Part I

Basic Elements of Linear Algebra

Although the subject of linear algebra is much broader than just vectors and matrices, most people’s first experience and day-to-day interaction with the subject is through these objects. Furthermore, and perhaps, a stronger argument as to why matrices and vectors are a fundamental place to start such a review is that much of modern computation is done with matrices and vectors, either as a direct representation of a model problem or through some type of discretization. Thus, a strong background and familiarity with these objects is necessary to successfully solve many problems.

1 Scalars

The most basic element in linear algebra is the scalar; scalars are just a number, and are typically denoted with Greek letters (α, β, γ, ε, δ, etc.). For the purpose of this review, and of most applications, scalars are taken from the field of real numbers, denoted, \( \mathbb{R} \). On occasion, people may use complex numbers as the field, \( \mathbb{C} \), or indeed any other field.

2 Vectors

Vectors are a finite, ordered set of numbers (of the same field as the scalars we are considering). These \( n \) numbers are indexed, and may be denoted, \( x = \{v_i\}_{i=1}^n \); vectors are typically denoted with lower case Roman letters (a,b,x,y,z, etc.) and are typically referred to as elements in \( \mathbb{R}^n \) (e.g. \( x \in \mathbb{R}^n \) or \( y \in \mathbb{R}^m \)).

There are two orientations of vectors, the more common, column vector,

\[
x = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}
\]

and the row vector

\[
x = [v_1 \ v_2 \ \cdots \ v_n]
\]

both are elements of \( \mathbb{R}^n \). Unless otherwise stated, all vectors as assumed to be column vectors. The transpose of a vector \( x \), denoted \( x^T \), changes a vector from a column vector to a row vector (or vice versa).

There are many examples of vectors. One common example is as a point in \( \mathbb{R}^2 \), where \( v_1 \) is the \( x \) coordinate and \( v_2 \) is the \( y \) coordinate. Vectors can also denote a direction and magnitude (see below), such as a force. Another common example of a vector is as a set of probabilities; here each component, say \( \pi \), gives
the probability of the \( i \)th event. Typically the components of these vectors sum to 1 (and are non-negative); in such cases these vectors are called stochastic.

2.1 Arithmetic Operations with Vectors

2.1.1 The Scalar Product

Let \( \alpha \in \mathbb{R} \) and \( x = \{ \nu_i : 1 \leq i \leq n \} \in \mathbb{R}^n \), then the scalar product is

\[
\alpha x = \{ \alpha \nu_i : 1 \leq i \leq n \}
\]

and is also in \( \mathbb{R}^n \).

2.1.2 Vector Addition

Let \( a = \{ \alpha_i : 1 \leq i \leq n \} \) and \( b = \{ \beta_i : 1 \leq i \leq n \} \) be vectors in \( \mathbb{R}^n \). The vector sum is defined as

\[
a + b = \{ \alpha_i + \beta_i : 1 \leq i \leq n \} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} = \begin{bmatrix} \alpha_1 + \beta_1 \\ \alpha_2 + \beta_2 \\ \vdots \\ \alpha_n + \beta_n \end{bmatrix}.
\]

The vector \( a + b \) is in \( \mathbb{R}^n \) as well. Thus \( \mathbb{R}^n \) is closed under addition and scalar multiplication.

2.1.3 The Dot Product

Let \( a \) and \( b \) be two column vectors in \( \mathbb{R}^n \). The dot product, denoted

\[
a \cdot b = a^T b = \sum_{i=1}^{n} \alpha_i \beta_i
\]

which is a scalar.

2.2 The Vector Norm

A norm is a mapping \( f : \mathbb{R}^n \rightarrow \mathbb{R}^+ \) (here \( \mathbb{R}^+ \) is the set of non-negative real numbers) with the following properties:

1. \( f(x) = 0 \) if and only if \( x = 0 \).
2. \( f(\alpha x) = |\alpha| f(x) \) for \( \alpha \in \mathbb{R} \) and \( x \in \mathbb{R}^n \).
3. \( f(x + y) \leq f(x) + f(y) \) for \( x, y \in \mathbb{R} \).

A norm \( f(x) \) is often denoted by \( \|x\| \). Let \( x = \{ \nu_i : 0 \leq i \leq n \} \). There are three commonly used vector norms in linear algebra:

1. The 2-Norm:

\[
f(x) = \|x\|_2 = \sqrt{\sum_{i=1}^{n} \nu_i^2}.
\]

This is probably the most commonly used norm.
2. The 1-Norm:

\[ f(x) = \|x\|_1 = \sum_{i=1}^{n} |\nu_i| . \]

This is sometimes referred to as the “Manhattan Norm” as it can be viewed as counting the number of blocks you would have to walk on a \( n \)-dimensional grid.

3. The \( \infty \)-Norm:

\[ f(x) = \|x\|_\infty = \max_{1 \leq i \leq n} |\nu_i| . \]

This gives the component with largest magnitude.

In all cases, the norm is a measure of length; the 2-Norm corresponds to the Euclidean length of the vector.

2.3 Inequalities

Let \( a = \{\alpha_i : 1 \leq i \leq n\} \) and \( b = \{\beta_i : 1 \leq i \leq n\} \) be elements in \( \mathbb{R}^n \). We say

\[ a \leq b \] if and only if \( \alpha_i \leq \beta_i \) for all \( 1 \leq i \leq n \).

The Cauchy-Schwartz Inequality states that

\[ |x^T y| \leq \|x\|_2 \|y\|_2 . \]

2.4 Special Vectors

There are two “special” notations used commonly in linear algebra. The first is the vector \( e = \{\epsilon_i = 1 : 1 \leq i \leq n\} \), that is the vector of all ones:

\[ e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} . \]

The second is the vector \( e_i \) which is vector of all zeros, but with a 1 in the \( i \)th position. \( e_i = \{\epsilon_j : 1 \leq j \leq n\} \), where

\[ e_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} . \]

3 Matrices

Matrices are two dimensional objects and are a collection of \( mn \) numbers; they are typically denoted by capitol letters (\( A, B, Q, M \), etc.) and are referred to as elements of \( \mathbb{R}^{m \times n} \). They are often specified in the following way: \( A = \{a_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\} \):

\[
A = \begin{bmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix} .
\]

\( m \) is the number of rows; \( n \) is the number of columns. Alternatively, matrices may also be thought of as a collection of \( n \) column vectors \( \{a_1, a_2, \ldots, a_n\} \) where each \( a_i \in \mathbb{R}^m \):

\[
A = \begin{bmatrix}
    | & | & \cdots & | \\
    a_1 & a_2 & \cdots & a_n
\end{bmatrix} ;
\]
or a collection of rows vectors \( \{b_1, b_2, \cdots, b_m\} \) where each \( b_i \in \mathbb{R}^n \):

\[
A = \begin{bmatrix}
-b_1 \\
-b_2 \\
\vdots \\
-b_m
\end{bmatrix}.
\]

Examples of matrices can be found throughout engineering. As we will see shortly, matrices are a concise way to express linear systems or linear constraints. Finite difference approximations can often be conveniently represented in matrices as well as transition probabilities for many classes of discrete stochastic processes. A more concrete example of a matrix is the \( 2 \times 2 \) rotation matrix; suppose we wish to rotate a point in \( \mathbb{R}^2 \) \( \theta \) degrees about the origin. The matrix

\[
Q = \begin{bmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{bmatrix}.
\]

when applied to a vector in \( \mathbb{R}^2 \) will rotate the vector by \( \theta \) degrees. Finally, vectors themselves are special cases of matrices: column vectors, in \( \mathbb{R}^n \), are matrices in \( \mathbb{R}^{n \times 1} \); row vectors are matrices in \( \mathbb{R}^{1 \times n} \).

### 3.1 The Matrix Transpose

Although the transpose of a matrix has much deeper mathematical meaning that what we will get into here, it can be convenient to introduce it here, as a symbolic manipulation. Let \( A = \{\alpha_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\} \). Then transpose of the matrix, denoted \( A^T = B = \{\beta_{ij} : 1 \leq i \leq n, 1 \leq j \leq m\} \) where the elements of \( B \) are given by

\[
\beta_{ij} = \alpha_{ji}.
\]

Thus if \( A \in \mathbb{R}^{m \times n} \) then \( B \in \mathbb{R}^{n \times m} \). The transpose has the following properties:

- \((A^T)^T = A\)
- \((A + B)^T = A^T + B^T\)
- \((AB)^T = B^T A^T\)
- \((\alpha A)^T = \alpha A^T\).

### 3.2 Arithmetic Operations with Matrices

Let \( \gamma \in \mathbb{R} \) and \( A = \{\alpha_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\} \) and \( B = \{\beta_{ij} : 1 \leq i \leq n, 1 \leq j \leq n\} \) be elements of \( \mathbb{R}^{m \times n} \).

#### 3.2.1 Scalar Multiplication

Multiplying a matrix \( A \) by the scalar \( \gamma \) scales the elements of \( A \) by \( \gamma \):

\[
\gamma A = \{\gamma \alpha_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}.
\]

#### 3.2.2 Matrix Addition

For two matrices in the same space \( \mathbb{R}^{m \times n} \), like \( A \) and \( B \) defined above, matrix addition is defined by \( A + B = \{\alpha_{ij} + \beta_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\} \), or

\[
A + B = \begin{bmatrix}
\alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\
\alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{m1} & \alpha_{m2} & \cdots & \alpha_{mn}
\end{bmatrix} + \begin{bmatrix}
\beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\
\beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\beta_{m1} & \beta_{m2} & \cdots & \beta_{mn}
\end{bmatrix} = \begin{bmatrix}
\alpha_{11} + \beta_{11} & \alpha_{12} + \beta_{12} & \cdots & \alpha_{1n} + \beta_{1n} \\
\alpha_{21} + \beta_{21} & \alpha_{22} + \beta_{22} & \cdots & \alpha_{2n} + \beta_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\alpha_{m1} + \beta_{m1} & \alpha_{m2} + \beta_{m2} & \cdots & \alpha_{mn} + \beta_{mn}
\end{bmatrix}.
\]
3.2.3 Matrix-Matrix Products

Let $A = \{\alpha_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$ be in $\mathbb{R}^{m \times n}$ and $B = \{\beta_{ij} : 1 \leq i \leq p, 1 \leq j \leq m\}$ be in $\mathbb{R}^{p \times m}$. The matrix product between an matrix in $\mathbb{R}^{p \times m}$ and a matrix in $\mathbb{R}^{m \times n}$, denoted $B \times A = C = \{\gamma_{ij} : 1 \leq i \leq p, 1 \leq j \leq n\}$, has elements given by:

$$
\gamma_{ij} = \sum_{k=1}^{m} \beta_{ik} \alpha_{kj}, \quad 1 \leq i \leq p, 1 \leq j \leq n,
$$
or $\gamma_{ij}$ is the dot product of the $i$th row of $B$ with the $j$th column of $A$. $C$ is an element of $\mathbb{R}^{p \times n}$. The number of columns of $B$ is called its *inner dimension* and the number of rows of $A$ is called its *inner dimension*; the matrix product is only defined when the inner dimension is the same.

3.2.4 The Outer Product

Let $x \in \mathbb{R}^{n}$ and $y \in \mathbb{R}^{m}$ be column vectors. Their outer product, $yx^T = A = \{\alpha_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$, is an element of $\mathbb{R}^{m \times n}$ with elements given by

$$
\alpha_{ij} = y_i x_j.
$$
It can be seen that the outer product is a special case of the matrix-matrix product.

3.2.5 Matrix-Vector Products

The matrix-vector product too can be viewed as a special case of the matrix-matrix product, however it ubiquitous presence in linear algebra makes it worthy of special mention. Let $A = \{\alpha_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\} = (a_1 a_2 \cdots a_n) (b_1 b_2 \cdots b_m)^T$ be an element in $\mathbb{R}^{m \times n}$ and $a_i \in \mathbb{R}^m$ and $b_i \in \mathbb{R}^n$ be column vectors; let $x \in \mathbb{R}^n$ be a column vector. The vectors $\{a_i\}$ are the columns of $A$ and $\{b_i^T\}$ are the rows of $A$. The product $Ax = c = \{\gamma_i : 1 \leq i \leq m\}$ is a column vector in $\mathbb{R}^m$. So

$$
Ax = c = \begin{bmatrix}
\sum_{j=1}^{n} \alpha_{1j} \nu_j \\
\sum_{j=1}^{n} \alpha_{2j} \nu_j \\
\vdots \\
\sum_{j=1}^{n} \alpha_{mj} \nu_j
\end{bmatrix} = \begin{bmatrix}
\nu_1 a_1 + \nu_2 a_2 + \cdots + \nu_n a_n
\end{bmatrix} = \begin{bmatrix}
b_1^T x \\
b_2^T x \\
\vdots \\
b_m^T x
\end{bmatrix}.
$$

Thus we see the two different interpretations of the matrix-vector product. The first says that the components of $c$, the $\gamma_i$s, are specific combinations of the components of $x$, where the weights are given by the elements in $i$th row of $A$. The second, is that the product $Ax$ is a weighted combination of the columns of $A$, where weights are given by the components of $x$.

3.3 Matrix Norms

Let $A$ and $B$ be elements of $\mathbb{R}^{m \times n}$ and $\gamma \in \mathbb{R}$. A norm is a function $f : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^+$ with the following properties:

1. $f(A) = 0$ if and only if $A = 0$.
2. $f(\gamma A) = |\gamma| f(A)$
3. $f(A + B) \leq f(A) + f(B)$.

Further more, when $n = m$, if the norm has the property that $f(AB) \leq f(A)f(B)$, we call the norm sub-multiplicative. All the norms we will consider are sub-multiplicative. Like vector norms, matrix norms or often denoted by $f(A) = \|A\|$.
3.3.1 Induced Norms

A vector norm induces a matrix norm in the following way:

\[ \|A\| = \sup_{\|x\|=1} \|Ax\|^m \]

(where the superscripts on the norms specifies the space the norm is defined on; either \(\mathbb{R}^n\) or \(\mathbb{R}^m\)). For the three vectors norms discussed above, the matrix norm has an explicit form:

1. The 2-Norm:

\[ \|A\|_2 = \sigma_1(A) \]

where \(\sigma_1(A)\) is the largest singular value of \(A\). (Singular values will be discussed later).

2. The 1-Norm:

\[ \|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}| \]

or the maximal column-sum.

3. The \(\infty\)-Norm:

\[ \|A\|_{\infty} = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}| \]

or the maximal row-sum.

3.3.2 Element-wise Norms

Several norms are defined directly in terms of the elements of the matrix \(A\). The most used of these norms is the Frobenius Norm and is defined as follows

\[ \|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} \]

The trace of a square matrix is defined as follows:

\[ \text{tr}(A) = \sum_{i=1}^{n} a_{ii} \]

Thus the Frobenius Norm can be written as

\[ \|A\|_F = \text{tr}(A^T A) \]

(note \(A^T A\) is square).

3.4 Structured Matrices

There are several types of matrices with special structure that recur in linear algebra. The first, which we have already seen, is a square matrix, a matrix with the same number of rows and columns so that \(m = n\).

A diagonal matrix is a square matrix where the only non-zero entries are the \(a_{ii}\)s. More specifically, \(D = \{\delta_{ij} : 1 \leq i, j \leq n\}\) is diagonal if

\[ \delta_{ij} = 0 \quad i \neq j, \quad 1 \leq i, j \leq n. \]
The *identity* matrix is a diagonal matrix with $\delta_{ii} = 1$; it is often denoted by $I$ and has the property that $Ix = x$.

A *symmetric* matrix $A = \{a_{ij} : 1 \leq i, j \leq n\}$ is a square matrix with the property that $a_{ij} = a_{ji}$. Diagonal matrices are symmetric.

**Part II**

**Sets of Vectors**

A set $S \subset \mathbb{R}^n$ is a *subspace* if the following holds if for all $x, y \in S$ and $\alpha \in \mathbb{R}$:

$$x + \alpha y \in S.$$  

A *linear combination* of $p$ vectors $\{s_1, s_2, \ldots, s_p\}$ is specified by any $p$ real numbers $\{\alpha_1, \alpha_2, \ldots, \alpha_p\}$ and is given by

$$\alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_p s_p.$$  

If the vectors $\{s_i\}$ are made to be the columns of a matrix $S = (s_1 \ s_2 \ \cdots \ s_p)$, and the vector $a$ has components $\alpha_i$, then, as we have seen, the linear combination can be expressed as the matrix-vector product $Sa$. Let $S = \{s_1, s_2, \ldots, s_p\}$ be a set of $p$ vectors in $\mathbb{R}^n$. The *span* of $S$, denoted $\text{span}(S)$ is

$$\text{span}(S) = \{y = Sx : x \in \mathbb{R}^p\}$$  

is the set of all linear combinations of the vectors in $S$.

A set of vectors $S = \{s_1, s_2, \ldots, s_p\}$ is *linearly independent* if the only linear combination of $S$ that equals the zero vector is the trivial zero vector. In other words, the only way that

$$\alpha_1 s_1 + \alpha_2 s_2 + \cdots + \alpha_p s_p = 0$$

is if all of the $\alpha_i$’s are zero. For examples the vectors:

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

are linearly independent and $\text{span}(\{e_1, e_2\}) = \mathbb{R}^2$.

The *dimension*, denoted $\text{dim}(S)$, of a subspace is the largest number of vectors in the subspace that are linearly independent. A *basis* is any set of linearly independent vectors that spans the subset. The number of vectors in a basis for $S$ is $\text{dim}(S)$. The basis is not a unique set of vectors.

Some facts about subspaces:

- If $S \subset \mathbb{R}^m$ is a subspace, then its dimension is at most $m$.
- Every vector in a subspace can be written as a unique linear combination of the basis vectors.
Part III
Linear Functions, Rank, Null Space, Linear Systems and Inverses

4 Linear Functions
A function \( f : A \rightarrow B \) is said to be linear if
\[
f(x + \alpha y) = f(x) + \alpha f(y)
\]
for \( x, y \in A \) and \( \alpha \in \mathbb{R} \).

The function \( f : \mathbb{R}^n \rightarrow \mathbb{R}^m \) defined by \( f(x) = Ax \) where \( A \in \mathbb{R}^{m \times n} \) is a linear function.

5 Range Space, Null Space and Rank
Recall that for a function \( f : \mathbb{R}^n \rightarrow \mathbb{R}^m \) to be one-to-one, each \( x \in \mathbb{R}^n \) must map to a unique \( y \in \mathbb{R}^m \). Thus for \( Ax \) to be one-to-one, it is necessary that \( n \leq m \) and that the columns of \( A \) be linearly independent. If \( A \) does not have linearly independent columns, then there is vector \( z \in \mathbb{R}^n \) such that \( Az = 0 \), and thus due to the linearity, there are an infinite number of vectors that map to zero. This set of vectors is called the null space of \( A \) and is denoted
\[
N(A) = \{ x \in \mathbb{R}^n : Ax = 0 \}.
\]
Note that due to the linearity of the mapping \( x \mapsto Ax \), \( N(A) \) is a subspace.

Now consider the range of the function \( Ax \). The range
\[
\mathcal{R}(A) = \{ Ax : x \in \mathbb{R}^n \}.
\]
is the set of vectors mapped to in \( \mathbb{R}^m \) by \( x \mapsto Ax \). For an arbitrary \( x \in \mathbb{R}^n \), \( y = Ax \) is a linear combination of the columns of \( A \). Thus we see that the range of \( Ax \) is the span of the columns of \( A \):
\[
\mathcal{R}(A) = \text{span}\{a_1, a_2, \ldots, a_n\}.
\]
The column-rank is the dimension of \( \mathcal{R}(A) \) and the row-rank is the size of \( \mathcal{R}(A^T) \). Observe that the column-rank of a matrix is equal to its row-rank and is called the rank of the matrix. A matrix is said to have full rank if \( \text{rank}(A) = \min(m, n) \). The function defined by \( x \mapsto Ax \) is one-to-one if \( m \geq n \) and \( A \) has full rank. This yields the useful property that if \( A \) is full rank and \( m \geq n \)
\[
Ax = Ay \Rightarrow x = y
\]
(which allows these terms to be canceled from algebraic equations).

For the mapping \( x \mapsto Ax \) to be onto, the column-rank must be \( m \) (however, the mapping may not be one-to-one if \( n \geq m \)). Thus a square matrix is full rank if and only if it is both one-to-one and onto. Such a matrix is called non-singular. A square matrix with rank less than \( n \) is called singular. Furthermore, since a non-singular matrix is a bijection (both onto and one-to-one) a unique inverse exists (this will be useful later).

5.1 The Fundamental Theorem of Linear Algebra
The fundamental theorem of linear algebra states that for a matrix \( A \in \mathbb{R}^{m \times n} \),
\[
\dim(\mathcal{R}(A)) + \dim(N(A)) = n.
\]
6 Linear Systems and the Matrix Inverse

Linear systems arise from problems of the following form: given a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$, what vector $x \in \mathbb{R}^n$ solves:

$$Ax = b$$

This is a central problem in linear algebra. One example of where this problem arises is the following: suppose we have $m$ foods, and $n$ nutrients. Food $i$ provides $\alpha_{ij}$ amount of nutrient $j$ per kilogram. Our goal is to determine what allocation of food matches a given nutrient prescription (the elements of $b$, $\beta_i$, correspond to the amount of nutrient $i$ needed by a person). This can be posed as finding the solution to a linear system of the above form.

The solution to the above problem need not exist, or may not be unique. However, when $A$ is square and $n = m$, and is of full rank, there is a unique solution to the problem. A solution does not exist if $b / \in \mathbb{R}(A)$ (then there is no linear combination of the columns of $A$ equal to $b$). The solution may be non-unique when the system is not full rank (suppose $x$ is a solution and $z$ is a vector in the null space of $A$, then $(x + \delta z)$ is a solution to the problem as well).

6.1 The Matrix Inverse

As stated above, when the system is square ($m = n$) and full rank, the mapping $x \mapsto Ax$ is a bijection and thus an inverse is uniquely defined, denoted $A^{-1}$, and the solution is given by $x = A^{-1}b$. This inverse satisfies the following property: $A^{-1}A = AA^{-1} = I$. The inverse is only defined for square, full-rank matrices. Suppose $A, B \in \mathbb{R}^{n \times n}$, if the inverse of $A$ and $B$, $A^{-1}$ and $B^{-1}$ respectively, exist, then the inverse of $AB$ is given by

$$(AB)^{-1} = B^{-1}A^{-1}.$$ 

As an example, consider the matrix

$$A = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}.$$ 

Then, provided $\alpha \delta - \beta \gamma \neq 0$ (and thus the matrix has full rank) the inverse is given by

$$A^{-1} = \frac{1}{\alpha \delta - \beta \gamma} \begin{bmatrix} \delta & -\beta \\ -\gamma & \alpha \end{bmatrix}.$$ 

6.2 A Simple Example

Consider

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \text{and} \quad c = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$ 

(Note that the formula for the matrix inverse doesn’t work here.) There are an infinite number of solutions to the problem

$$Ax = b$$

and any solution takes the form of

$$x = \begin{bmatrix} 2 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$ 

(Note that $(-1, 1)^T$ is in the null space of $A$) Secondly, the problem

$$Ax = c$$

has no solution as $(0, 1)^T \notin \mathbb{R}(A)$. 

9
Part IV
Inner Products, Orthogonality and Least Squares

7 The Inner Product

Let \( x, y \in \mathbb{R}^n \) and \( A \in \mathbb{R}^{n \times n} \) where \( A = A^T \) and \( z^T A z \geq 0 \) for all \( z \neq 0 \). Then, the inner product, with respect to \( A \), of \( x \) and \( y \) is given by

\[
\langle x, y \rangle_A = x^T A y = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j a_{ij}.
\]

The inner product has the following properties:

- \( \langle x, x \rangle_A \geq 0 \)
- \( \langle x, x \rangle_A = 0 \) if and only if \( x = 0 \).
- \( \langle x, y + w \rangle_A = \langle x, y \rangle_A + \langle x, w \rangle_A \)
- \( \langle x, \alpha y \rangle_A = \alpha \langle x, y \rangle_A \)
- \( \langle x, y \rangle_A = \langle y, x \rangle_A \).

The dot product can be defined as the inner product with respect to the identity matrix.

The inner product can be used to define a norm:

\[
\| x \|_A = \sqrt{\langle x, x \rangle_A}
\]

and is called the \( A \)-norm. When \( A = I \), this is just the 2-norm. When \( A = D \), a diagonal matrix, the components of \( D \) weight the contribution of the components of \( x \) to the norm. Let \( D = \{ \delta_{ij} \} \) where \( \delta_{ij} = 0 \) when \( i \neq j \) and \( \delta_{ii} > 0 \), then the weighted norm is given by:

\[
\| x \|_A = \sqrt{\sum_{i=1}^{n} \delta_{ii} x_i^2}.
\]

This is often used in statistics to decrease the effect of noisy or uncertain data points when estimating parameters.

The angle \( \theta \in [0, \pi] \), between two vectors \( x \) and \( y \) can be defined by:

\[
\cos \theta = \frac{x^T y}{\| x \|_2 \| y \|_2}.
\]

This can be thought of as telling us how much of vector \( x \) lies in the direction \( y \). It projects \( x \) onto \( y \).
8 Orthogonality

Two vectors $x$ and $y$ are said to be orthogonal (or $A$-orthogonal) if $x^T y = 0$ (or $\langle x, y \rangle_A = 0$). As set of vectors $\{u_1, \ldots, u_m\}$ is called an orthonormal set if $\|u_i\| = 1$, for all $i$, and

$$ u_i^T u_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}. $$

Every orthonormal set is a basis for the space it spans (and is linearly independent). An example of orthonormal vectors are the columns of the identity matrix. A second example in $\mathbb{R}^2$ are the following,

$$ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}. $$

A matrix $Q$ is said to be orthogonal if its columns form an orthonormal basis. When $Q \in \mathbb{R}^{n \times n}$ the following property holds:

$$ Q^T Q = I = QQ^T $$

and thus $Q^T$ is the inverse of $Q$.

Orthogonal matrices are often referred to a reflections or rotations. This is because they do not change the Euclidean length of a vector. Specifically

$$ \|Qx\|_2 = \|x\|_2. $$

8.1 The Gram-Schmidt Orthogonalization Process

Given a set of linearly independent vectors $\{y_1, y_2, \ldots, y_n\}$ the vectors can be used to construct an orthonormal basis for $\text{span}(\{y_1, \ldots, y_n\})$. The process is as follows: The first vector, $q_1$, is obtained by normalizing the first vector in the linearly independent set:

$$ q_1 = \frac{y_1}{\|y_1\|_2}. $$

To form the $k$th orthogonal vector, from the $k$th basis vector, project $y_k$ onto each of the existing orthogonal vectors $(\{q_1, \ldots, q_{k-1}\})$, giving

$$ \alpha_{ki} = \langle y_k, q_i \rangle, \quad 1 \leq i \leq k - 1 $$

and remove the component of $q_i$ from $y_k$ by subtracting $\alpha_{ki} q_i$ form $y_k$:

$$ \hat{q}_k = y_k - \sum_{i=1}^{k-1} \alpha_{ki} q_i = y_k - \sum_{i=1}^{k-1} \langle y_k, q_i \rangle q_i. $$

The $k$th orthonormal vector is then computed by normalizing $\hat{q}_k$:

$$ q_k = \frac{\hat{q}_k}{\|\hat{q}_k\|_k}. $$

8.2 The QR Matrix Factorization

The Gram Schmidt Orthogonalization process is computationally unstable (after computing a few vectors, the set $\{q_1, \ldots, q_n\}$ stops being orthogonal), an alternative is the QR factorization. Let $A \in \mathbb{R}^{m \times n}$, where $m \geq n$. Then there exists a factorization of $A = QR$ such that $Q$ is an orthogonal matrix, and $R$ is upper triangular. (A matrix $R = \{\rho_{ij}\}$ is upper triangular if $\rho_{ij} = 0$ when $i > j$; in other words, all the elements
below the diagonal are zero. The matrix $Q$ is $m \times m$ and the first $r$ columns of $Q$ form a basis for $\mathcal{R}(A)$ (where $r$ is the rank of $A$). Thus, to obtain an orthonormal basis for $\{y_1, \ldots, y_n\}$, we can define the matrix

$$A = \begin{bmatrix} | & | & | \\ y_1 & y_2 & \cdots & y_n \end{bmatrix}$$

and compute the QR factorization of $A$. Then the columns of $Q$ are an orthonormal basis for $\{y_1, \ldots, y_n\}$. The QR algorithm is implemented in MATLAB as `qr`.

To further motivate why a QR factorization is useful, consider the problem of solving the linear system $Ax = b$ where $b \in \mathcal{R}(A)$ and $A$ is a full rank $n \times n$ matrix. Substituting $QR$ for $A$ gives,

$$Ax = b \Rightarrow QRx = b \Rightarrow Rx = Q^Tb = \hat{b},$$

where the last equality is due to the fact that $Q^{-1} = Q^T$. Now we need only solve the system $Rx = \hat{b}$ which is easy: Consider the last row of the equation, which reads:

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & r_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = r_{nn}x_n = \hat{b}_n \Rightarrow x_n = \frac{\hat{b}_n}{r_{nn}}.$$

Now consider the second to last row, $n - 1$:

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & r_{n-1n-1} & r_{n-1n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = r_{n-1n-1}x_{n-1} + r_{n-1n}x_n = \hat{b}_{n-1} \Rightarrow x_{n-1} = \frac{\hat{b}_n - r_{n-1n}x_n}{r_{n-1n-1}}.$$

However, we know the value $x_n$ from the previous row. This process can be continued up the matrix, solving row $i$ only after row $i + 1$; done in this way, there is only one unknown being solved for, and we can easily compute the solution vector $x$.

### 8.3 The Orthogonal Complement

Let $S$ be a subset of $\mathbb{R}^n$. Define $S^\perp$, the orthogonal complement of $S$, as

$$S^\perp = \{ x \in \mathbb{R}^n : y^T x = 0 \forall y \in S \}.$$ 

$S^\perp$ is a subspace of $\mathbb{R}^n$. (Proof: Suppose $x, y \in S^\perp$, then take $z \in S$, $z^T(x + \alpha y) = z^T x + \alpha z^Ty = 0$ since $x$ and $y$ are, by the definition of $S^\perp$, orthogonal to any $z \in S$.)

Every vector in $z \in \mathbb{R}^n$ can be uniquely decomposed into $z = x + y$ with $x \in S$ and $y \in S^\perp$. Thus the inner product of a vector decomposes as follows:

$$z^Tz = (x + y)^T(x + y) = x^T x + 2x^T y + y^T y = x^T x + y^T y$$

as $x$ and $y$ are orthogonal.
9 Least Squares Problems

Recall that for an arbitrary matrix $A \in \mathbb{R}^{m \times n}$ a solution to the problem

$$Ax = b$$

does not necessarily exist; $b$ may not be in $\mathcal{R}(A)$. Now suppose we wish to find the $x$ that makes $Ax$ as close to $b$ as possible. In the example given at the end of the section on linear systems, we would be looking for an allocation of foods that comes as close as possible to matching the nutrient requirements.

To formalize this, define $r$, the residual vector, as follows:

$$r = Ax - b$$

and we seek to minimize $\|r\|_2$. Note that the minimizer of $\|r\|_2$ is the same as the minimizer of $\|r\|_2^2$, which is often a more convenient form to work with. We now decompose $b$ into the part that lies in $\mathcal{R}(A)$, say, $b_1$, and the part that is in $\mathcal{R}(A)^\perp$, say $b_2$. Plugging this in we see that

$$\|r\|_2^2 = \|b_1 - Ax + b_2\|_2^2 + \|b_1 - Ax\|_2^2 + \|b_2\|_2^2.$$

Thus $\|r\|_2^2$ is minimized when $Ax = b_1$ and the minimal value is $\|b_2\|_2^2$. So the solution is minimal when $r = b - Ax \in \mathcal{N}(A^T)$.

To see how we might actually solve such a problem, $r \in \mathcal{N}(A^T)$ means that $A^Tr = 0$ or

$$A^T(Ax - b) = 0 \Rightarrow A^TAx = A^Tb.$$

These are called the normal equations. When $A$ has full column rank, $A^TA$ has full rank and we can solve the above linear system. When $A$ doesn’t have full column rank, there are an infinite number of solutions to the above problem. As with many of the techniques discussed in these notes, the solution of the normal equations is numerically unstable. Using the ‘backslash’ command in MATLAB, will compute the least square solution to $\min \|Ax - b\|$ in a stable way.

Part V
Eigenvectors, Eigenvalues and the SVD

10 Eigenvectors and Eigenvalues

Let $A$ be an $n \times n$ matrix. A vector $x$ and a complex number $\lambda$ are called an eigenvector / eigenvalue pair if $x \neq 0$ and

$$Ax = \lambda x.$$

For example, the matrix

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

has 3 as an eigenvalue and the associated eigenvector is $[1 \ 1]^T$:

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

A given eigenvalue may have more than one eigenvector associated with it.
From the definition, we see that if $\lambda$ is an eigenvalue of $A$ with associated eigenvector $x$ that
\[ Ax = \lambda x \Rightarrow (A - \lambda I)x = 0 \Rightarrow x \in \mathcal{N}(A - \lambda I). \]
The space $\mathcal{N}(A - \lambda I)$ is called the eigenspace associated with eigenvalue $\lambda$. It is a subspace, and is denoted $E_\lambda$. The \textit{geometric multiplicity} of an eigenvalue $\lambda$ is the size of the associated null space, $\dim(\mathcal{N}(A - \lambda I))$. This subspace is closed under multiplication with $A$, so that if $x \in E_\lambda$, then $Ax \in E_\lambda$. This can be generalized to say that if $x \in \cap_{\lambda \in \Lambda} E_\lambda$ then so is the product $Ax$ (where $\Lambda$ is a collection of eigenvalues).

\section{10.1 The Determinant}

Let $A$ be an $n \times n$ matrix. Define the $(n - 1) \times (n - 1)$ matrix $\tilde{A}_{ij}$ as the matrix $A$ but with row $i$ and column $j$ removed. The \textit{determinant} of a matrix $A$, $\det(A)$, is computed by
\[
\det \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \alpha \delta - \beta \gamma,
\]
for $n = 2$ and when $n > 2$, by
\[
\det(A) = \sum_{j=1}^{n} (-1)^{i+j} A_{ij} \det(\tilde{A}_{ij})
\]
for any $i$. The determinant of an upper triangular matrix is the product of its diagonal elements (and, as a special case, so is the determinant of a diagonal matrix). The determinant has the following properties:

- If two columns (or rows) are the same, then $\det(A) = 0$.
- $\det(AB) = \det(A) \det(B)$
- $\det(A^T) = \det(A)$
- $\det(Q) = 1$ when $Q$ is an orthogonal matrix.
- $\det(A^{-1}) = \det(A)^{-1}$, when $A^{-1}$ exists.
- $\det(A) = 0$ if and only if $A$ is singular.

Thus if $\lambda$ is an eigenvalue of $A$, $\det(A - \lambda I) = 0$.

As an example, take the matrix:
\[
A = \begin{bmatrix}
1 & 0 & -2 \\
2 & -1 & 3 \\
1 & 2 & 4 \\
\end{bmatrix}.
\]
We will take $i = 1$. Then
\[
\tilde{A}_{11} = \begin{bmatrix}
-1 & 3 \\
2 & 4 \\
\end{bmatrix} \quad \tilde{A}_{12} = \begin{bmatrix}
2 & 3 \\
1 & 4 \\
\end{bmatrix} \quad \text{and} \quad \tilde{A}_{13} = \begin{bmatrix}
2 & -1 \\
1 & 2 \\
\end{bmatrix}.
\]
and
\[
\det(\tilde{A}_{11}) = (-1) \times 4 - 2 \times 4 = -10 \quad \det(\tilde{A}_{12}) = 2 \times 4 - 3 \times 3 = 5 \quad \text{and} \quad \det(\tilde{A}_{13}) = 2 \times 2 - (-1) \times 1 = 5.
\]
So
\[
\det(A) = \sum_{i=1}^{3} (-1)^{i+1} A_{ij} \det(\tilde{A}_{ij}) = (-1)^{1+1} A_{11} \det(\tilde{A}_{11}) + (-1)^{1+2} A_{12} \det(\tilde{A}_{12}) + (-1)^{1+3} A_{13} \det(\tilde{A}_{13})
\]
\[= 1 \times 1 \times (-10) + (-1) \times 0 \times 5 + 1 \times (-2) \times 5 = -20.\]
10.2 The Characteristic Polynomial and Computing Eigenvalues

The function
\[ \phi(\lambda) = \det(A - \lambda I) \]
is called the *characteristic function* of the matrix \( A \). \( \lambda \) is an eigenvalue of \( A \) if \( \phi(\lambda) = 0 \), or, in other words, \( \lambda \) is a root of \( \phi(\lambda) \). The characteristic function is a polynomial of degree \( n \) in \( \lambda \) and thus there are at most \( n \) unique eigenvalues. Observe that the roots of this polynomial may be complex numbers; when there are complex roots they appear in conjugate pairs. The *algebraic multiplicity* of \( \lambda \) is the number of times \( \lambda \) is a root \( \phi(\lambda) \). The geometric multiplicity is always less than or equal to the algebraic multiplicity.

To see how this is used to compute eigenvalues, let us return to the matrix at the beginning of this section:
\[
A = \begin{bmatrix}
1 & 2 \\
2 & 1
\end{bmatrix}
\]
so that
\[
\det(A - \lambda I) = (1 - \lambda)^2 - 4 = \lambda^2 - 2\lambda - 3 = \phi(\lambda).
\]
Setting \( \det(A - \lambda I) = 0 \) we have quadratic equation to solve in order to determine the eigenvalues. In this case we see that the eigenvalue are \( \lambda_1 = 3 \) and \( \lambda_2 = -1 \). We can now solve for the eigenvector associated with \( \lambda = -1 \):
\[
\begin{bmatrix}
1 & 2 \\
2 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} = -1 \begin{bmatrix}
x_1 \\
x_2
\end{bmatrix} \Rightarrow x_1 + 2x_2 = -x_1 \Rightarrow x_1 = -x_2,
\]
where the first implication is from the first row of the equation. Plugging this in to the second row we have that
\[
2x_1 + x_2 = -x_2 \Rightarrow -2x_2 = x_2
\]
which holds for all \( x_2 \), so we take \( x_2 = -1 \) and \( x_1 = 1 \). Thus the eigenvector associated with \( \text{lambda} = -1 \) is \( [1, -1]^T \). Note that the two eigenvectors are orthogonal to each other.

10.3 Diagonalizability

Two square matrices \( A \) and \( B \) are *similar* if they have the same eigenvalues. Suppose \( W \) is a non-singular matrix and that \( B = WAW^{-1} \); then \( A \) and \( B \) are similar:
\[
Ax = \lambda x \Rightarrow W^{-1}BWx = \lambda x \Rightarrow BWx = \lambda Wx
\]
so that \( B \) has eigenvalue \( \lambda \) and eigenvector \( Wx \). A matrix is *diagonalizable* whenever it is similar to a diagonal matrix, so that \( A = WDW^{-1} \). When this is so, the columns of \( W \) are the eigenvectors of \( A \) and the diagonal elements of \( D \) are the eigenvalues. The diagonal matrix is often denoted by \( \Lambda \) so that \( A = W\Lambda W^{-1} \) is the eigenvalue decomposition of \( A \).

When \( A \) is a symmetric matrix (\( A = A^T \)) then all of the eigenvalues are real and the eigenvectors are orthonormal. This was seen with the example matrix given before. In that case, the eigenvectors we computed were orthogonal to each other. In this case, the eigenvalue decomposition of \( A \) is
\[
A = QAQ^T
\]
and \( Q \) is an orthogonal matrix, and \( \Lambda \) is a diagonal matrix with elements that are the eigenvalues of \( A \).

If \( A^{-1} \) exists, its eigenvalues are the reciprocals of the eigenvalues of \( A \), and the eigenvectors are the same:
\[
A = QAQ^T \Rightarrow A^{-1} = Q\Lambda^{-1}Q^T, (\Lambda^{-1})_{ii} = \frac{1}{\Lambda_{ii}}.
\]
\( A^{-1} \) exists if and only if all of the eigenvalues of \( A \) are non-zero.
10.4 Definiteness

Let $A$ be a symmetric matrix then

- If all of the eigenvalues of $A$ are positive, $\lambda_i > 0$, the matrix is called positive definite. This equivalent to $x^T AX > 0$ for all $x \neq 0$.
- If all of the eigenvalues of $A$ are non-negative, $\lambda_i \geq 0$, the matrix is call positive semi-definite.
- If all of the eigenvalues of $A$ are negative, $\lambda_i < 0$, the matrix is call negative definite. This is equivalent to $x^T Ax < 0$ for all $x \neq 0$.
- If the eigenvalues of $A$ are of mixed sign, the matrix is called indefinite.

10.5 The Gershgorin Circle Theorem

Let $A$ be a square matrix. Define $r_i = \sum_{j=1, j \neq i}^{n} |a_{ij}|$ and define the sets 

$$S_i = \{ x \in \mathbb{C} : |a_{ii} - x| \leq r_i \}.$$ 

The eigenvalue of $A$ lie in the union of the sets $S_i$: 

$$\lambda(A) \in \bigcup_{i=1}^{n} S_i.$$ 

As an example, consider the matrix:

$$A = \begin{bmatrix} 1 & 0.5 & -0.25 \\ 0.5 & 2 & -1 \\ -0.25 & -1 & 3 \end{bmatrix}.$$ 

Note that $A$ is symmetric so the eigenvalues are real. Thus ours sets will be intervals on the real line. Then 

$$r_1 = |a_{12}| + |a_{13}| = |0.5| + |-0.25| = 0.75$$

$$r_2 = |a_{21}| + |a_{23}| = |0.5| + |-1| = 1.5$$

$$r_3 = |a_{31}| + |a_{32}| = |-0.25| + |-1| = 1.25$$

and the sets are defined as

$$S_1 = [a_{11} - r_1, a_{11} + r_1] = [1 - 0.75, 1 + 0.75] = [0.25, 1.75]$$

$$S_2 = [a_{22} - r_2, a_{22} + r_2] = [2 - 1.5, 2 + 1.5] = [0.5, 3.5]$$

$$S_3 = [a_{33} - r_3, a_{33} + r_3] = [3 - 1.25, 3 + 1.25] = [1.75, 4.25].$$

The union of these sets, $\bigcup_{i=1}^{n} S_i = [0.25, 4.25]$. Thus the eigenvalues of $A$ lie in the interval $[0.25, 4.25]$. As this region is strictly positive, the eigenvalues of $A$ must be strictly positive and the matrix is therefore positive definite.

11 The Singular Value Decomposition

Let $A \in \mathbb{R}^{m \times n}$ be an arbitrary matrix. The singular value decomposition or SVD is a generalization of the eigenvalue decomposition to non-square matrices. Every matrix $A$ can be decomposed as follows:

$$A = U \Sigma V^T$$
where $U$ is an orthogonal $m \times m$ matrix, $V$ is an orthogonal $n \times n$ matrix and $\Sigma$ is “diagonal” $m \times n$ matrix. The elements of $\Sigma$ are all zero except for $\sigma_{ii}$ which are non-negative and are typically denoted $\sigma_i$. The $\sigma_i$s are called the singular values of $A$ and are ranked so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_{\min(m,n)} = 0$. The rank of $A$ is the number of non-zero singular values and $\mathcal{R}(A) = \text{span}(\{u_1, \ldots, u_r\})$ and $\mathcal{N}(A) = \text{span}(\{v_{r+1}, \ldots, v_n\})$. For a symmetric matrix, the singular values are the absolute values of the eigenvalues, and in general, are related to the eigenvalues of $A^T A$ by

$$\sigma_i^2(A) = \lambda_i(A^T A).$$

To see this, observe that $A^T A$ is symmetric and thus has an eigenvalue decomposition $Q \Lambda Q^T$. $A^T A$ is also equal to $U^T \Sigma^2 V V^T \Sigma^2 U = U^T \Sigma^2 \Sigma U$ and that $\Sigma^2 \Sigma$ is a diagonal matrix with entries given by $\sigma_i^2$. Thus since the eigenvalue decomposition is unique, $\sigma_i^2 = \lambda_i$. The SVD decomposes the matrix into a rotations, followed by a scaling and then another rotation.