D

Linear Algebra Concepts and Notation

In this appendix, we review some basic concepts and standard notation used in linear algebra.

D.1 Elementary Concepts

The two concepts from linear algebra that we are most concerned with are *vectors* and *matrices* because of their usefulness in compressing complicated expressions into a compact notation. The vectors and matrices in this text are most often *real*, since they consist of real numbers. These concepts easily generalize to *complex* vectors and matrices.

Vectors

A **vector** $\boldsymbol{x} \in \mathbb{R}^n$ can be thought of as a one-dimensional array of numbers and is written as

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where x_i is called the ith **element**, **entry**, or **component**. An alternative notation that is useful in pseudocodes is $\boldsymbol{x} = (x_i)_n$. Sometimes the vector $\boldsymbol{x}$ displayed above is said to be a **column vector** to distinguish it from a **row vector** $\boldsymbol{y}$ written as

$$\boldsymbol{y} = [y_1, y_2, \ldots, y_n]$$

For example, here are some vectors:

$$\begin{bmatrix} \frac{1}{5} \\ 3 \\ -\frac{5}{6} \\ \frac{2}{7} \end{bmatrix} \qquad [\pi, \quad e, \quad 5, \quad -4] \qquad \begin{bmatrix} \frac{1}{2} \\ \frac{1}{3} \end{bmatrix}$$

To save space, a column vector $\boldsymbol{x}$ can be written as a row vector such as

$$\boldsymbol{x} = [x_1, x_2, \ldots, x_n]^T \qquad \text{or} \qquad \boldsymbol{x}^T = [x_1, x_2, \ldots, x_n]$$

by adding a T (for **transpose**) to indicate that we are interchanging or transposing a row or column vector. As an example, we have

$$[1 \quad 2 \quad 3 \quad 4]^T = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Many operations involving vectors are component-by-component operations. For vectors $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

the following definitions apply.

Equality $\boldsymbol{x} = \boldsymbol{y}$ if and only if $x_i = y_i$ for all $i\,(1 \leqq i \leqq n)$

Inequality $\boldsymbol{x} < \boldsymbol{y}$ if and only if $x_i < y_i$ for all $i\,(1 \leqq i \leqq n)$

Addition/Subtraction

$$\boldsymbol{x} \pm \boldsymbol{y} = \begin{bmatrix} x_1 \pm y_1 \\ x_2 \pm y_2 \\ \vdots \\ x_n \pm y_n \end{bmatrix}$$

Scalar Product

$$\alpha\boldsymbol{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix} \qquad \text{for } \alpha \text{ a constant or scalar}$$

Here is an example:

$$\begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} = 2 \begin{bmatrix} 0 \\ 2 \\ 0 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 \\ 0 \\ 6 \\ 0 \end{bmatrix}$$

For m vectors $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots, \boldsymbol{x}^{(m)}$ and m scalars $\alpha_1, \alpha_2, \ldots, \alpha_m$, we define a **linear combination** as

$$\sum_{i=1}^{m} \alpha_i \boldsymbol{x}^{(i)} = \alpha_1 \boldsymbol{x}^{(1)} + \alpha_2 \boldsymbol{x}^{(2)} + \cdots + \alpha_m \boldsymbol{x}^{(m)} = \begin{bmatrix} \sum_{i=1}^{m} \alpha_i x_1^{(i)} \\ \sum_{i=1}^{m} \alpha_i x_2^{(i)} \\ \vdots \\ \sum_{i=1}^{m} \alpha_i x_n^{(i)} \end{bmatrix}$$

Special vectors are the standard **unit vectors:**

$$e^{(1)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad e^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \cdots \qquad e^{(n)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

Clearly,

$$\sum_{i=1}^{n} \alpha_i e^{(i)} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Hence, any vector x can be written as a linear combination of the standard unit vectors

$$x = x_1 e^{(1)} + x_2 e^{(2)} + \cdots + x_n e^{(n)} = \sum_{i=1}^{n} x_i e^{(i)}$$

As an example, notice that

$$\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} + 4 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

The **dot product** or **inner product** of vectors x and y is the number

$$x^T y = [x_1, x_2, \ldots, x_n] \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

As an example, we see that

$$[1, 1, 1, 1] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = 4$$

Matrices

A **matrix** is a two-dimensional array of numbers that can be written as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

where a_{ij} is called the **element** or **entry** in the ith row and jth column. An alternative notation is $A = (a_{ij})_{n \times m}$. A column vector is also an $n \times 1$ matrix and a row vector is also a $1 \times m$ matrix. For example, here are three matrices:

$$\begin{bmatrix} \frac{1}{5} & \frac{2}{7} & -1 \\ 3 & 2 & \frac{1}{8} \\ -\frac{5}{6} & \frac{2}{5} & 3 \end{bmatrix} \qquad \begin{bmatrix} 1 & 6 & \frac{9}{8} & -5 \end{bmatrix} \qquad \begin{bmatrix} \frac{11}{2} & \frac{4}{9} \\ \frac{2}{3} & -\frac{7}{8} \\ \pi & e \\ \frac{1}{\pi} & \frac{1}{e} \end{bmatrix}$$

The entries in $\boldsymbol{A}$ can be grouped into column vectors:

$$\boldsymbol{A} = \left[\begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix} \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix} \cdots \begin{bmatrix} a_{1m} \\ a_{2m} \\ \vdots \\ a_{nm} \end{bmatrix} \right] = \left[\boldsymbol{a}^{(1)} \quad \boldsymbol{a}^{(2)} \quad \cdots \quad \boldsymbol{a}^{(m)} \right]$$

where $\boldsymbol{a}^{(j)}$ is the jth column vector. Also, $\boldsymbol{A}$ can be grouped into row vectors:

$$\boldsymbol{A} = \begin{bmatrix} [a_{11} & a_{12} & \cdots & a_{1m}] \\ [a_{21} & a_{22} & \cdots & a_{2m}] \\ & & \vdots & \\ [a_{n1} & a_{n2} & \cdots & a_{nm}] \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^{(1)} \\ \boldsymbol{A}^{(2)} \\ \vdots \\ \boldsymbol{A}^{(n)} \end{bmatrix}$$

where $\boldsymbol{A}^{(i)}$ is the ith row vector. Notice that

$$\begin{bmatrix} 1 & 5 & 9 & 13 \\ 2 & 6 & 10 & 14 \\ 3 & 7 & 11 & 15 \\ 4 & 8 & 12 & 16 \end{bmatrix} = \left[\begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} \begin{bmatrix} 9 \\ 10 \\ 11 \\ 12 \end{bmatrix} \begin{bmatrix} 13 \\ 14 \\ 15 \\ 16 \end{bmatrix} \right] = \begin{bmatrix} [1 & 5 & 9 & 13] \\ [2 & 6 & 10 & 14] \\ [3 & 7 & 11 & 15] \\ [4 & 8 & 12 & 16] \end{bmatrix}$$

An $n \times n$ matrix of special importance is the **identity** matrix, denoted by $\boldsymbol{I}$, composed of all 0's except that the main diagonal consists of 1's:

$$\boldsymbol{I} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \left[\boldsymbol{e}^{(1)} \quad \boldsymbol{e}^{(2)} \quad \cdots \quad \boldsymbol{e}^{(n)} \right]$$

A matrix of this same general form with entries d_i on the main diagonal is called a **diagonal** matrix and is written as

$$\boldsymbol{D} = \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix} = \operatorname{diag}(d_1, d_2, \ldots, d_n)$$

where the blank space indicates 0 entries. A **tridiagonal** matrix is a square matrix of the form

$$\boldsymbol{T} = \begin{bmatrix} d_1 & c_1 & & & & \\ a_1 & d_2 & c_2 & & & \\ & a_2 & d_3 & c_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & a_{n-2} & d_{n-1} & c_{n-1} \\ & & & & a_{n-1} & d_n \end{bmatrix}$$

where the diagonal entries $\{a_i\}$, $\{d_i\}$, and $\{c_i\}$ are called the **subdiagonal**, **main diagonal**, and **superdiagonal**, respectively.

For the general $n \times n$ matrix $\boldsymbol{A} = (a_{ij})$, $\boldsymbol{A}$ is a diagonal matrix if $a_{ij} = 0$ whenever $i \neq j$, and $\boldsymbol{A}$ is a tridiagonal matrix if $a_{ij} = 0$ whenever $|i - j| \geqq 2$. The matrix $\boldsymbol{A}$ is a **lower triangular matrix** whenever $a_{ij} = 0$ for all $i < j$ and is an **upper triangular matrix** whenever $a_{ij} = 0$ for all $i > j$. Examples of identity, diagonal, tridiagonal, lower triangular, and upper triangular matrices, respectively, are as follows:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix} \qquad \begin{bmatrix} 5 & 3 & 0 & 0 & 0 \\ 2 & 5 & 3 & 0 & 0 \\ 0 & 2 & 9 & 2 & 0 \\ 0 & 0 & 3 & 7 & 2 \\ 0 & 0 & 0 & 3 & 7 \end{bmatrix}$$

$$\begin{bmatrix} 6 & 0 & 0 & 0 \\ 3 & 6 & 0 & 0 \\ 4 & -2 & 7 & 0 \\ 5 & -3 & 9 & 21 \end{bmatrix} \qquad \begin{bmatrix} 1 & -1 & 2 & 1 \\ 0 & 5 & -5 & 1 \\ 0 & 0 & 9 & -3 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

As with vectors, many operations involving matrices correspond to component operations. For matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \qquad \boldsymbol{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nm} \end{bmatrix}$$

the following definitions apply:

Equality $\boldsymbol{A} = \boldsymbol{B}$ if and only if $a_{ij} = b_{ij}$ for all $i\,(1 \leqq i \leqq n)$ and all $j\,(1 \leqq j \leqq m)$

Inequality $\boldsymbol{A} < \boldsymbol{B}$ if and only if $a_{ij} < b_{ij}$ for all $i\,(1 \leqq i \leqq n)$ and all $j\,(1 \leqq j \leqq m)$

Addition/Subtraction

$$\boldsymbol{A} \pm \boldsymbol{B} = \begin{bmatrix} a_{11} \pm b_{11} & a_{12} \pm b_{12} & \cdots & a_{1m} \pm b_{1m} \\ a_{21} \pm b_{21} & a_{22} \pm b_{22} & \cdots & a_{2m} \pm b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} \pm b_{n1} & a_{n2} \pm b_{n2} & \cdots & a_{nm} \pm b_{nm} \end{bmatrix}$$

Scalar Product

$$\alpha\boldsymbol{A} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \cdots & \alpha a_{1m} \\ \alpha a_{21} & \alpha a_{22} & \cdots & \alpha a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha a_{n1} & \alpha a_{n2} & \cdots & \alpha a_{nm} \end{bmatrix} \quad \text{for } \alpha \text{ a constant}$$

As an example, we have

$$\begin{bmatrix} \frac{1}{5} & \frac{7}{5} & -1 \\ -3 & 2 & -8 \\ \frac{6}{5} & \frac{2}{5} & -3 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} 1 & 7 & 0 \\ 0 & 10 & 0 \\ 6 & 2 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 1 \\ 3 & 0 & 8 \\ 0 & 0 & 3 \end{bmatrix}$$

Matrix-Vector Product

The product of an $n \times m$ matrix $\boldsymbol{A}$ and an $m \times 1$ vector $\boldsymbol{b}$ is of special interest. Considering the matrix $\boldsymbol{A}$ in terms of its columns, we have

$$\begin{aligned}\boldsymbol{Ab} &= \begin{bmatrix}\boldsymbol{a}^{(1)} & \boldsymbol{a}^{(2)} & \cdots & \boldsymbol{a}^{(m)}\end{bmatrix}\begin{bmatrix}b_1 \\ b_2 \\ \vdots \\ b_m\end{bmatrix} \\ &= b_1\boldsymbol{a}^{(1)} + b_2\boldsymbol{a}^{(2)} + \cdots + b_m\boldsymbol{a}^{(m)} \\ &= \sum_{i=1}^{m} b_i\boldsymbol{a}^{(i)}\end{aligned}$$

Thus, $\boldsymbol{Ab}$ is a vector and can be thought of as a linear combination of the columns of $\boldsymbol{A}$ with coefficients the entries of $\boldsymbol{b}$. Considering matrix $\boldsymbol{A}$ in terms of its rows, we have

$$\boldsymbol{Ab} = \begin{bmatrix}\boldsymbol{A}^{(1)} \\ \boldsymbol{A}^{(2)} \\ \vdots \\ \boldsymbol{A}^{(n)}\end{bmatrix}\boldsymbol{b} = \begin{bmatrix}\boldsymbol{A}^{(1)}\boldsymbol{b} \\ \boldsymbol{A}^{(2)}\boldsymbol{b} \\ \vdots \\ \boldsymbol{A}^{(n)}\boldsymbol{b}\end{bmatrix}$$

Thus, the jth element of $\boldsymbol{Ab}$ can be viewed as the dot product of the jth row of $\boldsymbol{A}$ and the vector $\boldsymbol{b}$.

Matrix Product

The product of the matrix $\boldsymbol{A} = (a_{ij})_{n\times m}$ and the matrix $\boldsymbol{B} = (b_{ij})_{m\times r}$ is the matrix $\boldsymbol{C} = (c_{ij})_{n\times r}$ such that

$$\boldsymbol{AB} = \boldsymbol{C}$$

where

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj} = \sum_{k=1}^{m} a_{ik}b_{kj} \qquad (1 \leqq i \leqq n,\ 1 \leqq j \leqq r)$$

The element c_{ij} is the dot product of the ith row vector of $\boldsymbol{A}$

$$\boldsymbol{A}^{(i)} = [a_{i1}, a_{i2}, \ldots, a_{im}]$$

and the jth column vector of $\boldsymbol{B}$

$$\boldsymbol{b}^{(j)} = \begin{bmatrix}b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj}\end{bmatrix}$$

that is,

$$c_{ij} = \boldsymbol{A}^{(i)}\boldsymbol{b}^{(j)}$$

Similarly, the matrix product $\boldsymbol{AB}$ can be thought of in two different ways. We can write either

$$\begin{aligned}\boldsymbol{AB} &= \boldsymbol{A}\begin{bmatrix}\boldsymbol{b}^{(1)} & \boldsymbol{b}^{(2)} & \cdots & \boldsymbol{b}^{(r)}\end{bmatrix} \\ &= \begin{bmatrix}\boldsymbol{Ab}^{(1)} & \boldsymbol{Ab}^{(2)} & \cdots & \boldsymbol{Ab}^{(r)}\end{bmatrix} \\ &= \boldsymbol{C}\end{aligned} \tag{1}$$

or

$$\boldsymbol{AB} = \begin{bmatrix}\boldsymbol{A}^{(1)} \\ \boldsymbol{A}^{(2)} \\ \vdots \\ \boldsymbol{A}^{(n)}\end{bmatrix}\boldsymbol{B} = \begin{bmatrix}\boldsymbol{A}^{(1)}\boldsymbol{B} \\ \boldsymbol{A}^{(2)}\boldsymbol{B} \\ \vdots \\ \boldsymbol{A}^{(n)}\boldsymbol{B}\end{bmatrix} = \boldsymbol{C} \tag{2}$$

Equation (1) implies that the jth column of $\boldsymbol{C} = \boldsymbol{AB}$ is

$$\boldsymbol{c}^{(j)} = \boldsymbol{Ab}^{(j)}$$

That is, each column of $\boldsymbol{C}$ is the result of *postmultiplying* $\boldsymbol{A}$ by the jth column of $\boldsymbol{B}$. In other words, each column of $\boldsymbol{C}$ can be obtained by taking inner products of a column of $\boldsymbol{B}$ with all rows of $\boldsymbol{A}$:

$$\boldsymbol{c}^{(j)} = \boldsymbol{Ab}^{(j)} = \begin{bmatrix}\longleftarrow \\ \longleftarrow \\ \vdots \\ \longleftarrow\end{bmatrix}\begin{bmatrix}b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj}\end{bmatrix} = \begin{bmatrix}c_{1j} \\ c_{2j} \\ \vdots \\ c_{nj}\end{bmatrix}$$

The long left-arrow means an inner product is formed across the elements in the row—that is, $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$. Equation (2) implies that the ith row of the result $\boldsymbol{C}$ of multiplying $\boldsymbol{A}$ times $\boldsymbol{B}$ is

$$\boldsymbol{C}^{(i)} = \boldsymbol{A}^{(i)}\boldsymbol{B}$$

That is, each row of $\boldsymbol{C}$ is the result of *premultiplying* $\boldsymbol{B}$ by the ith row of $\boldsymbol{A}$. In other words, each row of $\boldsymbol{C}$ can be obtained by taking inner products of a row of $\boldsymbol{A}$ with all columns of $\boldsymbol{B}$:

$$\begin{aligned}\boldsymbol{C}^{(i)} = \boldsymbol{A}^{(i)}\boldsymbol{B} &= [a_{i1} \quad a_{i2} \quad \cdots \quad a_{im}]\begin{bmatrix}\Big\uparrow & \Big\uparrow & \cdots & \Big\uparrow\end{bmatrix} \\ &= [c_{i1} \quad c_{i2} \quad \cdots \quad c_{ir}]\end{aligned}$$

The long up-arrow means an inner product is formed from the elements in the column—that is, $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$.

As an example, we can determine the matrix product columnwise as

$$\begin{bmatrix}3 & 1 & 7 \\ 2 & 4 & -5 \\ 1 & -3 & 2\end{bmatrix}\begin{bmatrix}-1 & -3 & 2 \\ 1 & 1 & 1 \\ -3 & -2 & 1\end{bmatrix} = \begin{bmatrix}\boldsymbol{c}^{(1)} & \boldsymbol{c}^{(2)} & \boldsymbol{c}^{(3)}\end{bmatrix}$$

where

$$c^{(1)} = \begin{bmatrix} 3 & 1 & 7 \\ 2 & 4 & -5 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -23 \\ 17 \\ -10 \end{bmatrix}$$

$$c^{(2)} = \begin{bmatrix} 3 & 1 & 7 \\ 2 & 4 & -5 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -22 \\ 8 \\ -10 \end{bmatrix}$$

$$c^{(3)} = \begin{bmatrix} 3 & 1 & 7 \\ 2 & 4 & -5 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 14 \\ 3 \\ 1 \end{bmatrix}$$

or we can determine it rowwise as

$$\begin{bmatrix} 3 & 1 & 7 \\ 2 & 4 & -5 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} -1 & -3 & 2 \\ 1 & 1 & 1 \\ -3 & -2 & 1 \end{bmatrix} = \begin{bmatrix} C^{(1)} \\ C^{(2)} \\ C^{(3)} \end{bmatrix}$$

where

$$C^{(1)} = [3 \quad 1 \quad 7] \begin{bmatrix} -1 & -3 & 2 \\ 1 & 1 & 1 \\ -3 & -2 & 1 \end{bmatrix} = [-23 \quad -22 \quad 14]$$

$$C^{(2)} = [2 \quad 4 \quad -5] \begin{bmatrix} -1 & -3 & 2 \\ 1 & 1 & 1 \\ -3 & -2 & 1 \end{bmatrix} = [17 \quad 8 \quad 3]$$

$$C^{(3)} = [1 \quad -3 \quad 2] \begin{bmatrix} -1 & -3 & 2 \\ 1 & 1 & 1 \\ -3 & -2 & 1 \end{bmatrix} = [-10 \quad -10 \quad 1]$$

Other Concepts

The **transpose** of the $n \times m$ matrix A, denoted A^T, is obtained by interchanging the rows and columns of $A = (a_{ij})_{n \times m}$:

$$A^T = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(n)} \end{bmatrix}^T = \begin{bmatrix} A^{(1)^T} & A^{(2)^T} & \cdots & A^{(n)^T} \end{bmatrix}$$

or

$$A^T = \begin{bmatrix} a^{(1)} & a^{(2)} & \cdots & a^{(m)} \end{bmatrix}^T = \begin{bmatrix} a^{(1)^T} \\ a^{(2)^T} \\ \vdots \\ a^{(m)^T} \end{bmatrix}$$

Hence, A^T is the $m \times n$ matrix:

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{bmatrix} = (a_{ji})_{m \times n}$$

As an example, we have

$$\begin{bmatrix} 2 & 4 & 9 \\ 5 & 7 & 3 \\ 10 & 6 & 2 \end{bmatrix}^T = \begin{bmatrix} 2 & 5 & 10 \\ 4 & 7 & 6 \\ 9 & 3 & 2 \end{bmatrix}$$

An $n \times n$ matrix A is **symmetric** if $a_{ij} = a_{ji}$ for all i $(1 \leqq i \leqq n)$ and all j $(1 \leqq j \leqq n)$. In other words, A is symmetric if $A = A^T$.

Some useful properties for matrices of compatible sizes are as follows:

PROPERTIES ***Elementary Consequences of the Definitions***

1. $AB \neq BA$ (in general)
2. $AI = IA = A$
3. $A0 = 0A = 0$
4. $(A^T)^T = A$
5. $(A + B)^T = A^T + B^T$
6. $(AB)^T = B^T A^T$

If A and B are square matrices that satisfy $AB = BA = I$, then B is said to be the **inverse** of A, which is denoted A^{-1}.

To illustrate Property 1, form the following products, and observe that matrix multiplication is not commutative:

$$\begin{bmatrix} 3 & 1 & 7 \\ 2 & 4 & -5 \\ 1 & -3 & 2 \end{bmatrix} \begin{bmatrix} -1 & -3 & 2 \\ 1 & 1 & 1 \\ -3 & -2 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -3 & 2 \\ 1 & 1 & 1 \\ -3 & -2 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 7 \\ 2 & 4 & -5 \\ 1 & -3 & 2 \end{bmatrix}$$

Also, verify that $AA^{-1} = A^{-1}A = I$ for

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

and

$$A^{-1} = \begin{bmatrix} -1 & 0 & 1 \\ -5 & 1 & 3 \\ 7 & -1 & -4 \end{bmatrix}$$

As our final set of examples, we have the product of a matrix times a vector and of two matrices:

$$\begin{bmatrix} 3 & 2 & -1 \\ 5 & 3 & 2 \\ -1 & 1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3x_1 & + & 2x_2 & - & x_3 \\ 5x_1 & + & 3x_2 & + & 2x_3 \\ -x_1 & + & x_2 & - & 3x_3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ -\frac{5}{3} & 1 & 0 \\ -8 & 5 & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & -1 \\ 5 & 3 & 2 \\ -1 & 1 & -3 \end{bmatrix} = \begin{bmatrix} 3 & 2 & -1 \\ 0 & -\frac{1}{3} & \frac{11}{3} \\ 0 & 0 & 15 \end{bmatrix}$$

The reader should verify them and note how they relate to solving the following problem using naive Gaussian elimination (see Section 7.1):

$$\begin{cases} 3x_1 + 2x_2 - x_3 = 7 \\ 5x_1 + 3x_2 + 2x_3 = 4 \\ -x_1 + x_2 - 3x_3 = -1 \end{cases}$$

As well, compute the products shown and relate them to this problem:

$$\begin{bmatrix} 1 & 0 & 0 \\ -\frac{5}{8} & 1 & 0 \\ -8 & 5 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 4 \\ -1 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{3} & 2 & -\frac{7}{15} \\ 0 & -3 & \frac{11}{15} \\ 0 & 0 & \frac{1}{15} \end{bmatrix} \begin{bmatrix} 3 & 2 & -1 \\ 0 & -\frac{1}{3} & \frac{11}{3} \\ 0 & 0 & 15 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{3} & 2 & -\frac{7}{15} \\ 0 & -3 & \frac{11}{15} \\ 0 & 0 & \frac{1}{15} \end{bmatrix} \begin{bmatrix} 7 \\ -\frac{23}{3} \\ -37 \end{bmatrix}$$

Cramer's Rule

The solution of a 2×2 linear system of the form

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix}$$

is given by

$$x = \frac{1}{D} \text{Det} \begin{bmatrix} f & c \\ g & d \end{bmatrix} = \frac{1}{D}(fd - gc)$$

$$y = \frac{1}{D} \text{Det} \begin{bmatrix} a & f \\ b & g \end{bmatrix} = \frac{1}{D}(ag - bf)$$

where

$$D = \text{Det} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = ad - bc \neq 0$$

D.2 Abstract Vector Spaces

The vectors that have been considered so far in this appendix are members of a particular vector space $\mathbb{R}^n$. There is a general concept of an abstract vector space that will include $\mathbb{R}^n$ as a particular case. An abstract vector space (a linear space) is a quadruple $(X, F, +, *)$, where X is a set of elements called **vectors**, F is a **field**, $+$ is an operation, and $*$ is an operation. There are ten axioms to be satisfied, and all of them are familiar to any reader who has worked with the special case of $\mathbb{R}^n$. First, let's fix the field to be $\mathbb{R}$. (The other field that is often needed is $\mathbb{C}$, but fields other than these two are rarely used in this situation.)

■ THEOREM 1 **AXIOMS FOR A VECTOR SPACE**

1. If $\boldsymbol{x}$ and $\boldsymbol{y}$ belong to X, then $\boldsymbol{x} + \boldsymbol{y}$ also belongs to X.
2. For $\boldsymbol{x}$ and $\boldsymbol{y}$ in X, $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x}$.
3. For $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{z}$ in X, $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z})$.
4. The set X contains a special element $\mathbf{0}$ such that $\boldsymbol{x} + \mathbf{0} = \boldsymbol{x}$ for all $\boldsymbol{x}$ in X.
5. For each $\boldsymbol{x}$, there is an element $-\boldsymbol{x}$ with the property that $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$.
6. If $a \in \mathbb{R}$, then for each $\boldsymbol{x}$ in X, $a\boldsymbol{x} \in X$. ($a\boldsymbol{x}$ means $a * \boldsymbol{x}$.)
7. If $a \in \mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in X$, then $a(\boldsymbol{x} + \boldsymbol{y}) = a\boldsymbol{x} + a\boldsymbol{y}$.
8. If $a, b \in \mathbb{R}$ and $\boldsymbol{x} \in X$, then $(a + b)\boldsymbol{x} = a\boldsymbol{x} + b\boldsymbol{x}$.
9. If $a, b \in \mathbb{R}$ and $\boldsymbol{x} \in X$, then $a(b\boldsymbol{x}) = (ab)\boldsymbol{x}$.
10. For $\boldsymbol{x} \in X$, $1\boldsymbol{x} = \boldsymbol{x}$.

From these axioms, one can prove many additional properties, such as the following:

■ PROPERTIES ***Immediate Consequences of the Axioms***

1. The zero element, $\mathbf{0}$, of X is unique.
2. $0\boldsymbol{x} = \mathbf{0}$ and $a\mathbf{0} = \mathbf{0}$ for $a \in \mathbb{R}$. (Notice the different zeros here!)
3. For each $\boldsymbol{x}$ in X, the element $-\boldsymbol{x}$ in Axiom **5** is unique.
4. For each $\boldsymbol{x}$ in X, $(-1)\boldsymbol{x} = -\boldsymbol{x}$.
5. If $a\boldsymbol{x} = \mathbf{0}$ and $a \neq 0$, then $\boldsymbol{x} = \mathbf{0}$.

A good example of a vector space (other than $\mathbb{R}^n$) is the set of all polynomials. We know that the sum of two polynomials is another polynomial and that a scalar multiple of a polynomial is a polynomial. All other axioms for a vector space are quickly verified. The zero element is the polynomial that we define by the equation $\mathbf{0}(t) = 0$ for all real values of t.

Subspaces

If U is a subset of the vector space X and if U is a vector space also (with the same definitions of $+$ and $*$ as used in X), then we call U a subspace of X. In checking to determine whether a given subset U is a subspace, one need only verify Axioms 1 and 6. Indeed, once that has been done, Axiom 6 and Property 4 above yield $-\boldsymbol{u} \in U$ when $\boldsymbol{u} \in U$. Then Axiom 1 yields $\mathbf{0} = \boldsymbol{u} + (-\boldsymbol{u}) \in \mathbf{U}$. The remaining axioms are true for U simply because $U \subset X$.

Linear Independence

A finite ordered set of points $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ in a vector space is said to be **linearly dependent** if there is a nontrivial equation of the form

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

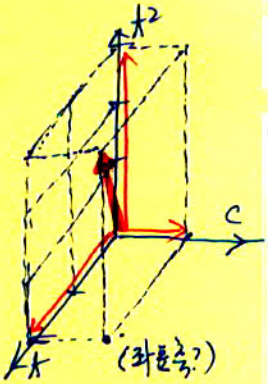
The term *nontrivial* means that not all the coefficients a_i are zero. For example, if $n = 3$ and $\boldsymbol{x}_1 = 3\boldsymbol{x}_2 - \boldsymbol{x}_3$, then the ordered set $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ is linearly dependent. If $n = 3$ and $\boldsymbol{x}_3 = \boldsymbol{x}_1$, which is permitted in an ordered set, then $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ is linearly dependent. Note that if these were interpreted as plain sets, we would have $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_1\} = \{\boldsymbol{x}_1, \boldsymbol{x}_2\}$, because in describing a plain set the repeated entry can be dropped without changing the set! This explains the necessity of dealing with ordered sets or indexed sets in defining linear dependence. (The difficulty arises only for indexed sets in which two elements are equal but bear different indices.) A finite set consisting of n (distinct) elements $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ is **linearly independent** if the equation

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

is true only when all the coefficients a_i are zero. An arbitrary set, possibly infinite, is linearly independent if every finite subset of that set is linearly independent.

To illustrate linear independence, consider the three polynomials $\boldsymbol{p}_1(t) = t^3 - 2t$, $\boldsymbol{p}_2(t) = t^2 + 4$, and $\boldsymbol{p}_3(t) = 2t^2 + t$. Is the set $\{\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3\}$ linearly independent? To find out, suppose that $a_1\boldsymbol{p}_1 + a_2\boldsymbol{p}_2 + a_3\boldsymbol{p}_3 = 0$. Then for all t,

$$a_1\left(t^3 - 2t\right) + a_2\left(t^2 + 4\right) + a_3\left(2t^2 + t\right) = 0$$



Collecting terms, we have

$$a_3t^3 + (a_2 + 2a_3)t^2 + (-2a_1 + a_3)t + 4a_2 = 0 \qquad (t \in \mathbb{R})$$

Since a cubic polynomial can have at most three roots (if it is not zero), the coefficients of each power of t in the preceding equation must be zero:

$$a_3 = a_2 + 2a_3 = -2a_1 + a_3 = 4a_2 = 0$$

$f_1(t) = 1$
$f_2(t) = 2t$
$f_3(t) = 3t^2$
$f_4(t) = 1 + 2t + 3t^2$

Hence, all a_i must be zero. The set is linearly independent.

THEOREM 2 **THEOREM ON LINEAR DEPENDENCE**

A finite, ordered, set $\{x_1, x_2, \ldots, x_n\}$, with $n \geqq 2$, is linearly dependent if and only if some element of the set, say, x_k, is a linear combination of its predecessors in the set:

$$x_k = \sum_{i=1}^{k-1} a_i x_i$$

Bases

A **basis** for a vector space is a **maximal linearly independent** set in the vector space. **Maximal** means that no vector can be added to the set without spoiling the linear independence. For example, a basis for the space of all polynomials is given by the functions $u_i(t) = t^i$ for $i = 0, 1, 2, \ldots$. To see that this is a maximal linearly independent set, suppose we add to the set any polynomial p. Let the degree of p be n. Then the set $\{u_0, u_1, \ldots, u_n, p\}$ is linearly dependent. Indeed, one element (namely, p) is a linear combination of its predecessors in the set, and the above theorem applies.

If a vector space X has a finite basis, $\{u_1, u_2, \ldots, u_n\}$, then every basis for X has n elements. This number is called the **dimension** of X, and we say that X is **finite-dimensional**. Each x in X has a unique representation $x = \sum_{i=1}^{n} a_i u_i$. The existence of this representation is a consequence of the maximality, and the uniqueness is a consequence of the linear independence of the basis.

Linear Transformations

If X and Y are vector spaces and if L is a mapping of X into Y such that

$$\boxed{L(au + bv) = aL(u) + bL(v)}$$

for all scalars a and b and for all vectors u and v in X, then we say that L is **linear**. Many familiar operations that are studied in mathematics are linear. For example, differentiation is a linear operator:

$$(f + g)' = f' + g' \qquad (af)' = af'$$

The Laplace transform is linear, and so is the map $f \mapsto \int_1^b f(t)\,dt$.

If the space X is finite-dimensional and if we select a basis $\{u_1, u_2, \ldots, u_n\}$ for X, then a linear map $L : X \to Y$ is completely known if the n vectors $Lu_1, Lu_2, \ldots, Lu_n$ are known. Indeed, any vector x in X is representable in terms of the basis, $x = \sum_{j=1}^{n} c_j u_j$, and from this, we get $Lx = \sum_{j=1}^{n} c_j Lu_j$. Going further, suppose that Y is also finite-dimensional. Select a basis for Y, say, $\{v_1, v_2, \ldots, v_m\}$. Then each image Lu_j is expressible in terms of the basis selected for Y, and we have, for suitable coefficients a_{ij},

$$Lu_j = \sum_{i=1}^{m} a_{ij} v_i$$

From this, it follows that

$$Lx = L\left(\sum_{j=1}^{n} c_j u_j\right) = \sum_{j=1}^{n} c_j Lu_j = \sum_{j=1}^{n} c_j \sum_{i=1}^{m} a_{ij} v_i$$

In this way, a matrix $A = (a_{ij})$ is associated with L, but only after the choice of bases in X and Y has been made.

The special case in which $Y = X$ and the same basis is used in both roles leads to these equations:

$$x = \sum_{j=1}^{n} c_j u_j$$

$$Lu_j = \sum_{i=1}^{n} a_{ij} u_i$$

$$Lx = \sum_{j=1}^{n} c_j \sum_{i=1}^{n} a_{ij} u_i$$

Eigenvalues and Eigenvectors

Let A be an $n \times n$ matrix. If x is a nonzero vector with the property that Ax is a scalar multiple of x, then we call x an **eigenvector** of A. When this occurs, the equation

$$\boxed{Ax = \lambda x}$$

고유값

is satisfied for some scalar λ (which may be zero). The scalar λ is then called an **eigenvalue** of A. Since we have a nonzero solution of the equation $Ax - \lambda x = 0$, the matrix $A - \lambda I$ must be singular. Hence, its determinant is zero. The equation

$$\mathrm{Det}(A - \lambda I) = 0$$

is called the **characteristic equation** of A. As a function of λ, the left side of this equation is a polynomial of degree n, which has exactly n roots if we count each root with its multiplicity.

Change of Basis and Similarity

If L is a linear transformation taking an n-dimensional vector space into itself, then, having selected a basis $\{u_1, u_2, \ldots, u_n\}$, we can assign a matrix A to L. Thus, we have

$$Lu_j = \sum_{i=1}^{n} A_{ij} u_i$$

If another basis for X is chosen, say, $\{v_1, v_2, \ldots, v_n\}$, then another matrix, B, arises in the same way, and we have

$$Lv_j = \sum_{i=1}^{n} B_{ij} v_i$$

What is the relationship between A and B? Define the matrix P by the equation

$$u_k = \sum_{i=1}^{n} P_{ik} v_i \qquad 1 \leqq k \leqq n$$

Then

$$B = PAP^{-1}$$

To prove this, we must establish that

$$\boldsymbol{L}\boldsymbol{v}_j = \sum_{i=1}^{n} (\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1})_{ij}\boldsymbol{v}_i$$

The equations already recorded above justify the steps in the following calculation:

$$\begin{aligned}
\sum_{i=1}^{n} (\boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1})_{ij}\boldsymbol{v}_i &= \sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{r=1}^{n} \boldsymbol{P}_{ik}\boldsymbol{A}_{kr}\boldsymbol{P}^{-1}_{rj}\boldsymbol{v}_i \\
&= \sum_{k=1}^{n}\sum_{r=1}^{n} \boldsymbol{A}_{kr}\boldsymbol{P}^{-1}_{rj}\boldsymbol{u}_k \\
&= \sum_{r=1}^{n} \boldsymbol{P}^{-1}_{rj}\boldsymbol{L}\boldsymbol{u}_r \\
&= \boldsymbol{L}\left(\sum_{r=1}^{n} \boldsymbol{P}^{-1}_{rj}\boldsymbol{u}_r\right) \\
&= \boldsymbol{L}\left(\sum_{r=1}^{n}\sum_{i=1}^{n} \boldsymbol{P}^{-1}_{rj}\boldsymbol{P}_{ir}\boldsymbol{v}_i\right) \\
&= \boldsymbol{L}\left(\sum_{i=1}^{n} \boldsymbol{I}_{ij}\boldsymbol{v}_i\right) = \boldsymbol{L}\boldsymbol{v}_j
\end{aligned}$$

Orthogonal Matrices and Spectral Theorem

A matrix $\boldsymbol{Q}$ is said to be **orthogonal** if

$$\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I} \quad \Rightarrow Q^T = Q^{-1}$$

This forces $\boldsymbol{Q}$ to be square and nonsingular. Furthermore,

$$\boldsymbol{Q}^{-1} = \boldsymbol{Q}^T$$

With this concept available, we can state one of the principal theorems of linear algebra: the spectral theorem for symmetric matrices.

THEOREM 3 **SPECTRAL THEOREM FOR SYMMETRIC MATRICES**

If $\boldsymbol{A}$ is a symmetric real matrix, then there exists an orthogonal matrix $\boldsymbol{Q}$ such that $\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q}$ is a diagonal matrix.

* columns are orthogonal vectors to each other.

The equation

$$\boldsymbol{Q}^T\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{D}$$

is equivalent to

$$\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{D}$$

If $\boldsymbol{D}$ is diagonal, the columns $\boldsymbol{v}_i$ of $\boldsymbol{Q}$ obey the equation

$$\boldsymbol{A}\boldsymbol{v}_i = d_{ii}\boldsymbol{v}_i$$

In other words, the columns of Q form an orthonormal system of eigenvectors of A, and the diagonal elements of D are the eigenvalues of A.

Norms

A **vector norm** on a vector space X is a real-valued function on X, written $x \mapsto \|x\|$ and having these three properties:

PROPERTIES *Properties of Vector Norms*

1. $\|x\| > 0$ for all nonzero vectors x.
2. $\|ax\| = |a|\|x\|$ for all vectors x and all scalars a.
3. $\|x + y\| \leqq \|x\| + \|y\|$ for all vectors x and y.

On $\mathbb{R}^n$, the simplest vector norms are

$$\|x\|_1 = |x_1| + |x_2| + \cdots + |x_n| \qquad (\ell_1\text{-vector norm})$$

$$\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \qquad (\text{Euclidean}/\ell_2\text{-vector norm})$$

$$\|x\|_\infty = \max\{|x_1|, |x_2|, \dots, |x_n|\} \qquad (\ell_\infty\text{-vector norm})$$

Here, x_i denotes the ith component of the vector. Any norm can be thought of as assigning a *length* to each vector. It is the Euclidean norm that corresponds directly to our usual concept of length, but other norms are sometimes much more convenient for our purposes. For example, if we know that $\|x - y\|_\infty < 10^{-8}$, then we know that each component of x differs from the corresponding component of y by at most 10^{-8} and that the converse is also true. When we solve a system of linear equations $Ax = b$ numerically, we shall want to know (among other things) how big the residual vector is. That is conveniently measured by $\|Ax - b\|$, where some norm has been specified.

For $n \times n$ matrices, we can also have matrix norms, subject to the following requirements:

PROPERTIES *Properties of Matrix Norms*

1. $\|A\| > 0$ if $A \neq 0$
2. $\|\alpha A\| = |\alpha|\|A\|$
3. $\|A + B\| \leqq \|A\| + \|B\|$ **(triangular inequality)**

for matrices A, B and scalars α.

We usually prefer matrix norms that are related to a vector norm. When a vector norm has been specified on $\mathbb{R}^n$, there is a standard way of introducing a related **matrix norm** for $n \times n$ matrices, namely,

$$\|A\| = \sup\{\|Ax\| : x \in \mathbb{R}^n, \quad \|x\| \leqq 1\}$$

We say that this matrix norm is the **subordinate norm** to the given vector norm or the **norm induced** by the given vector norm. The close relationship between the two is useful, because it leads to the following inequality, which is true for all vectors $\boldsymbol{x}$:

$$\|\boldsymbol{A}\boldsymbol{x}\| \leqq \|\boldsymbol{A}\| \, \|\boldsymbol{x}\|$$

The matrix norms subordinate to the vector norms discussed above are, respectively,

$$\|\boldsymbol{A}\|_1 = \max_{1 \leqq j \leqq n} \textstyle\sum_{i=1}^{n} |a_{ij}| \qquad \textbf{(ℓ_1-matrix norm)}$$

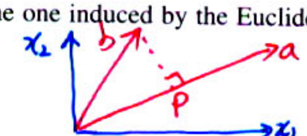
$$\|\boldsymbol{A}\|_2 = \max_{1 \leqq k \leqq n} \sigma_k \qquad \textbf{(Spectral/ℓ_2-matrix norm)}$$

$$\|\boldsymbol{A}\|_\infty = \max_{1 \leqq i \leqq n} \textstyle\sum_{j=1}^{n} |a_{ij}| \qquad \textbf{(ℓ_∞-matrix norm)}$$

Here, σ_k are the singular values of $\boldsymbol{A}$. (Refer to Section 8.2 for definitions.) Note from above that the matrix norm subordinate to the Euclidean vector norm is not what most students think that it should be, namely,

$$\|\boldsymbol{A}\|_{\mathrm{F}} = \left\{ \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^2 \right\}^{1/2} \qquad \textbf{(Frobenius norm)}$$

This is indeed a matrix norm; however, it is not the one induced by the Euclidean vector norm.



Gram-Schmidt Process

The **projection operator** is defined to be

$$\operatorname{proj}_{\boldsymbol{y}} \boldsymbol{x} = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\langle \boldsymbol{y}, \boldsymbol{y} \rangle} \boldsymbol{y}$$

$a^T(b - \bar{x}a) = 0$

$a^T b = \bar{x} a^T a$

$\bar{x} = \frac{a^T b}{a^T a} = \frac{\langle a, b \rangle}{\langle a, a \rangle}$

that projects the vector $\boldsymbol{x}$ orthogonally onto the vector $\boldsymbol{y}$. The Gram-Schmidt process can be written as

$$\begin{aligned}
z_1 &= v_1, & q_1 &= \frac{z_1}{\|z_1\|} \\
z_2 &= v_2 - \operatorname{proj}_{z_1} v_2, & q_2 &= \frac{z_2}{\|z_2\|} \\
z_3 &= v_3 - \operatorname{proj}_{z_1} v_3 - \operatorname{proj}_{z_2} v_3, & q_3 &= \frac{z_3}{\|z_3\|}
\end{aligned}$$

In general, the k step is

$$z_k = v_k - \sum_{j=1}^{k-1} \operatorname{proj}_{v_j} v_k, \qquad q_k = \frac{z_k}{\|z_k\|}$$

Here $\{z_1, z_2, z_3, \ldots, z_k\}$ is an orthogonal set and $\{q_1, q_2, q_3, \ldots, q_k\}$ is an orthonormal set. When implemented on a computer, the Gram-Schmidt process is numerically unstable because the vectors z_k may not be exactly orthogonal due to roundoff errors. By a minor modification, the Gram-Schmidt process can be stabilized. Instead of computing the vectors u_k as above, it can be computed a term at a time. A computer algorithm for the modified Gram-Schmidt process

```
for j = 1 to k
    for i = 1 to j − 1
        s ← ⟨v_j, v_i⟩
        v_j ← v_j − s v_i
    end for
v_i ← v_j/||v_j||
end for
```

Here the vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are replaced with orthonormal vectors that span the same subspace. The i-loop removes components in the $\boldsymbol{v}_i$ direction followed by normalization of the vector. In exact arithmetic, this computation gives the same results as the original form above. However, it produces smaller errors in finite-precision computer arithmetic.

EXAMPLE 1 Consider the vectors $\boldsymbol{v}_1 = (1, \varepsilon, 0, 0)$, $\boldsymbol{v}_1 = (1, 0, \varepsilon, 0)$, and $\boldsymbol{v}_1 = (1, 0, 0, \varepsilon)$. Assume ε is a small number. Carry out the standard Gram-Schmidt procedure and the modified Gram-Schmidt procedure. Check the orthogonality conditions of the resulting vectors.

Solution Using the classical Gram-Schmidt process, we obtain $\boldsymbol{u}_1 = (1, \varepsilon, 0, 0)$, $\boldsymbol{u}_2 = (0, -1, 1, 0)/\sqrt{2}$, and $\boldsymbol{u}_3 = (0, -1, 0, 1)/\sqrt{2}$. Using the modified Gram-Schmidt process, we find $z_1 = (1, \varepsilon, 0, 0)$, $z_2 = (0, -1, 1, 0)/\sqrt{2}$, and $z_3 = (0, -1, -1, 2)/\sqrt{6}$. Checking orthogonality, we find $\langle \boldsymbol{u}_2, \boldsymbol{u}_3 \rangle = \frac{1}{2}$ and $\langle z_2, z_3 \rangle = 0$. ■

Gram-Schmidt Process

- a method for orthogonalizing a set of vectors in an inner product space.
- takes a finite, linearly independent set $S = \{v_1, \ldots v_k\}$ for $k \le n$ and generates an orthogonal set $S' = \{u_1, \ldots u_k\}$ that spans the same k-dimensional subspace of $\mathbb{R}^n$ as S.

Hilbert Space

the generalizations of $\mathbb{R}^3$ which possess an analog of the inner-product and are "complete" in the sense that the limits needed to perform calculus exist.

Functional

a function that takes a vector as its argument and returns a scalar.