EE731 Lecture Notes: Matrix Computations for Signal Processing

James P. Reilly©
Department of Electrical and Computer Engineering
McMaster University

September 24, 2006

0 Preface

This collection of ten chapters of notes will give the reader an introduction to the fundamental principles of linear algebra for application in many disciplines of modern engineering and science, including signal processing, control theory, process control, applied statistics, robotics, etc. We assume the reader has an equivalent background to a freshman course in linear algebra, some introduction to probability and statistics, and a basic knowledge of the Fourier transform.

The first chapter, some fundamental ideas required for the remaining portion of the course are established. First, we look at some fundamental ideas of linear algebra such as linear independence, subspaces, rank, nullspace, range, etc., and how these concepts are interrelated.

In chapter 2, the most basic matrix decomposition, the so–called eigendecomposition, is presented. The focus of the presentation is to give an intuitive insight into what this decomposition accomplishes. We illustrate how the eigendecomposition can be applied through the Karhunen-Loeve transform. In this way, the reader is made familiar with the important properties of this decomposition. The Karhunen-Loeve transform is then generalized to the broader idea of transform coding. The ideas of autocorrelation, and the covariance matrix of a signal, are discussed and interpreted.

In chapter 3, we develop the *singular value decomposition* (SVD), which is closely related to the eigendecomposition of a matrix. We develop the relationships between these two decompositions and explore various properties of the SVD.

Chapter 4 deals with the quadratic form and its relation to the eigendecomposition, and also gives an introduction to error mechanisms in floating point number systems. The condition number of a matrix, which is a critical part in determining a lower bound on the relative error in the solution of a system of linear equations, is also developed.

Chapters 5 and 6 deal with solving linear systems of equations by Gaussian elimination. The Gaussian elimination process is described through a bigger–block matrix approach, that leads to other useful decompositions, such as the Cholesky decomposition of a square symmetric matrix.

Chapters 7–10 deal with solving least–squares problems. The standard least squares problem and its solution are developed in Chapter 7. In Chapter 8, we develop a generalized "pseudoinverse" approach to solving the least–squares problem. The QR decomposition in developed in Chapter 9, and its application to the solution of linear least squares problems is discussed in Chapter 10.

Finally, in Chapter 11, the solution of Toeplitz systems of equations and its underlying theory is developed.

1 Fundamental Concepts

The purpose of this lecture is to review important fundamental concepts in linear algebra, as a foundation for the rest of the course. We first discuss the fundamental building blocks, such as an overview of matrix multiplication from a "big block" perspective, linear independence, subspaces and related ideas, rank, etc., upon which the rigor of linear algebra rests. We then discuss vector norms, and various interpretations of the matrix multiplication operation. We close the chapter with a discussion on determinants.

1.1 Notation

Throughout this course, we shall indicate that a matrix $\mathbf{A}$ is of dimension $m \times n$, and whose elements are taken from the set of real numbers, by the notation $\mathbf{A} \in \mathbb{R}^{m \times n}$. This means that the matrix $\mathbf{A}$ belongs to the Cartesian product of the real numbers, taken $m \times n$ times, one for each element of $\mathbf{A}$. In a similar way, the notation $\mathbf{A} \in \mathbb{C}^{m \times n}$ means the matrix is of dimension $m \times n$, and the elements are taken from the set of complex numbers. By the matrix dimension "$m \times n$", we mean $\mathbf{A}$ consists of m rows and n columns.

Similarly, the notation $\mathbf{a} \in \mathbb{R}^m (\mathbb{C}^m)$ implies a vector of m elements which are taken from the set of real (complex) numbers. When referring to a single vector, we use the term *dimension* to denote the number of elements.

Also, we shall indicate that a scalar a is from the set of real (complex) numbers by the notation $a \in \mathbb{R}(\mathbb{C})$. Thus, an upper case bold character denotes a *matrix*, a lower case bold character denotes a vector, and a lower case non-bold character denotes a scalar.

By convention, a vector by default is taken to be a *column* vector. Further, for a matrix $\mathbf{A}$, we denote its ith column as $\mathbf{a}_i$. We also imply that its jth row is $\mathbf{a}_j^T$, even though this notation may be ambiguous, since it may also be taken to mean the transpose of the jth column. The context of the discussion will help to resolve the ambiguity.

1.2 Fundamental Linear Algebra

1.2.1 Vector Spaces

Formally, a vector space is defined as follows:

A vector space $\mathcal{S}$ satisfies two requirements:

1. If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathcal{S}$, then $\mathbf{x} + \mathbf{y}$ is still in $\mathcal{S}$.
2. If we multiply any vector $\mathbf{x}$ in $\mathcal{S}$ by a scalar c, then $c\mathbf{x}$ is still in $\mathcal{S}$.

This definition implies that if a set of vectors are in a vector space, then any linear combination of these vectors are also in the space. We now expand on this definition of a vector space, as follows.

Suppose we have a set of vectors $[\mathbf{a}_1, \ldots, \mathbf{a}_n]$, where $\mathbf{a}_i \in \mathbb{R}^m, i = 1, \ldots, n$, and a set of scalars $c_i \in \mathbb{R}, i = 1, \ldots, n$. Then the vector $\mathbf{y} \in \mathbb{R}^m$ defined by

$$\mathbf{y} = \sum_{i=1}^{n} c_i \mathbf{a}_i \tag{1}$$

is referred to as a *linear combination* of the vectors $\mathbf{a}_i$. (Note that in this section and in the sequel, all column vectors are assumed to be of length m, unless stated otherwise).

We wish to see if the above equation can be represented as a matrix–vector multiplication, which has the advantage of being more compact. We note that each coefficient c_i in (1) multiplies all elements in the corresponding vector $\mathbf{a}_i$.

$$\mathbf{y} = \underset{\mathbf{A}}{\begin{bmatrix} | & | & & | \\ | & | & & | \\ | & | & \cdots & | \end{bmatrix}} \underset{\mathbf{c}}{\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}} \tag{2}$$

Consider the accompanying diagram, depicting matrix–vector multiplication, where each vertical line represents an entire column $\mathbf{a}_i$ of $\mathbf{A}$. In a manner similar to (1), from the rules of matrix–vector multiplication, we see that each element c_i of $\mathbf{c}$ multiples only elements of the corresponding column $\mathbf{a}_i$; i.e., coefficient c_i interacts only with the column $\mathbf{a}_i$. Thus, (1) can be written in the form

$$\mathbf{y} = \mathbf{A}\mathbf{c} \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n} = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$, and $\mathbf{c} \in \mathbb{R}^n = [c_1, \ldots, c_n]^T$.

Instead of using (3) to define a single vector, we can use it to define a set of vectors, which we will denote as $\mathcal{S}$. Consider the expression

$$\mathcal{S} = \{\mathbf{y} \in \mathbb{R}^m | \mathbf{y} = \mathbf{A}\mathbf{c}, \mathbf{c} \in \mathbb{R}^n\} \tag{4}$$

where now it is implied that $\mathbf{c}$ takes on all possible values within $\mathbb{R}^n$. The set $\mathcal{S}$ defined in this way is referred to as a *vector space*, and is the set of all linear combinations of the vector set. The *dimension* of the vector space is the number of independent directions that span the space; e.g., the dimension of the universe is 3.

The dimension of the vector space $\mathcal{S}$ $\big($denoted as $\dim(\mathcal{S})\big)$ is not necessarily n, the number of vectors or columns of $\mathbf{A}$. In fact, $\dim(\mathcal{S}) \leq n$. The quantity $\dim(\mathcal{S})$ depends on the characteristics of the vectors $\mathbf{a}_i$. For example, the vector space defined by the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ in Fig. 1 below is the plane of the paper. The dimension of this vector space is 2:

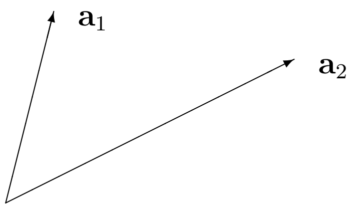

Figure 1: A vector set containing two vectors.

If a third vector $\mathbf{a}_3$ which is orthogonal to the plane of the paper were added to the set, then the resulting vector space would be the three–dimensional universe. A third example is shown in Figure 2. Here, since none of the vectors $\mathbf{a_1} \ldots, \mathbf{a}_3$ have a component which is orthogonal to the plane of the

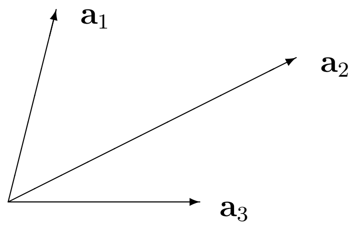

Figure 2: A vector set containing three vectors.

paper, all linear combinations of this vector set, and hence the corresponding vector space, lies in the plane of the paper. Thus, in this case, $\dim(\mathcal{S})$ is 2, even though there are three vectors in the set.

In this section we have defined the notion of a vector space and its dimension. Note that the term *dimension* when applied to a vector is the number of elements in the vector. The term *length* is also used to indicate the number of elements of a vector.

1.2.2 Linear Independence

A vector set $[\mathbf{a}_1, \ldots, \mathbf{a}_n]$ is linearly independent under the conditions

$$\sum_{j=1}^{n} c_j \mathbf{a}_j = \mathbf{0} \quad \text{if and only if} \quad c_1, \ldots, c_n = 0 \tag{5}$$

This means that means that a set of vectors is linearly independent if and only if the only zero linear combination of the vectors has coefficients which are all zero.

Let $\mathcal{S}$ be the vector space corresponding to the vector set $[\mathbf{a}_1, \ldots, \mathbf{a}_n]$. This set of n vectors is linearly independent if and only if $\dim(\mathcal{S}) = n$. If $\dim(\mathcal{S}) < n$, then the vectors are linearly dependent. Note that a set of vectors $[\mathbf{a}_1, \ldots, \mathbf{a}_n]$, where $n > m$ cannot be linearly independent. Further, a linearly dependent vector set can be made independent by removing vectors from the set.

Example 1

$$\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3] = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 3 & -1 \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

This set is linearly independent. On the other hand, the set

$$\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2\ \mathbf{b}_3] = \begin{bmatrix} 1 & 2 & -3 \\ 0 & 3 & -3 \\ 1 & 1 & -2 \end{bmatrix} \tag{7}$$

is not. This follows because the third column is a linear combination of the first two. (-1 times the first column plus -1 times the second equals the third column). Thus, the coefficients c_j in (5) resulting in zero are any scalar multiple of (1, 1, 1).

1.2.3 Span, Range, and Subspaces

In this section, we explore these three closely-related ideas. In fact, their mathematical definitions are almost the same, but the interpretation is different for each case.

Span:

The span of a vector set $[\mathbf{a}_1, \ldots, \mathbf{a}_n]$, written as span$[\mathbf{a}_1, \ldots, \mathbf{a}_n]$, is the vector space $\mathcal{S}$ corresponding to this set; i.e.,

$$\text{span}\,[\mathbf{a}_1, \ldots, \mathbf{a}_n] = \left\{ \mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \sum_{j=1}^{n} c_j \mathbf{a}_j, \quad c_j \in \mathbb{R} \right\} = \mathcal{S}. \tag{8}$$

Subspaces

A subspace is a subset of a vector space. More precisely, a k–dimensional subspace $\mathcal{U}$ of $\mathcal{S} = \text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n]$ is determined by span$[\mathbf{a}_{i_1}, \ldots \mathbf{a}_{i_k}]$, where the indices satisfy $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$.

Note that $[\mathbf{a}_{i1}, \ldots \mathbf{a}_{ik}]$ is not necessarily *a basis* for the subspace S. This set is a basis only if it is a maximally independent set. This idea is discussed shortly. The set $\{\mathbf{a}_i\}$ need not be linearly independent to define the span or subset.

For example, the vectors $[\mathbf{a}_1, \mathbf{a}_2]$ in Fig. 1 define a subspace (the plane of the paper) which is a subset of the three–dimensional universe $\mathbb{R}^3$.

$*$ What is the span of the vectors $[\mathbf{b}_1, \ldots, \mathbf{b}_3]$ in example 1?

Range:

The *range* of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted $R(\mathbf{A})$, is the vector space satisfying

$$R(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \mathbf{A}\mathbf{x}, \text{ for } \mathbf{x} \in \mathbb{R}^n\}. \tag{9}$$

Thus, we see that $R(\mathbf{A})$ is the vector space consisting of all linear combinations of the columns $\mathbf{a}_i$ of $\mathbf{A}$, whose coefficients are the elements x_i of $\mathbf{x}$. Therefore, $R(\mathbf{A}) \equiv \text{span}[\mathbf{a}_1, \ldots, \mathbf{a}_n]$. The distinction between *range* and *span* is that the argument of *range* is a matrix, while for *span* it is a set of vectors. If the columns of $\mathbf{A}$ are (not) linearly independent, then $R(\mathbf{A})$ will (not) span n dimensions. Thus, the dimension of the vector space $R(\mathbf{A})$ is less than or equal to n. Any vector $\mathbf{y} \in R(\mathbf{A})$ is of dimension (length) m.

Example 3:

$$\mathbf{A} = \begin{bmatrix} 1 & 5 & 3 \\ 2 & 4 & 3 \\ 3 & 3 & 3 \end{bmatrix} \text{(the last column is the average of the first two)} \tag{10}$$

$R(\mathbf{A})$ is the set of all linear combinations of any two columns of $\mathbf{A}$.

In the case when $n < m$ (i.e., $\mathbf{A}$ is a *tall* matrix), it is important to note that $R(\mathbf{A})$ is indeed a subspace of the m-dimensional "universe" $\mathbb{R}^m$. In this case, the dimension of $R(\mathbf{A})$ is less than or equal to n. Thus, $R(\mathbf{A})$ does not span the whole universe, and therefore is a subspace of it.

1.2.4 Maximally Independent Set

This is a vector set which cannot be made larger without losing independence, and smaller without remaining maximal; i.e. it is a set containing the maximum number of independent vectors spanning the space.

1.2.5 A Basis

A basis for a subspace is any maximally independent set within the subspace. It is not unique.

Example 4. A basis for the subspace S spanning the first 2 columns of

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ & 3 & -3 \\ & & 3 \end{bmatrix}, \text{ i.e., } S = \text{span}\left[\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}\right]$$

is

$$\begin{aligned} \mathbf{e}_1 &= (1,0,0)^T \\ \mathbf{e}_2 &= (0,1,0)^T. \end{aligned}$$

[1]or any other linearly independent set in span$[\mathbf{e}_1, \mathbf{e}_2]$.

Any vector in S is *uniquely* represented as a linear combination of the basis vectors.

1.2.6 Orthogonal Complement Subspace

If we have a subspace S of dimension n consisting of vectors $[\mathbf{a}_1, \ldots, \mathbf{a}_n], \mathbf{a}_i \in \mathbb{R}^m, i = 1, \ldots, n$, for $n \leq m$, the orthogonal complement subspace $S_\perp$ of S of dimension $m - n$ is defined as

$$S_\perp = \left\{ \mathbf{y} \in \mathbb{R}^m | \mathbf{y}^T \mathbf{x} = 0 \text{ for all } \mathbf{x} \in S \right\} \tag{11}$$

[1]A vector $\mathbf{e}_i$ is referred to as an *elementary* vector, and has zeros everywhere except for a 1 in the ith position.

i.e., any vector in $S_\perp$ is orthogonal to any vector in S. The quantity $S_\perp$ is pronounced "S–perp".

Example 5: Take the vector set defining S from Example 4:

$$S \equiv \begin{bmatrix} 1 & 2 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} \tag{12}$$

then, a basis for $S_\perp$ is

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{13}$$

1.2.7 Rank

Rank is an important concept which we will use frequently throughout this course. We briefly describe only a few basic features of rank here. The idea is expanded more fully in the following sections.

1. The rank of a matrix $\mathbf{A}$ (denoted rank($\mathbf{A}$)), is the maximum number of linearly independent rows or columns in $\mathbf{A}$. Thus, it is the dimension of $R(\mathbf{A})$. The symbol r is commonly used to denote rank; i.e., $r = \text{rank}(\mathbf{A})$.
2. if $\mathbf{A} = \mathbf{BC}$, and $r_1 = \text{rank}(\mathbf{B}), r_2 = \text{rank}(\mathbf{C})$, then $\text{rank}(\mathbf{A}) \leq \min(r_1, r_2)$.
3. A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to be *rank deficient* if its rank is less than $\min(m, n)$. Otherwise, it is said to be *full rank*.
4. If $\mathbf{A}$ is square and rank deficient, then $\det(\mathbf{A}) = 0$.
5. It can be shown that $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$. More is said on this point later.

A matrix is said to be *full column (row) rank* if its rank is equal to the number of columns (rows).

Example 6: The rank of **A** in Example 4 is 3, whereas the rank of **A** in Example 3 is 2.

Example 7: Consider the matrix multiplication $\mathbf{C} \in \mathbb{R}^{m\times n} = \mathbf{AB}$, where $\mathbf{A} \in \mathbb{R}^{m\times 2}$ and $\mathbf{B} \in \mathbb{R}^{2\times n}$, depicted by the following diagram:

$$\underset{\mathbf{C}}{\begin{bmatrix} | & | & | & | \\ | & | & | & | \\ | & | & | & | \end{bmatrix}} = \underset{\mathbf{A}}{\begin{bmatrix} | & | \\ | & | \\ | & | \end{bmatrix}} \underset{\mathbf{B}}{\begin{bmatrix} x & x & x & x \\ x & x & x & x \end{bmatrix}} . \qquad (14)$$

Then, the rank of **C** is at most two. To see this, we realize from our discussion on the relation between matrix multiplication and the operation of forming linear combinations that the ith column of **C** is a linear combination of the two columns of **A** whose coefficients are the ith column of **B**. Thus, all columns of **C** reside in the vector space $R(\mathbf{A})$. If the columns of **A** and the rows of **B** are linearly independent, then the dimension of this vector space is two, and hence $\text{rank}(\mathbf{C}) = 2$. If the columns of **A** or the rows of **B** are linearly *dependent*, then $\text{rank}(\mathbf{C}) = 1$. This example can be extended in an obvious way to matrices of arbitrary size.

1.2.8 Null Space of A

The null space $N(\mathbf{A})$ of **A** is defined as

$$N(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \neq \mathbf{0} \mid \mathbf{Ax} = \mathbf{0}\} . \qquad (15)$$

From previous discussions, the product $\mathbf{Ax}$ is a linear combination of the columns $\mathbf{a}_i$ of **A**, where the elements x_i of $\mathbf{x}$ are the corresponding coefficients. Thus, from (15), $N(\mathbf{A})$ is the set of non–zero coefficients of all zero linear combinations of the columns of **A**. If the columns of **A** are linearly independent, then $N(\mathbf{A}) = \emptyset$ by definition, because there can be no coefficients except zero which result in a zero linear combination. In this case, the dimension of the null space is zero, and **A** is full column rank. The null space is empty if and only if $\mathbf{A}$ is full column rank, and is non–empty when $\mathbf{A}$ is column rank deficient. Note that any vector in $N(\mathbf{A})$ is of dimension n. Any vector in $N((A))$ is orthogonal to the rows of $\mathbf{A}$, and is thus in the orthogonal complement of the span of the rows of $\mathbf{A}$.

Example 8: Let $\mathbf{A}$ be as before in Example 3. Then $N(\mathbf{A}) = c(1,1,-2)^T$, where $c \in \mathbb{R}$.

A further example is as follows. Take 3 vectors $[\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$ where $\mathbf{a}_i \in \mathbb{R}^3, i = 1,\ldots,3$, that are constrained to lie in a 2–dimensional plane. Then there exists a zero linear combination of these vectors. The coefficients of this linear combination define a vector $\mathbf{x}$ which is in the nullspace of $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3]$. In this case, we see that $\mathbf{A}$ is rank deficient.

Another important characterization of a matrix is its *nullity*. The nullity of $\mathbf{A}$ is the dimension of the nullspace of $\mathbf{A}$. In Example 6 above, the nullity of $\mathbf{A}$ is one. We then have the following interesting property:

$$\text{rank}(\mathbf{A}) + \text{nullity}(\mathbf{A}) = n. \tag{16}$$

1.3 Four Fundamental Subspaces of a Matrix

The four matrix subspaces of concern are: *the column space, the row space*, and their respective *orthogonal complements*. The development of these four subspaces is closely linked to $N(A)$ and $R(A)$. We assume for this section that $\mathbf{A} \in \mathbb{R}^{m \times n}$, $r \leq \min(m,n)$, where $r = \text{rank}\mathbf{A}$.

1.3.1 The Column Space

This is simply $R(\mathbf{A})$. Its dimension is r. It is the set of all linear combinations of the columns of $\mathbf{A}$. Any vector in $R(\mathbf{A})$ is of dimension m.

1.3.2 The Orthogonal Complement of the Column Space

This may be expressed as $R(\mathbf{A})_\perp$, with dimension $m - r$. It may be shown to be equivalent to $N(\mathbf{A}^T)$, as follows: By definition, $N(\mathbf{A}^T)$ is the set $\mathbf{x}$ satisfying:

$$\underbrace{\begin{bmatrix} \rule{2cm}{0.4pt} \\ \rule{2cm}{0.4pt} \\ \rule{2cm}{0.4pt} \\ \rule{2cm}{0.4pt} \end{bmatrix}}_{\mathbf{A}^T} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \mathbf{0}, \tag{17}$$

where columns of $\mathbf{A}$ are the rows of $\mathbf{A}^T$. From (17), we see that $N(\mathbf{A}^T)$ is the set of $\mathbf{x} \in \mathbb{R}^m$ which is orthogonal to all columns of $\mathbf{A}$ (rows of $\mathbf{A}^T$). This by definition is the orthogonal complement of $R(\mathbf{A})$. Any vector in $R(\mathbf{A})_\perp$ is of dimension m.

1.3.3 The Row Space

The row space is defined simply as $R(\mathbf{A}^T)$, with dimension r. The row space is the range of the rows of $\mathbf{A}$, or the subspace spanned by the rows, or the set of all possible linear combinations of the rows of $\mathbf{A}$. Any vector in $R(\mathbf{A}^T)$ is of dimension n.

1.3.4 The Orthogonal Complement of the Row Space

This may be denoted as $R(\mathbf{A}^T)_\perp$. Its dimension is $n - r$. This set must be that which is orthogonal to all rows of $\mathbf{A}$: i.e., for $\mathbf{x}$ to be in this space, $\mathbf{x}$ must satisfy

$$\begin{matrix} \text{rows} \\ \text{of} \\ \mathbf{A} \end{matrix} \quad \rightarrow \begin{bmatrix} \rule{0.7cm}{0.4pt} \\ \rule{0.7cm}{0.4pt} \\ \rule{0.7cm}{0.4pt} \\ \vdots \\ \rule{0.7cm}{0.4pt} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{0}. \tag{18}$$

Thus, the set $\mathbf{x}$, which is the orthogonal complement of the row space satisfying (18), is simply $N(\mathbf{A})$. Any vector in $R(\mathbf{A}^T)_\perp$ is of dimension n.

We have noted before that $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$. Thus, the dimension of the row and column subspaces are equal. This is surprising, because it implies the number of linearly independent rows of a matrix is the same as the number of linearly independent columns. This holds regardless of the size or rank of the matrix. It is not an intuitively obvious fact and there is no immediately obvious reason why this should be so. Nevertheless, the rank of a matrix is the number of independent rows *or* columns.

1.4 "Bigger-Block" Interpretations of Matrix Multiplication

In this section, we take a look at matrix multiplication from the viewpoint that columns of a matrix product are linear combinations of the columns of the first matrix. To standardize our discussion, let us define the matrix product $\mathbf{C}$ as

$$\underset{m\times n}{\mathbf{C}} = \underset{m\times k}{\mathbf{A}} \; \underset{k\times n}{\mathbf{B}} \tag{19}$$

The three interpretations of this operation now follow:

1.4.1 Inner-Product Representation

If $\mathbf{a}$ and $\mathbf{b}$ are column vectors of the same length, then the scalar quantity $\mathbf{a}^T\mathbf{b}$ is referred to as the *inner product* of $\mathbf{a}$ and $\mathbf{b}$. If we define $\mathbf{a}_i^T \in \mathbb{R}^k$ as the ith row of $\mathbf{A}$ and $\mathbf{b}_j \in \mathbb{R}^k$ as the jth column of $\mathbf{B}$, then the element c_{ij} of $\mathbf{C}$ is defined as the inner product $\mathbf{a}_i^T\mathbf{b}_j$. This is the conventional small-block representation of matrix multiplication.

1.4.2 Column Representation

This is the next bigger–block view of matrix multiplication. Here we look at forming the product one column at a time. The jth column $\mathbf{c}_j$ of $\mathbf{C}$ may be expressed as a linear combination of columns $\mathbf{a}_i$ of $\mathbf{A}$ with coefficients which are the elements of the jth column of $\mathbf{B}$. Thus,

$$\mathbf{c}_j = \sum_{i=1}^{k} \mathbf{a}_i b_{ij}, \qquad j = 1, \ldots, n. \tag{20}$$

This operation is identical to the inner–product representation above, except we form the product one column at a time. For example, if we evaluate only the pth element of the jth column $\mathbf{c}_j$, we see that (20) degenerates into $\sum_{i=1}^{k} a_{pi} b_{ij}$. This is the inner product of the pth row and jth column of $\mathbf{A}$ and $\mathbf{B}$ respectively, which is the required expression for the (p, j)th element of $\mathbf{C}$.

1.4.3 Outer–Product Representation

This is the largest–block representation. Let us define a column vector $\mathbf{a} \in \mathbb{R}^m$ and a row vector $\mathbf{b}^T \in \mathbb{R}^n$. Then the *outer product* of $\mathbf{a}$ and $\mathbf{b}$ is an $m \times n$ matrix of rank one and is defined as $\mathbf{a}\mathbf{b}^T$.

Now let $\mathbf{a}_i$ and $\mathbf{b}_i^T$ be the ith column and row of $\mathbf{A}$ and $\mathbf{B}$ respectively. Then the product $\mathbf{C}$ may also be expressed as

$$\mathbf{C} = \sum_{i=1}^{k} \mathbf{a}_i \mathbf{b}_i^T. \tag{21}$$

By looking at this operation one column at a time, we see this form of matrix multiplication performs exactly the same operations as the column representation above. For example, the jth column $\mathbf{c}_j$ of the product is determined from (21) to be $\mathbf{c}_j = \sum_{i=1}^{k} \mathbf{a}_i b_{ij}$, which is identical to (20) above.

1.4.4 Matrix Multiplication Again

Here we give an alternate interpretation for matrix multiplication by comparing this operation to that of forming linear combinations. Consider a matrix $\mathbf{A}$ *pre–multiplied* by $\mathbf{B}$ to give $\mathbf{Y} = \mathbf{BA}$. ($\mathbf{A}$ and $\mathbf{B}$ are assumed to have conformable dimensions). Let us assume $\mathbf{Y} \in \Re^{m \times n}$. Then we can interpret this operation in two ways:

- Each column $\mathbf{y}_i, i = 1, \ldots, n$ of $\mathbf{Y}$ is a linear combination of the *columns* of $\mathbf{B}$, whose coefficients are the ith column $\mathbf{a}_i$ of $\mathbf{A}$; i.e.,

$$\mathbf{y}_i = \sum_{k=1}^{n} \mathbf{b}_k a_{ki} = \mathbf{B}\mathbf{a}_i \tag{22}$$

  This operation is very similar to the column representation for matrix multiplication.

- Each row $\mathbf{y}_j^T, j = 1, \ldots, m$ of $\mathbf{Y}$ is a linear combination of the *rows* of $\mathbf{A}$, whose coefficients are the jth row of $\mathbf{B}$; i.e.,

$$\mathbf{y}_j^T = \sum_{k=1}^{m} b_{jk} \mathbf{a}_k^T = \mathbf{b}_j^T \mathbf{A}. \tag{23}$$

  Using this idea, can matrix multiplication be cast in a *row* representation format?

- A further related idea is as follows. Consider an orthonormal matrix $\mathbf{Q}$ of appropriate dimension. We know that multiplication of a vector by an orthonormal matrix results in a rotation of the vector. The operation $\mathbf{QA}$ rotates each column of $\mathbf{A}$. The operation $\mathbf{AQ}$ rotates each row.

1.5 Vector Norms

A *vector norm* is a means of expressing the length or distance associated with a vector. A norm on a vector space $\mathbb{R}^n$ is a *function* f, which maps a point in

$\mathbb{R}^n$ into a point in $\mathbb{R}$. Formally, this is stated mathematically as $f : \mathbb{R}^n \rightarrow \mathbb{R}$. The norm has the following properties:

1. $f(\mathbf{x}) \geq 0 \quad$ for all $\mathbf{x} \in \mathbb{R}^n$.
2. $f(x) = 0 \quad$ if and only if $\quad \mathbf{x} = 0$.
3. $f(\mathbf{x} + \mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y}) \quad$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
4. $f(a\mathbf{x}) = |a| f(\mathbf{x}) \quad$ for $a \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n$.

We denote the function $f(\mathbf{x})$ as $||\mathbf{x}||$.

The p-norms: This is a useful class of norms, generalizing on the idea of the Euclidean norm. They are defined by

$$||\mathbf{x}||_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p}. \tag{24}$$

If $p = 1$:

$$||\mathbf{x}||_1 = \sum_i |x_i|$$

which is simply the sum of absolute values of the elements.

If $p = 2$:

$$||\mathbf{x}||_2 = \left(\sum_i {x_i}^2 \right)^{\frac{1}{2}} = (\mathbf{x}^T \mathbf{x})^{\frac{1}{2}}$$

which is the familiar Euclidean norm.

If $p = \infty$:

$$||\mathbf{x}||_\infty = \max_i |x_i|$$

which is the largest element of $\mathbf{x}$. This may be shown in the following way. As $p \rightarrow \infty$, the largest term within the round brackets in (24) dominates all the others. Therefore (24) may be written as

$$
\begin{aligned}
||\mathbf{x}||_\infty = \lim_{p\to\infty} \left[\sum_{i=1}^{n} x_i^p\right]^{\frac{1}{p}} &= \lim_{p\to\infty} [x_k^p]^{\frac{1}{p}} \\
&= x_k \qquad (25)
\end{aligned}
$$

where k is the index corresponding to the largest element x_i.

Note that the $p = 2$ norm has many useful properties, but is expensive to compute. Obviously, the 1– and ∞–norms are easier to compute, but are more difficult to deal with algebraically. All the p–norms obey all the properties of a vector norm.

1.6 Determinants

Consider a square matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$. We can define the matrix $\mathbf{A}_{ij}$ as the submatrix obtained from $\mathbf{A}$ by deleting the ith row and jth column of $\mathbf{A}$. The scalar number $\det(\mathbf{A}_{ij})$ (where $\det(\cdot)$ denotes *determinant*) is called the *minor* associated with the element a_{ij} of $\mathbf{A}$. The signed minor $c_{ij} \triangleq (-1)^{j+i}\det(\mathbf{A}_{ij})$ is called the *cofactor* of a_{ij}.

The determinant of $\mathbf{A}$ is the m-dimensional volume contained within the columns (rows) of $\mathbf{A}$. This interpretation of determinant is very useful as we see shortly. The determinant of a matrix may be evaluated by the expression

$$
\det(\mathbf{A}) = \sum_{j=1}^{m} a_{ij}c_{ij}, \qquad i \in (1\ldots m). \qquad (26)
$$

or

$$
\det(\mathbf{A}) = \sum_{i=1}^{m} a_{ij}c_{ij}, \qquad j \in (1\ldots m). \qquad (27)
$$

Both the above are referred to as the *cofactor expansion* of the determinant. Eq. (26) is along the ith *row* of $\mathbf{A}$, whereas (27) is along the jth *column*. It is indeed interesting to note that both versions above give exactly the same number, regardless of the value of i or j.

Eqs. (26) and (27) express the $m \times m$ determinant det$\mathbf{A}$ in terms of the cofactors c_{ij} of $\mathbf{A}$, which are themselves $(m-1) \times (m-1)$ determinants. Thus, $m-1$ recursions of (26) or (27) will finally yield the determinant of the $m \times m$ matrix $\mathbf{A}$.

From (26) it is evident that if $\mathbf{A}$ is triangular, then $\det(\mathbf{A})$ is the product of the main diagonal elements. Since diagonal matrices are in the upper triangular set, then the determinant of a diagonal matrix is also the product of its diagonal elements.

Properties of Determinants

Before we begin this discussion, let us define the volume of a parallelopiped defined by the set of column vectors comprising a matrix as the *principal volume* of that matrix.

We have the following properties of determinants, which are stated without proof:

1. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}) \qquad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times m}$.
   The principal volume of the product of matrices is the product of principal volumes of each matrix.

2. $\det(\mathbf{A}) = \det(\mathbf{A}^T)$
   This property shows that the characteristic polynomials [2] of $\mathbf{A}$ and $\mathbf{A}^T$ are identical. Consequently, as we see later, eigenvalues of $\mathbf{A}^T$ and $\mathbf{A}$ are identical.

3. $\det(c\mathbf{A}) = c^m \det(\mathbf{A}) \qquad c \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{m \times m}$.
   This is a reflection of the fact that if each vector defining the principal volume is multiplied by c, then the resulting volume is multiplied by c^m.

4. $\det(\mathbf{A}) = 0 \rightleftharpoons \mathbf{A}$ is singular.
   This implies that at least one dimension of the principal volume of the corresponding matrix has collapsed to zero length.

[2] The characteristic polynomial of a matrix is defined in Chapter 2.

5. $\det(\mathbf{A}) = \prod_{i=1}^{m} \lambda_i$, where λ_i are the eigen (singular) values of $\mathbf{A}$.
   This means the parallelopiped defined by the column or row vectors of a matrix may be transformed into a regular rectangular solid of the same $m-$ dimensional volume whose edges have lengths corresponding to the eigen (singular) values of the matrix.

6. The determinant of an orthonormal[3] matrix is ± 1.
   This is easy to see, because the vectors of an orthonormal matrix are all unit length and mutually orthogonal. Therefore the corresponding principal volume is ± 1.

7. If $\mathbf{A}$ is nonsingular, then $\det(\mathbf{A}^{-1}) = [\det(\mathbf{A})]^{-1}$.

8. If $\mathbf{B}$ is nonsingular, then $\det(\mathbf{B}^{-1}\mathbf{A}\mathbf{B}) = \det(\mathbf{A})$.

9. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by interchanging any two rows (or columns), then $\det(\mathbf{B}) = -\det(\mathbf{A})$.

10. If $\mathbf{B}$ is obtained from $\mathbf{A}$ by by adding a scalar multiple of one row to another (or a scalar multiple of one column to another), then $\det(\mathbf{B}) = \det(\mathbf{A})$.

A further property of determinants allows us to compute the *inverse* of $\mathbf{A}$. Define the matrix $\tilde{\mathbf{A}}$ as the *adjoint* of $\mathbf{A}$:

$$\tilde{\mathbf{A}} = \begin{bmatrix} c_{11} & \ldots & c_{1m} \\ \vdots & & \vdots \\ c_{m1} & \ldots & c_{mm} \end{bmatrix}^T \tag{28}$$

where the c_{ij} are the cofactors of $\mathbf{A}$. According to (26) or (27), the ith row $\tilde{\mathbf{a}}_i^T$ of $\tilde{\mathbf{A}}$ times the ith column $\mathbf{a}_i$ is $\det(\mathbf{A})$; i.e.,

$$\tilde{\mathbf{a}}_i^T \mathbf{a}_i = \det(\mathbf{A}), \qquad i = 1, \ldots, m. \tag{29}$$

It can also be shown that

$$\tilde{\mathbf{a}}_i^T \mathbf{a}_j = 0, \qquad i \neq j. \tag{30}$$

[3]An *orthonormal* matrix is defined in Chapter 2.

Then, combining (29) and (30) for $i, j \in \{1, \ldots, m\}$ we have the following interesting property:

$$\tilde{\mathbf{A}}\mathbf{A} = \det(\mathbf{A})\mathbf{I}, \tag{31}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix. It then follows from (31) that the inverse $\mathbf{A}^{-1}$ of $\mathbf{A}$ is given as

$$\mathbf{A}^{-1} = [\det(\mathbf{A})]^{-1}\tilde{\mathbf{A}}. \tag{32}$$

Neither (26) nor (32) are computationally efficient ways of calculating a determinant or an inverse respectively. Better methods which exploit the properties of various matrix decompositions are made evident later in the course.