APPENDIX

LINEAR ALGEBRA

A

A.1 PROPERTIES OF MATRICES

In this appendix, some useful properties and formulas from linear algebra are summarized. The readers are assumed to have taken a first course related to linear algebra. Hence, the material here is only for a quick reference and not for "learning" the topic.

Inner product: Given two l-dimensional vectors, $\boldsymbol{x} \in \mathbb{R}^l$ and $\boldsymbol{y} \in \mathbb{R}^l$, their inner product is defined as

$$\boldsymbol{x}^T \boldsymbol{y} = x_1 y_1 + \cdots + x_l y_l = \sum_{i=1}^{l} x_i y_i.$$

Obviously, by its definition, the inner product is a symmetric operation, i.e., $\boldsymbol{x}^T \boldsymbol{y} = \boldsymbol{y}^T \boldsymbol{x}$.

In the case of complex vectors, $\boldsymbol{x} \in \mathbb{C}^l$ and $\boldsymbol{y} \in \mathbb{C}^l$, the inner product is defined as

$$\boldsymbol{x}^H \boldsymbol{y} = x_1^* y_1 + \cdots + x_l^* y_l = \sum_{i=1}^{l} x_i^* y_i.$$

where H denotes the Hermitian operation, i.e., transposition combined with conjugation.

Linear independence: Given a set of $m \leq l$ l-dimensional vectors, $\boldsymbol{x}_i \in \mathbb{R}^l$, $i = 1, 2, \ldots, m$, we say that they are linearly independent if

$$\sum_{i=1}^{m} a_i \boldsymbol{x}_i = 0$$

can only be true if $a_i = 0$, $i = 1, 2, \ldots, m$. If some or all of the as are nonzero, the above summation cannot be zero. If, on the other hand, there exist some a_i which are not equal to zero and the above sum is zero, then the vectors are linearly dependent (one can be expressed in terms of the rest).

Furthermore, if $m > l$, then the m vectors are necessarily linearly dependent.

Span of a set of vectors: Consider a set of, say, m vectors, $\boldsymbol{x}_i$, $i = 1, 2, \ldots, m$. Form the set of *all* possible linear combinations of them, i.e.,

$$S = \left\{ \boldsymbol{x} : \boldsymbol{x} = \sum_{i=1}^{m} a_i \boldsymbol{x}_i, \ \forall a_i \in \mathbb{R}, \ i = 1, 2, \ldots, m \right\}.$$

The set S is known as the span of the m vectors, and we write

$$S = \text{span}\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m\}.$$

Moreover, S is a subspace $S \subseteq \mathbb{R}^l$. Also, if the m vectors are linearly independent, the dimension of S is equal to m.

Basis of a subspace: Consider a set of vectors in a subspace $V \subseteq \mathbb{R}^l$, i.e., $\boldsymbol{e}_i \in V$, $i = 1, 2, \ldots, m$. We say that this set comprises a basis of V if:

- the vectors are linearly independent, and
- $\text{span}\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m\} = V$.

Furthermore, if they are mutually orthogonal, i.e., $\boldsymbol{e}_i^T \boldsymbol{e}_j = 0$, $i \neq j$, we say that the basis is orthogonal. Moreover, if their respective norms are equal to one, i.e., $\|\boldsymbol{e}_i\| = 1$, $i = 1, 2, \ldots, m$, the basis is known as orthonormal. Obviously, in this case, the dimension of V is equal to m.

Rank of a matrix: Consider the $l \times l$ matrix A. We define the rank of the matrix as the maximum number, r, of linearly independent rows or columns of the matrix. If $r = l$ we say that the matrix is a full rank one. The definition applies also to matrices that are not square; that is, for $m \times l$ matrices. In such cases, $r \leq \min\{m, l\}$.

Inverse of a matrix: The inverse of a square $l \times l$ matrix A, denoted as A^{-1}, is the $l \times l$ matrix such that the following matrix products hold true:

$$A^{-1}A = AA^{-1} = I,$$

where I is the identity matrix, having all its elements equal to zero, except those in the main diagonal, which are all equal to 1. For the inverse of a square matrix to exist, its rows (columns) must be linearly independent. That is, it must be a full rank matrix.

Transpose of a matrix: The transpose of a matrix A, denoted as A^T, is the new matrix whose columns are the rows of A. That is, if

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1l} \\ a_{21} & a_{22} & \ldots & a_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \ldots & a_{ll} \end{bmatrix},$$

its transpose matrix is equal to

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{l1} \\ a_{12} & a_{22} & \ldots & a_{l2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1l} & a_{2l} & \ldots & a_{ll} \end{bmatrix}.$$

In another notation, if matrix A is written as a concatenation of its column vectors, i.e.,

$$A = [\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_l],$$

its transpose is written as

$$A^T = \begin{bmatrix} \boldsymbol{a}_1^T \\ \boldsymbol{a}_2^T \\ \vdots \\ \boldsymbol{a}_l^T \end{bmatrix}.$$

Trace of a matrix: The trace of an $l \times l$ matrix is defined as the sum of the elements across its main diagonal, i.e.,

$$\text{trace}\{A\} = \sum_{i=1}^{l} a_{ii}.$$

Orthogonal projection to a subspace: Let us consider a set of linearly independent vectors $\boldsymbol{a}_i \in \mathbb{R}^l$, $i = 1, 2, \dots, m$, where $m \leq l$. These vectors generate the subspace span$\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_m\} \subseteq \mathbb{R}^l$. Consider a vector $\boldsymbol{x} \in \mathbb{R}^l$, which, in general, does not lie on this subspace. Then the orthogonal projection of the vector to the subspace is the vector given by (e.g., [3])

$$P_{\{\boldsymbol{a}_i\}}(\boldsymbol{x}) = A(A^T A)^{-1} A^T \boldsymbol{x},$$

where A is the $m \times l$ matrix with as columns the above vectors, i.e.,

$$A = [\boldsymbol{a}_1, \dots, \boldsymbol{a}_m].$$

Below some important matrix properties and formulas are summarized.

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

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

Note that if a matrix is given as the outer product of two vectors, i.e., $A = \boldsymbol{a}\boldsymbol{b}^T$, then, based on the previous property, its trace is equal to the inner product of the involved vectors, i.e.,

$$\text{trace}\{A\} = \text{trace}\{\boldsymbol{b}^T \boldsymbol{a}\} = \boldsymbol{b}^T \boldsymbol{a} = \boldsymbol{a}^T \boldsymbol{b},$$

which is also easily checked by the definition of the trace and the inner product operation.

- $\det(AB) = \det(A)\det(B)$, 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 $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^l$.

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 $\Delta := B - CA^{-1}D$, $\;E := A^{-1}D$, $\;F := CA^{-1}$, and

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

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

EIGENVALUES AND EIGENVECTORS

The notion of eigenvalues and eigenvectors of a square matrix is among the most important ones and they represent a basic tool in linear algebra. Eigenvalues/eigenvectors are used in various parts in the book.

Let us consider a square $l \times l$ matrix A. Any pair $(\lambda \in \mathbb{C}, \boldsymbol{u} \in \mathbb{C}^l)$ is said to comprise an eigenvalue-eigenvector pair if it satisfies the following equation:

$$A\boldsymbol{u} = \lambda \boldsymbol{u}.$$

The above can equivalently be written as

$$(A - \lambda I)\boldsymbol{u} = \mathbf{0},$$

where I is the identity matrix. In order for the above equation to be true for nonzero vectors $\boldsymbol{u}$, the determinant of the corresponding matrix must necessarily be zero. If this were not zero, the matrix would be invertible and then the solution for $\boldsymbol{u}$ would be zero. Hence, the goal is to find those values of λ that make the corresponding determinant zero, i.e.,

$$|A - \lambda I| = 0.$$

The above can be shown to be an lth-degree polynomial of λ, and hence it is an equation with l roots, i.e., $\lambda_1, \dots, \lambda_l$. These are known as the l eigenvalues of matrix A. The roots/eigenvalues can be discrete or have a multiplicity, as is always the case with the roots of a polynomial. They can also be real or complex. For each eigenvalue, one can now define a corresponding eigenvector so that

$$A\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \; i = 1, 2, \dots, l.$$

Because one can multiply both sides of the above by a scalar and the equation is still valid, it is common to consider the unit norm, normalized eigenvectors, i.e., $\frac{\boldsymbol{u}_i}{\|\boldsymbol{u}_i\|}$. That is, for the eigenvectors, it is their direction in space and not their norm that is of significance.

Two important properties of the eigenvalues/eigenvectors of interest to us in the book are the following:

$$\operatorname{trace}\{A\} = \sum_{i=1}^{l} \lambda_i, \tag{A.1}$$

$$|A| = \prod_{i=1}^{l} \lambda_i. \tag{A.2}$$

Hence, if a matrix is invertible (full rank), then all its eigenvalues must be nonzero.

MATRIX DERIVATIVES AND GRADIENTS

Matrix derivatives and gradients play an important role in various parts of the book. Given a function of a vector quantity, i.e., $f(\boldsymbol{x})$, its gradient or derivative with respect to $\boldsymbol{x}$ is the vector of the partial derivatives of f with respect to all the components of the vector, i.e.,

$$\nabla_{\boldsymbol{x}} f := \frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} := \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_l} \end{bmatrix}.$$

The following formulas hold true:

- $\frac{\partial \boldsymbol{a}^T \boldsymbol{x}}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{x}^T \boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a}$.
- $\frac{\partial \boldsymbol{x}^T A \boldsymbol{x}}{\partial \boldsymbol{x}} = (A + A^T)\boldsymbol{x}$,
  which becomes $2A\boldsymbol{x}$ if A is symmetric.
- $\frac{\partial (AB)}{\partial x} = \frac{\partial A}{\partial x} B + A \frac{\partial B}{\partial x}$.
- $\frac{\partial A^{-1}}{\partial x} = -A^{-1} \frac{\partial A}{\partial x} A^{-1}$.
- $\frac{\partial \ln |A|}{\partial x} = \text{trace}\{A^{-1} \frac{\partial A}{\partial x}\}$,
  where $|\cdot|$ denotes the determinant, and matrices A and B are functions of the scalar x.
- $\frac{\partial \, \text{trace}\{AB\}}{\partial A} = B^T$,
  where the derivative of a function f with respect to a matrix is defined as the matrix whose elements are the corresponding partial derivatives of the function with respect to each one of the elements of the matrix, i.e., $\left[\frac{\partial f}{\partial A}\right]_{ij} := \frac{\partial f}{\partial A(i,j)}$.
- $\frac{\partial \, \text{trace}\{A^T B\}}{\partial A} = B$.
- $\frac{\partial \, \text{trace}\{ABA^T\}}{\partial A} = A(B + B^T)$.
- $\frac{\partial \ln |A|}{\partial A} = (A^T)^{-1}$.
- $\frac{\partial A\boldsymbol{x}}{\partial \boldsymbol{x}} = A^T$,
  where, by definition, the derivative of a vector quantity with respect to another vector is defined as the matrix $\left[\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}}\right]_{ij} = \frac{\partial y_i}{\partial x_j}$. In a more analytic form,

$$\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_l} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x_l} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_l}{\partial x_1} & \frac{\partial y_l}{\partial x_1} & \cdots & \frac{\partial y_l}{\partial x_l} \end{bmatrix}.$$

  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 $\boldsymbol{x} \in \mathbb{R}^l$, the following is true:

$$\boldsymbol{x}^T A \boldsymbol{x} > 0. \tag{A.3}$$

If equality with zero is also 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 the eigenvalues of such a matrix are positive real numbers. We will give the proof working with real vectors. The generalization to the complex number case is straightforward by replacing transposition with the Hermitian operation.
  Let λ_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,

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

  or

  $$0 < \boldsymbol{u}_i^T A \boldsymbol{u}_i = \lambda_i. \tag{A.5}$$

  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

  $$A\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, \tag{A.6}$$
  $$A\boldsymbol{u}_j = \lambda_j \boldsymbol{u}_j. \tag{A.7}$$

  Multiplying Eq. (A.6) on the left by $\boldsymbol{u}_j^T$ and the transpose of Eq. (A.7) on the right by $\boldsymbol{u}_i$, we obtain

  $$\boldsymbol{u}_j^T A \boldsymbol{u}_i - \boldsymbol{u}_j^T A^T \boldsymbol{u}_i = 0 = (\lambda_i - \lambda_j)\boldsymbol{u}_j^T \boldsymbol{u}_i. \tag{A.8}$$

  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 the previous property, it is now straightforward to show that a symmetric matrix A can be diagonalized by the similarity transformation

  $$U^T A U = \Lambda, \tag{A.9}$$

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

  $$U = [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_l], \tag{A.10}$$

  and Λ 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 $U U^T = I$; that is, U is an orthogonal matrix, $U^T = U^{-1}$. The proof is similar for Hermitian complex matrices.

A.3 WIRTINGER CALCULUS

Consider a function

$$f : \mathbb{C} \longmapsto \mathbb{C}, \tag{A.11}$$

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

$$\frac{\partial f}{\partial z}(c) := \frac{1}{2}\left(\frac{\partial f_r}{\partial x}(c) + \frac{\partial f_i}{\partial y}(c)\right) + \frac{j}{2}\left(\frac{\partial f_i}{\partial x}(c) - \frac{\partial f_r}{\partial y}(c)\right), \tag{A.12}$$

and the conjugate Wirtinger derivative, or CW-derivative, is defined as

$$\frac{\partial f}{\partial z^*}(c) := \frac{1}{2}\left(\frac{\partial f_r}{\partial x}(c) - \frac{\partial f_i}{\partial y}(c)\right) + \frac{j}{2}\left(\frac{\partial f_i}{\partial x}(c) + \frac{\partial f_r}{\partial y}(c)\right), \tag{A.13}$$

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: We have
$$\frac{\partial (fg)}{\partial z}(c) = \frac{\partial f}{\partial z}(c)g(c) + f(c)\frac{\partial g}{\partial z}(c)$$

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. The proof for the CW-derivative is similar.

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, http://www2.imm.dtu.dk/pubdb/p.php?3274, 2013.
[3] G. Strang, Introduction to Linear Algebra, 5th ed., Wellesley-Cambridge Press and SIAM, 2016.