APPENDIX

# LINEAR ALGEBRA

# A

## CHAPTER OUTLINE



## A.1 PROPERTIES OF MATRICES

Let $A, B, C$, and $D$ be matrices of appropriate sizes. Invertibility is always assumed, whenever a matrix inversion is performed. The following properties hold true.

- $(AB)^T = B^T A^T$.
- $(AB)^{-1} = B^{-1}A^{-1}$.
- $(A^T)^{-1} = (A^T)^{-1}$.
- $\text{trace}\{AB\} = \text{trace}\{BA\}$.
- From the previous, we readily get

$$\text{trace}\{ABC\} = \text{trace}\{CAB\} = \text{trace}\{BCA\}.$$

- $\det(AB) = \det(A)\det(B)$, where $\det(\cdot)$ denotes the determinant of a square matrix. As a consequence, the following is also true.
- $\det(A^{-1}) = \frac{1}{\det(A)}$.
- Let $A$ and $B$ be two $m \times l$ matrices. Then

$$\det(I_m + AB^T) = \det(I_l + A^T B).$$

  A by-product is the following

$$\det(I + \boldsymbol{a}\boldsymbol{b}^T) = 1 + \boldsymbol{a}^T \boldsymbol{b},$$

  where $\boldsymbol{a},\ \boldsymbol{b} \in \mathbb{R}^l$.


**Machine Learning. http://dx.doi.org/10.1016/B978-0-12-801522-3.09984-5**

### *Matrix inversion lemmas*

- Woodbury's identity:

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

- $(I + AB)^{-1}A = A(I + BA)^{-1}$.
- $(A^{-1} + B^T C^{-1}B)^{-1}B^T C^{-1} = AB^T(BAB^T + C)^{-1}$.
- The following two inversion lemmas for partitioned matrices are particularly useful:

$$\begin{bmatrix} A & D \\ C & B \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + E\Delta^{-1}F & -E\Delta^{-1} \\ -\Delta^{-1}F & \Delta^{-1} \end{bmatrix}$$

  where $\Delta := B - CA^{-1}D$, $E := A^{-1}D$, $F := CA^{-1}$.
  Also

$$\begin{bmatrix} A & D \\ C & B \end{bmatrix}^{-1} = \begin{bmatrix} \Delta^{-1} & -\Delta^{-1}E \\ -F\Delta^{-1} & B^{-1} + F\Delta^{-1}E \end{bmatrix},$$

  where $\Delta := A - DB^{-1}C$, $E := DB^{-1}$, $F := B^{-1}C$. Matrix $\Delta$ is also known as the *Schur complement*.
  For complex matrices, the transposition becomes the Hermitian one.

### *Matrix derivatives*

- $\frac{\partial \boldsymbol{a}^T\boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^T\boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a}$.
- $\frac{\partial \boldsymbol{x}^T A\boldsymbol{x}}{\partial \boldsymbol{x}} = (A + A^T)\boldsymbol{x}$,
  which becomes $2A\boldsymbol{x}$ if $A$ is symmetric.
- $\frac{\partial (AB)}{\partial x} = \frac{\partial A}{\partial x}B + A\frac{\partial B}{\partial x}$.
- $\frac{\partial A^{-1}}{\partial x} = -A^{-1}\frac{\partial A}{\partial x}A^{-1}$.
- $\frac{\partial \ln|A|}{\partial x} = \mathrm{trace}\{A^{-1}\frac{\partial A}{\partial x}\}$,
  where $|\cdot|$ denotes the determinant, and matrices $A$ and $B$ are functions of $x$.
- $\frac{\partial\, \mathrm{trace}\{AB\}}{\partial A} = B^T$,
  where $\left[\frac{\partial}{\partial A}\right]_{ij} := \frac{\partial}{\partial A(i,j)}$.
- $\frac{\partial\, \mathrm{trace}\{A^T B\}}{\partial A} = B$.
- $\frac{\partial\, \mathrm{trace}\{ABA^T\}}{\partial A} = A(B + B^T)$.
- $\frac{\partial \ln|A|}{\partial A} = (A^T)^{-1}$.
- $\frac{\partial A\boldsymbol{x}}{\partial \boldsymbol{x}} = A^T$,
  where by definition $\left[\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right]_{ij} = \frac{\partial y_i}{\partial x_j}$.
  More on matrix identities can collectively be found in [2].

## A.2 POSITIVE DEFINITE AND SYMMETRIC MATRICES

- An $l \times l$ real symmetric matrix $A$ is called *positive definite* if, for *every* nonzero vector $\boldsymbol{x}$, the following is true:

$$\boldsymbol{x}^T A \boldsymbol{x}. \tag{A.1}$$

If equality with zero is allowed, $A$ is called *positive semidefinite*. The definition is extended to complex Hermitian symmetric matrices, $A$, if $\forall \boldsymbol{x} \in \mathbb{C}$,

$$\boldsymbol{x}^H A \boldsymbol{x} > 0.$$

- It is easy to show that all eigenvalues of such a matrix are positive. Indeed, let $\lambda_i$ be one eigenvalue and $\boldsymbol{u}_i$ the corresponding unit norm eigenvector ($\boldsymbol{u}_i^T \boldsymbol{u}_i = 1$). Then, by the respective definitions

$$A\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \tag{A.2}$$

or

$$0 < \boldsymbol{u}_i^T A \boldsymbol{u}_i = \lambda_i. \tag{A.3}$$

Since the determinant of a matrix is equal to the product of its eigenvalues, we conclude that the determinant of a positive definite matrix is also positive.
- Let $A$ be an $l \times l$ symmetric matrix, $A^T = A$. Then the eigenvectors corresponding to distinct eigenvalues are orthogonal. Indeed, let $\lambda_i \neq \lambda_j$ be two such eigenvalues. From the definitions we have

$$A\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \tag{A.4}$$

$$A\boldsymbol{u}_j = \lambda_j \boldsymbol{u}_j. \tag{A.5}$$

Multiplying Eq. (A.4) on the left by $\boldsymbol{u}_j^T$ and the transpose of Eq. (A.5) on the right by $\boldsymbol{u}_i$, we obtain

$$\boldsymbol{u}_j^T A \boldsymbol{u}_i - \boldsymbol{u}_j^T A^T \boldsymbol{u}_i = 0 = (\lambda_i - \lambda_j)\boldsymbol{u}_j^T \boldsymbol{u}_i. \tag{A.6}$$

Thus, $\boldsymbol{u}_j^T \boldsymbol{u}_i = 0$. Furthermore, it can be shown that even if the eigenvalues are not distinct, we can still find a set of orthogonal eigenvectors. The same is true for Hermitian matrices, in case we deal with more general complex-valued matrices.
- Based on this, it is now straightforward to show that a symmetric matrix $A$ can be diagonalized by the similarity transformation

$$U^T A U = \Lambda, \tag{A.7}$$

where matrix $U$ has as its columns the unit norm eigenvectors ($\boldsymbol{u}_i^T \boldsymbol{u}_i = 1$) of $A$, that is,

$$U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_l], \tag{A.8}$$

and $\Lambda$ is the diagonal matrix with elements being the corresponding eigenvalues of $A$. From the orthonormality of the eigenvectors, it is obvious that $U^T U = I$ and $UU^T = I$; that is, $U$ is an orthogonal matrix, $U^T = U^{-1}$. The proof is similar for Hermitian complex matrices as well.

## A.3 WIRTINGER CALCULUS

Let a function

$$f : \mathbb{C} \longmapsto \mathbb{C}, \tag{A.9}$$

and let

$$f(z) = f_r(x, y) + jf_i(x, y), \quad z = x + jy,\ x, y \in \mathbb{R}.$$

Then, the Wirtinger derivative or $W$-derivative of $f$ at a point $c \in \mathbb{C}$ is defined as

$$\frac{\partial f}{\partial z}(c) := \frac{1}{2}\left(\frac{\partial f_r}{\partial x}(c) + \frac{\partial f_i}{\partial y}(c)\right) + \frac{j}{2}\left(\frac{\partial f_i}{\partial x}(c) - \frac{\partial f_r}{\partial y}(c)\right), \tag{A.10}$$

and the conjugate Wirtinger derivative or $CW$-derivative as

$$\frac{\partial f}{\partial z^*}(c) := \frac{1}{2}\left(\frac{\partial f_r}{\partial x}(c) - \frac{\partial f_i}{\partial y}(c)\right) + \frac{j}{2}\left(\frac{\partial f_i}{\partial x}(c) + \frac{\partial f_r}{\partial y}(c)\right), \tag{A.11}$$

provided that the involved derivatives exist. In this case, we say that $f$ is differentiable in the real sense. This definition has been extended to gradients, for vector-valued functions as well as to Frechét derivatives in complex Hilbert spaces [1]. The following properties are valid:

- If $f$ has a Taylor series expansion with respect to $z$ (i.e., it is holomorphic) around $c$, then

$$\frac{\partial f}{\partial z^*}(c) = 0.$$

- If $f$ has a Taylor series expansion with respect to $z^*$ around $c$, then

$$\frac{\partial f}{\partial z}(c) = 0.$$

- $\left(\frac{\partial f}{\partial z}(c)\right)^* = \frac{\partial f^*}{\partial z^*}(c)$.
- $\left(\frac{\partial f}{\partial z^*}(c)\right)^* = \frac{\partial f^*}{\partial z}(c)$.
- Linearity: If $f$ and $g$ are differentiable in the real sense, then

$$\frac{\partial(af + bg)}{\partial z}(c) = a\frac{\partial f}{\partial z}(c) + b\frac{\partial g}{\partial z}(c),$$

  and

$$\frac{\partial(af + bg)}{\partial z^*}(c) = a\frac{\partial f}{\partial z^*}(c) + b\frac{\partial g}{\partial z^*}(c).$$

- Product rule:

$$\frac{\partial(fg)}{\partial z}(c) = \frac{\partial f}{\partial z}(c)g(c) + f(c)\frac{\partial g}{\partial z}(c),$$

  and

$$\frac{\partial(fg)}{\partial z^*}(c) = \frac{\partial f}{\partial z^*}(c)g(c) + f(c)\frac{\partial g}{\partial z^*}(c).$$

- Division rule: If $g(c) \neq 0$,

$$\left.\frac{\partial \left(\frac{f}{g}\right)}{\partial z}\right|_c = \frac{\frac{\partial f}{\partial z}(c)g(c) - f(c)\frac{\partial g}{\partial z}(c)}{g^2(c)},$$

and

$$\left.\frac{\partial \left(\frac{f}{g}\right)}{\partial z^*}\right|_c = \frac{\frac{\partial f}{\partial z^*}(c)g(c) - f(c)\frac{\partial g}{\partial z^*}(c)}{g^2(c)}.$$

- Let

$$f : \mathbb{C} \longmapsto \mathbb{R}.$$

If $z_o$ is a local optimal of the real-valued $f$, then

$$\frac{\partial f}{\partial z}(z_o) = \frac{\partial f}{\partial z^*}(z_o) = 0.$$

Indeed, in this case $f_i = 0$ and the Wirtinger derivative becomes

$$\frac{\partial f}{\partial z}(z_o) = \frac{1}{2}\left(\frac{\partial f_r}{\partial x}(z_o) - j\frac{\partial f_r}{\partial y}(z_o)\right) = 0,$$

as at the optimal point both derivatives on the left hand side become zero. Similar is the proof for the $CW$-derivative.

## REFERENCES

[1] P. Bouboulis, S. Theodoridis, Extension of Wirtinger's calculus to reproducing kernel Hilbert spaces and the complex kernel LMS, IEEE Trans. Signal Process. 53(3) (2011) 964-978.
[2] K.B. Petersen, M.S. Pedersen, The Matrix Cookbook, 2013, http://www2.imm.dtu.dk/pubdb/p.php?3274.