A Brief Review of Linear Algebra and Optimization

**Notation:** Vectors will generally be represented in lower-case bold, e.g., \( \mathbf{x} \), and matrices in upper-case bold, e.g., \( \mathbf{A} \). Scalars will usually be represented in lower case italics, e.g., \( \alpha \), \( \mathbf{d} \).

1 Vectors

In this course, by default vectors will be column vectors such as

\[
\mathbf{x} = \begin{bmatrix} x_1 \\
x_2 \\
\vdots \\
x_N \end{bmatrix},
\]

which has length \( N \). We can define the transpose and Hermitian of this vector as

\[
\mathbf{x}^T = [x_1, x_2, \ldots, x_N],
\]

and

\[
\mathbf{x}^H = [x_1^*, x_2^*, \ldots, x_N^*],
\]

respectively. Here the superscript * represents the complex conjugate. Given two vectors, \( \mathbf{x} \) and \( \mathbf{y} \), of length \( N \) their *inner product* can be defined as

\[
\mathbf{x}^H \mathbf{y} = \sum_{n=1}^{N} x_n^* y_n.
\]

The vectors are said to be orthogonal (\( \mathbf{x} \perp \mathbf{y} \)) if \( \mathbf{x}^H \mathbf{y} = 0 \). We define the 2-norm of a vector as

\[
||\mathbf{x}|| = \sqrt{\mathbf{x}^H \mathbf{x}}.
\]

Based on properties of the inner product, we can write the Cauchy-Schwarz inequality

\[
|\mathbf{x}^H \mathbf{y}| \leq ||\mathbf{x}|| ||\mathbf{y}||
\]

where the two sides are equal iff \( \mathbf{x} = \alpha \mathbf{y} \).

A set of vectors, \( \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \) is said to linearly *independent* iff (if and only if)

\[
\sum_{n=1}^{N} \alpha_n \mathbf{x}_n = 0 \Rightarrow \alpha_n = 0.
\]

Given a set of linearly independent vectors \( \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \) one can define the set \( X \) of arbitrary linear combinations of these vectors, \( X = \{\mathbf{x} \mid \mathbf{x} = \sum_{n=1}^{N} \alpha_n \mathbf{x}_n\} \). \( X \) is a *vector space* and the set \( \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \) forms a (non-unique) basis for the vector space. If \( \mathbf{x}_i \perp \mathbf{x}_j, \forall i \neq j \) and \( ||\mathbf{x}_i|| = 1 \), the basis is said to be *orthonormal*. 
2 Matrices

Matrices can, in general, be rectangular such as

\[
A = \begin{bmatrix}
A_{11} & A_{12} & \cdots & A_{1M} \\
A_{21} & A_{22} & \cdots & A_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
A_{N1} & A_{N2} & \cdots & A_{NM}
\end{bmatrix}.
\]

As usual, the transpose and Hermitian can be defined as

\[
A^T = \begin{bmatrix}
A_{11} & A_{21} & \cdots & A_{N1} \\
A_{12} & A_{22} & \cdots & A_{N2} \\
\vdots & \vdots & \ddots & \vdots \\
A_{1M} & A_{2M} & \cdots & A_{NM}
\end{bmatrix},
\]

\[
A^H = \begin{bmatrix}
A_{11}^* & A_{21}^* & \cdots & A_{N1}^* \\
A_{12}^* & A_{22}^* & \cdots & A_{N2}^* \\
\vdots & \vdots & \ddots & \vdots \\
A_{1M}^* & A_{2M}^* & \cdots & A_{NM}^*
\end{bmatrix}.
\]

A square matrix A is said to be symmetric if A = A^T and to be Hermitian if A = A^H. Also, note

\[
[AB]^H = B^H A^H
\]

The rank of a rectangular matrix, A is the number of linearly independent rows (or columns) of A. Note that \(\text{rank}(A) = \text{rank}(A^H A) = \text{rank}(AA^H)\). The matrix A is full rank if \(\text{rank}(A) = \min\{N, M\}\).

If a square matrix has full rank, \(\exists\) (there exists) an inverse matrix, denoted as A^{-1}, such that \(A^{-1}A = AA^{-1} = I\), where I is the \(N \times N\) identity matrix. If the matrix is not full rank, the inverse matrix does not exist and the matrix is said to be singular.

Also, assuming inverses exist,

\[
[AB]^{-1} = B^{-1} A^{-1}.
\]

The determinant of a square \(N \times N\) matrix A is defined (recursively) as

\[
\det A = \sum_{k=1}^{N} (-1)^{i+k} A_{ik} \det A_{ik},
\]

where i is any value between 1 and N and \(A_{ik}\) is the \((N-1) \times (N-1)\) matrix formed by deleting the i-th row and k-th column of the original matrix A.

Some properties of the determinant:
• A matrix $A$ is non-singular (invertible) iff $\det A \neq 0$.

• $\det AB = \det A \det B$.

• $\det A^{-1} = \frac{1}{\det A}$.

• $\det A^T = \det A$.

• $\det aA = a^N \det A$.

The trace of a matrix is defined as the sum of the diagonal elements of the matrix,

$$\text{trace } A = \sum_{n=1}^{N} A_{nn}.$$  \hfill (14)

The column space of a $N \times M$ matrix $A$ is the vector space spanned by its columns (rows):

$$X_c = \{y | y = Ax\}$$

Note, $X_c \subseteq C^N$.

The null space of a matrix is the vector space which results in the zero vector:

$$X_n = \{x | Ax = 0\}$$

Note, $X_n \subseteq C^M$.

For a rectangular matrix $A$, the projection matrix $P_A$ is defined as

$$P_A = A(A^HA)^{-1}A^H,$$ \hfill (15)

with the properties,

$$P_Aa = a,$$ \hfill (16)

if $a$ belongs to the column space of $A$. If $a$ is orthogonal to this column space,

$$P_Aa = 0$$ \hfill (17)

Note that $P^2 = PP = P$, i.e., $P$ is an example of an idempotent matrix.

2.1 Quadratic Forms

An important term we will use in this course is the quadratic form

$$Q(x) = x^H Ax.$$ \hfill (18)
Symmetric matrices (Hermitian for complex matrices) are called *positive semi-definite* iff

\[ Q(x) \geq 0, \quad \forall x \neq 0 \]  \hspace{1cm} (19)

and *positive definite* iff

\[ Q(x) > 0, \quad \forall x \neq 0. \]  \hspace{1cm} (20)

Note that for any (possibly rectangular) matrix \( A \), the matrices \( AA^H \) and \( A^H A \) are positive semi-definite.

### 2.2 Eigenvalues

The *characteristic equation* of a square matrix \( A \) is

\[ Au = \lambda u \]  \hspace{1cm} (21)

Finding the scalars, \( \lambda \) that satisfy this equation is equivalent to solving the \( N \)-th order *characteristic polynomial* in \( \lambda \)

\[ \det [A - \lambda I] = 0 \]  \hspace{1cm} (22)

Since any \( N \)-th order polynomial has \( N \) complex roots, there are \( N \) solutions (*eigenvalues*) to this characteristic polynomial. Some of the roots may be repeated. The vectors \( x \) corresponding to these eigenvalues are *eigenvectors*. Some properties of eigenvalues are:

- if \( u \) is an eigenvector, \( \alpha u \) is also an eigenvector. Therefore one can always choose eigenvector with unit two-norm. We will generally assume eigenvectors have unit norm.

- eigenvectors corresponding to distinct eigenvalues are linearly independent

- if \( \text{rank}(A) = M < N \), there are \( (N - M) \) linearly independent vectors solving

\[ Au = 0, \]  \hspace{1cm} (23)

i.e. there are \( (N - M) \) linearly independent eigenvectors corresponding to the zero eigenvalue. These eigenvectors span the *null space* of \( A \).

- Hermitian matrices have real eigenvalues

- A matrix is positive definite iff all its eigenvalues are positive.

- \( \det A = \prod_{n=1}^{N} \lambda_n \), i.e. a matrix is invertible iff \( \lambda_n \neq 0, \forall n \).

- \( \text{trace}(A) = \sum_{n=1}^{N} \lambda_n \).
• Eigenvectors corresponding to distinct eigenvalues of a Hermitian matrix are orthogonal.

• The eigenvalues of $A^{-1}$ are $1/\lambda_n$, $n = 1, \ldots, N$.

• If matrix $A$ has eigenvalues $\lambda_n$, the matrix $A + \alpha I$ has eigenvalues $\lambda_n + \alpha$ with the same corresponding eigenvectors.

• The condition number of a matrix is the ratio of the largest to the smallest eigenvalue $(\lambda_{\text{max}}/\lambda_{\text{min}})$.

2.2.1 Eigendecomposition

If the eigenvectors of a $N \times N$ matrix $A$ are $u_n, n = 1, \ldots, N$, corresponding to eigenvectors $\lambda_n, n = 1, \ldots, N$, define an eigenvector matrix $U = [u_1, u_2, \ldots, u_N]$. Then

$$AU = UA$$

$$\Rightarrow A = U\Lambda U^{-1},$$

where, $\Lambda$ is the $N \times N$ diagonal matrix, $\text{diag}[\lambda_1, \lambda_2, \ldots, \lambda_N]$.

Since the eigenvectors of a Hermitian matrix $A$ are orthogonal, by making them unit norm, we have

$$U^H U = I$$

This is an example of a unitary matrix, i.e., $U^{-1} = U^H$. The eigendecomposition of this matrix is, therefore,

$$A = U\Lambda U^H,$$  \hspace{1cm} (27)

or

$$A = \sum_{n=1}^{N} \lambda_n u_n u_n^H,$$  \hspace{1cm} (28)

also,

$$A^{-1} = U\Lambda^{-1} U^H,$$  \hspace{1cm} (29)

$$= \sum_{n=1}^{N} \frac{1}{\lambda_n} u_n u_n^H,$$  \hspace{1cm} (30)
2.2.2 Singular Value Decomposition

For any \( N \times M \) rectangular matrix \( A \), one can use the singular value decomposition (SVD),

\[
A = U \Sigma V^H,
\]

\[
= \sum_{n=1}^{N} \lambda_n u_n v_n^H, \quad N < M
\]

\[
= \sum_{m=1}^{M} \lambda_m u_m v_m^H, \quad N > M
\]

The matrix \( U \) is the \((N \times N)\) right singular matrix with the eigenvectors of the matrix \( AA^H \) as its columns. The matrix \( V \) is the \((M \times M)\) left singular matrix with the eigenvectors of the matrix \( A^H A \) as its columns. The columns of \( U \) and \( V \) are chosen such that they have unit norm. Note that since \( A^H A \) and \( AA^H \) are Hermitian matrices, \( U \) and \( V \) are unitary matrices. The matrix \( \Sigma \) is a \( N \times M \) matrix such that \( \Sigma_{nm} = 0 \) if \( n \neq m \). The diagonal entries of \( \Sigma \) are the positive square-roots of the non-zero eigenvalues of \( A^H A \).

3 Differentiation With Respect to Complex Variables

Several times in this class we will have to find the solution to an optimization problem. Often we will have to take a derivative with respect to a vector of complex numbers. A quick review of the theory behind such differentiation.

Consider a scalar function \( f(w) \) of a complex variable \( w = x + jy \). The derivative of this function with respect to \( w \) and its complex conjugate \( w^* \) is defined as [1]

\[
\frac{df}{dw} = \frac{1}{2} \left( \frac{\partial f}{\partial x} - j \frac{\partial f}{\partial y} \right),
\]

\[
\frac{df}{dw^*} = \frac{1}{2} \left( \frac{\partial f}{\partial x} + j \frac{\partial f}{\partial y} \right).
\]

This definition is consistent with the expected \( d(w)/dw = 1 \). However, what is more interesting is that

\[
\frac{dw}{dw^*} = \frac{1}{2} \left( \frac{\partial w}{\partial x} + j \frac{\partial w}{\partial y} \right) = \frac{1}{2} (1 + j(j)) = 0.
\]

The same is true of \( d(w^*)/dw \), i.e. in taking derivatives of functions one can treat \( w \) and \( w^* \) as independent. This greatly simplifies many problems of interest.

Example: \( f(w) = |w|^2 = w w^* \). Then, \( \frac{df}{dw} = w^* \).

Consider now a scalar function \( f \) of a vector \( w = [w_1, w_2, \ldots, w_N]^T, \ w_n = x_n + j y_n \). By
definition, the gradient of this function with respect to \( w \) is the vector
\[
\nabla f = \begin{bmatrix}
\frac{\partial f}{\partial x_1} + j \frac{\partial f}{\partial y_1} \\
\frac{\partial f}{\partial x_2} + j \frac{\partial f}{\partial y_2} \\
\vdots \\
\frac{\partial f}{\partial x_N} + j \frac{\partial f}{\partial y_N}
\end{bmatrix}
= 2 \frac{df}{dw^*}
\] (37)

4 Constrained Optimization

The optimizations in this class will usually be *constrained*, i.e., of the form\(^1\)
\[
w^* = \arg \min_w f(w), \quad \text{such that } g(w) = 0.
\] (38)

Here, \( f(w) \) is the objective function while \( g(w) \) is the constraint. In its original form, this is difficult problem to solve. However, by using Lagrange multipliers one can convert this to an *unconstrained* problem. The Lagrange multiplier form is defined as
\[
\mathcal{L}(w, \lambda) = f(w) + \lambda g(w).
\] (39)

One now needs to find the vector \( w \) and \( \lambda \) that minimizes this function. Since this problem is unconstrained, one can “differentiate and set to zero”. Note that differentiating w.r.t. \( \lambda \) makes \( g(w) = 0 \) one equation we have to account for (the constraint).

For example, consider the optimization problem
\[
w^* = \arg \max_w [w^H R w] \quad w^H s = 1.
\] (40)

Using Lagrange multipliers,
\[
\mathcal{L}(w, \lambda) = w^H R w + \lambda (w^H s - 1),
\] (41)
\[
\Rightarrow \nabla_w \mathcal{L}(w, \lambda) = R w + \lambda s = 0,
\] (42)
\[
\Rightarrow w = -\lambda R^{-1} s.
\] (43)

Also,
\[
\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \Rightarrow w^H s = 1.
\] (44)

Using Eqns. (43) and (44)
\[
w = \frac{R^{-1} s}{s^H R^{-1} s}.
\] (45)

References


\(^1\)This discussion presents an extremely simplified discussion of constrained optimization. I recommend reading [2] for more details.