NUS | IE6001 : Foundations of Optimization

Linear Algebra Refresher – AY2022/2023

Below is some of the *keywords* or *concepts* that will show up in the second half of the course. Please make sure to familiarize yourselves with them beforehand. Suggested reference is provided at the end of this document.

Eigenvalues and eigenvectors: A vector $v \in \mathbb{R}^n \setminus \{0\}$ and a scalar $\lambda \in \mathbb{R}$ is said to be pair of an eigenvector and an eigenvalue of a matrix $A \in \mathbb{R}^{n\times n}$ if $Av = \lambda v$, or equivalently, $(A - \lambda\mathbb{I})v = 0$.

- If A is symmetric, then all of its eigenvalues are real numbers.

Eigenvalue decomposition: A symmetric matrix $A \in \mathbb{S}^n$ can be decomposed as

$$A = Q\Lambda Q^{-1} = Q\Lambda Q^\top,$$

where $\Lambda \in \mathbb{S}^n$ is a diagonal matrix whose entries are the eigenvalues of A and $Q \in \mathbb{R}^{n\times n}$ is an orthogonal matrix (*i.e.,* $Q^\top Q = QQ^\top = \mathbb{I}$) whose columns are the eigenvectors of A.

- $A^2 = (Q\Lambda Q^\top)(Q\Lambda Q^\top) = Q\Lambda^2 Q^\top$. To elaborate, when A is squared, the eigenvalues are squared whereas the eigenvectors remain the same.

Positive definieness: A symmetric matrix $A \in \mathbb{S}^n$ is said to be positive definite ($A \succ 0$) *if*

- all of its eigenvalues are strictly positive, or equivalently,
- $v^\top Av > 0$ for all $v \in \mathbb{R}^n \setminus \{0\}$.

Positive semidefinieness: A symmetric matrix $A \in \mathbb{S}^n$ is said to be positive semidefinite ($A \succeq 0$) *if*

- all of its eigenvalues are nonnegative, or equivalently,
- if $v^\top Av \geq 0$ for all $v \in \mathbb{R}^n$.

Principle square root: For any positive semidefinite matrix $A \in \mathbb{S}^n$, there exists a (unique) positive semidefinite $B \in \mathbb{S}^n$ which satisfies $A = BB$. We then say that B is the principle square root of A.

- Note that the principle square root of A can be written as $QSQ^\top$, where $Q\Lambda Q^\top$ is an eigenvalue decomposition of A and $S \in \mathbb{S}^n$ is a diagonal matrix with $s_{ii} = \sqrt{\lambda_{ii}}$, $i = 1, \ldots, n$.

Trace: For a square matrix matrix $A \in \mathbb{R}^{n\times n}$, $\operatorname{tr}(A)$ denotes the trace of A, which is defined as the sum of its main diagonal elements. In other words, $\operatorname{tr}(A) = a_{11} + a_{22} + \ldots + a_{nn}$.

- For any $A \in \mathbb{R}^{n\times n}$, $\operatorname{tr}(A) = \operatorname{tr}(A^\top)$.
- For any $A \in \mathbb{R}^{n\times n}$ and any $c \in \mathbb{R}$, $\operatorname{tr}(cA) = c\operatorname{tr}(A)$.
- For any $A, B \in \mathbb{R}^{n\times n}$, $\operatorname{tr}(A + B) = \operatorname{tr}(A) + \operatorname{tr}(B)$.
- For any $A, B \in \mathbb{R}^{m\times n}$, $\operatorname{tr}(A^\top B) = \operatorname{tr}(AB^\top) = \operatorname{tr}(B^\top A) = \operatorname{Tr}(BA^\top) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}b_{ij}$. Each of these quantities is often referred to as a dot/an inner product between matrices A and B.
- (Cyclic Property) For any $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times o}, C \in \mathbb{R}^{o\times m}$, $\operatorname{tr}(ABC) = \operatorname{tr}(CAB) = \operatorname{tr}(BCA)$. This property can be generalized when there is a greater number of matrices.
- For any $A \in \mathbb{R}^{n\times n}$, $\operatorname{tr}(A)$ equals to the sum of its eigenvalues. As a side note, the determinant of A equals to the product of its eigenvalues.

Schur complement: Let $M \in \mathbb{S}^{p+q}$ be a symmetric block matrix of the form

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

for some $A \in \mathbb{R}^{p\times p}, B \in \mathbb{R}^{p\times q}, C \in \mathbb{R}^{q\times q}$. $A - BC^{-1}B^\top$ is the Schur complement of C in M.

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

Common vector derivatives: For a fixed vector $x \in \mathbb{R}^n$ and a matrix $A \in \mathbb{R}^{n\times n}$, we have

- if $f(x) = c^\top x$, then $\nabla_x f(x) = c$,
- if $f(x) = x^\top x$, then $\nabla_x f(x) = 2x$,
- if $f(x) = x^\top Ax$, then $\nabla_x f(x) = (A + A^\top)x$.

References

[1] Petersen, K. B. and Pedersen, M. S. *The Matrix Cookbook*, 2012.