-1}B$.

The above equations can be expanded and written as

$$x = (A - BD^{-1}C)^{-1}c - (A - BD^{-1}C)^{-1}BD^{-1}d$$

$$\begin{aligned} y &= D^{-1}d - D^{-1}C(A - BD^{-1}C)^{-1}c + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1}d \\ &= -D^{-1}C(A - BD^{-1}C)^{-1}c + (D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1})d \end{aligned}$$

According to

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} c \\ d \end{bmatrix}$$

we can see that

$$
\begin{aligned}
\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} (A-BD^{-1}C)^{-1} & -(A-BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A-BD^{-1}C)^{-1} & (D^{-1}+D^{-1}C(A-BD^{-1}C)^{-1}BD^{-1}) \end{bmatrix} \\
&= \begin{bmatrix} (A-BD^{-1}C)^{-1} & 0 \\ -D^{-1}C(A-BD^{-1}C)^{-1} & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \\
&= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} (A-BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix}
\end{aligned}
$$

Since

$$(XYZ)^{-1} = Z^{-1}Y^{-1}X^{-1}$$

and

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

we can have

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} (A-BD^{-1}C) & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}$$

We can see that

$$\det M = \det(A-BD^{-1}C)\ \det D$$

If A is invertible, then we can use the Schur complement, $D-CA^{-1}B$ of A to obtain the following factorization of M:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

If $D-CA^{-1}B$ is invertible, we can invert all three matrices above and we get another formula for the inverse of M in terms of $D-CA^{-1}B$, namely

$$
\begin{aligned}
\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & (D-CA^{-1}B)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \\
&= \begin{bmatrix} A^{-1}+A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix}
\end{aligned}
$$

3.1 Matrix Inversion Lemma

If A, D and both Schur complements $A-BD^{-1}C$ and $D-CA^{-1}B$ are all invertible, by comparing the two expression for M^{-1}, we get the formula

$$(A-BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1}$$

and

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} (A-BD^{-1}C)^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{bmatrix}$$

If we set $D = I$ and change B to $-B$ we get

$$(A+BC)^{-1} = A^{-1} - A^{-1}B(I+CA^{-1}B)^{-1}CA^{-1}$$

a formula known as the matrix inversion lemma, or the Sherman-Woodbury-Morrison formula.

The $(A+BC)^{-1}$ can be used to solve the problem

$$(A+BC)x = b$$

where $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{p\times n}$. If p is small, the it gives us a method for solving $(A+BC)x = b$, which is same as problem

$$Ax + By = b, \quad y = Cx$$

or, in matrix form

$$\begin{bmatrix} A & B \\ C & -I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix}$$

It is more efficient to solve a larger set of equations whenA, B, C are much sparse than $A + BC$. From $Ax + By = b$, we get $x = A^{-1}(b - By)$, and substitute into the $y = Cx$, we have $y = C(A^{-1}(b - By))$ and

$$y = (I + CA^{-1}B)^{-1}CA^{-1}b$$

$$x = (A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1})b$$

Since $x = (A + BC)^{-1}b$, we have

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I + CA^{-1}B)^{-1}CA^{-1}$$

3.2 Additional Characteristics of SPD Matrices

Consider a matrix $X \in S^n$ partitioned as

$$X = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix}$$

where $D \in S^k$. If $\det D \neq 0$ (D is invertible), the Schur complement of D is matrix

$$S = A - BD^{-1}B^\top.$$

If $\det A \neq 0$ (A is invertible), the Schur complement of A is matrix

$$S = D - B^\top A^{-1}B.$$

Proposition 3.0.1. *For any symmetric matrix, M, of the form*

$$M = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix},$$

if D is invertible then the following properties hold:

1. *$M \succ 0$ if and only if $D \succ 0$ and $A - BD^{-1}B^\top \succ 0$.*
2. *If $D \succ 0$, then $M \succeq 0$ if and only if $A - BD^{-1}B^\top \succeq 0$.*

Proposition 3.0.2. *For any symmetric matrix, M, of the form*

$$M = \begin{bmatrix} A & B \\ B^\top & D \end{bmatrix},$$

if A is invertible then the following properties hold:

1. *$M \succ 0$ if and only if $A \succ 0$ and $D - B^\top A^{-1}B \succ 0$.*
2. *If $A \succ 0$, then $M \succeq 0$ if and only if $D - B^\top A^{-1}B \succeq 0$.*

4 Pseudo-Inverses and SVD

5 PCA and Multivariate Gaussian

References

[1] The Schur Complement and Symmetric Positive Semidefinite (and Definite) Matrices www.cis.upenn.edu/~jean/schur-comp.pdf.

[2] Appendix A.5.5 and C.4.3. Boyd, Stephen, and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.

[3] Quadratic Functions, Optimization, and Quadratic Forms https://ocw.mit.edu/courses/sloan-school-of-management/15-084j-nonlinear-programming-spring-2004/lecture-notes/lec4_quad_form.pdf

[4]