LINEAR ALGEBRA

Numerical linear algebra underlays much of computational mathematics, including such topics as optimization, the numerical solution of ordinary and partial differential equations, approximation theory, the numerical solution of integral equations, computer graphics, and many others. We begin by doing a fairly rapid presentation of topics from linear algebra (Chapter 7), and then we study the solution of linear systems (Chapter 8) and the solution of the matrix eigenvalue problem (Chapter 9).

VECTOR SPACES

We will use the *vector spaces* $\mathbb{R}^n$ and $\mathbb{C}^n$, the spaces of n-tuples of real and complex numbers, respectively. We usually regard these as column vectors,

$$v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

but to save space, I will often write them in their transposed form,

$$v = \begin{bmatrix} v_1, & \cdots, & v_n \end{bmatrix}^T$$

We can add them and multiply them by scalars, in the standard manner. There are other vector spaces of interest. For example,

$$\mathcal{P}_n = \{p \mid p(x) \text{ a polynomial of degree } \leq n\}$$

and

$$C[a, b] = \{f \mid f(x) \text{ continuous on } [a, b]\}$$

Vector spaces satisfy certain properties with respect to their operations of vector addition and scalar multiplication, which include standard properties such as the commutative and distributive laws, as well as others we do not review here.

For $\mathcal{W}$ to be a *subspace* of a vector space $\mathcal{V}$ means that $\mathcal{W} \subset \mathcal{V}$ and that $\mathcal{W}$ is a vector space on its own with respect to the operations of vector addition and scalar multiplication inherited from $\mathcal{V}$.

We say $v_1, ..., v_m$ are *dependent* if there is a set of scalars $c_1, ...c_m$ (at least some of which are nonzero) for which

$$c_1 v_1 + \cdots + c_m v_m = 0$$

The vectors $v_1, ..., v_m$ are *independent* if they are not dependent.

We say $\{v_1, ..., v_m\}$ is a *basis* for a vector space if for every $v \in \mathcal{V}$, there is a unique choice of coefficients $\{c_1, ...c_m\}$ for which

$$v = c_1 v_1 + \cdots + c_m v_m$$

If such a basis exists, we say the space is finite dimensional, and we define m to be the dimension.

Theorem: If a vector space $\mathcal{V}$ has a finite basis $\{v_1, ..., v_m\}$, then every basis for $\mathcal{V}$ has exactly m elements.

Examples: (1) $\mathbb{R}^n$ and $\mathbb{C}^n$ have dimension n. The standard basis for both is

$$e_i = [0, ..., 1, 0, ..., 0]^T, \quad i = 1, ..., n$$

with the 1 in position i.

(2) The polynomials of degree$\leq n$, denoted earlier by $\mathcal{P}_n$, has dimension $n+1$. The standard basis is

$$\left\{1, x, x^2, ..., x^n\right\}$$

The Legendre polynomials $\{P_0(x), ..., P_n(x)\}$ and Chebyshev polynomials $\{T_0(x), ..., T_n(x)\}$ are often a preferable basis for many applications.

(3) The multivariable polynomials

$$\sum_{i+j\leq n} a_{i,j} x^i y^j$$

of degree$\leq n$ are a vector space of dimension

$$\frac{1}{2}(n+1)(n+2)$$

(4) $C[a, b]$ is infinite dimensional.

MATRICES

Matrices are rectangular arrays of real or complex numbers, onto which we impose arithmetic operations that are generalizations of the operations of addition and multiplication for real and complex numbers. The general form a matrix of order $m \times n$ is

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}$$

I assume you are well-acquainted with matrices and their arithmetic properties, at least those aspects usually covered in beginning linear algebra courses. You should review the material in the text in Chapter 7, partly to learn the notation used in the text when working with matrices and vectors.

The arithmetic properties of matrix addition and multiplication are listed on page 466, and some of them require some work to show. For example, consider showing the distributive law for matrix multiplication,

$$(AB)\,C = A\,(BC)$$

with A, B, C matrices of respective orders $m\times n, n\times p$, and $p \times q$, respectively. Writing this out, we want to show

$$\sum_{k=1}^{p} (AB)_{i,k}\,C_{k,l} = \sum_{j=1}^{n} A_{i,j}\,(BC)_{j,l}$$

for $1 \le i \le m$, $1 \le l \le q$.

With new situations, we often use notation to suggest what should be true. But this is done only after deciding what actually is true.

LINEAR SYSTEMS

The linear system

$$\begin{aligned} a_{1,1}x_1 + \cdots + a_{1,n}x_n &= b_1 \\ &\vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n &= b_m \end{aligned}$$

is a system of m linear equations in the n unknowns $x_1, ..., x_n$. This can be written in matrix-vector notation as

$$Ax = b$$

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}, \ x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \ v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

During most of this course, we will consider only the case of square systems, in which $n = m$, as these occur most frequently in practice.

THEOREM

Let A be an $n \times n$ square matrix with elements in $\mathbb{R}$ (or $\mathbb{C}$), and let the vector space $\mathcal{V}$ be $\mathbb{R}^n$ (or $\mathbb{C}^n$). Then the following are equivalent statements.

(a) $Ax = b$ has a unique solution $x \in \mathcal{V}$ for every $b \in \mathcal{V}$.
(b) $Ax = b$ has a solution $x \in \mathcal{V}$ for every $b \in \mathcal{V}$.
(c) $Ax = 0$ implies $x = 0$
(d) A^{-1} exists
(e) $\det A \neq 0$
(f) $\operatorname{rank} A = n$.

There are many ways of defining $\det A$, the determinant of A; and it is a part of all beginning linear algebra courses. The student should review the properties of determinants, including their evaluation by *expansion by minors* or *cofactors*.

For rank A, define the *row rank* of A to be the number of independent rows in A, regarding them as elements of $\mathbb{R}^n$ (or $\mathbb{C}^n$). Similarly, define the *column rank* of A to be the number of independent columns in A, again regarding them as elements of $\mathbb{R}^n$ (or $\mathbb{C}^n$). It can be shown that the row rank and column rank are equal, and this number is called the *rank* of A. [These definitions generalize to general rectangular matrices, and again the row rank and column rank are equal. This leads to a well defined concept of rank for all rectangular matrices.]

The proof of the theorem takes up a goodly portion of introductory linear algebra courses. One of the useful things to note in proving some of the equivalences is

$$Ax = x_1 A_{*,1} + \cdots + x_n A_{*,n}$$

with $A_{*,1}, ..., A_{*,n}$ the columns of A.

INNER PRODUCT SPACES

We say a vector space $\mathcal{V}$ is an *inner product space* if we can introduce a sense of *angle* into it. More precisely, we introduce a generalization of the *dot product* from vector field theory.

For $x, y \in \mathbb{R}^n$, define

$$(x, y) = \sum_{i=1}^{n} x_i y_i = x^T y = y^T x$$

For $x, y \in \mathbb{C}^n$, define

$$(x, y) = \sum_{i=1}^{n} x_i \overline{y}_i = y^* x$$

Define the *Euclidean norm* of $x \in \mathbb{R}^n$ or $\mathbb{C}^n$ by

$$\|x\|_2 = (x, x)^{\frac{1}{2}} = \left[\sum_{i=1}^{n} |x_i|^2 \right]^{\frac{1}{2}}$$

Cauchy-Schwartz inequality: For $x, y \in \mathcal{V}$ ($\mathbb{R}^n$ or $\mathbb{C}^n$),

$$|(x, y)| \le \|x\|_2 \, \|y\|_2$$

PROOF: We prove it for only the case $\mathcal{V} = \mathbb{R}^n$, and there is a similar proof for the complex case. Assume $x, y \neq 0$, as otherwise the result is easily true (both sides equal 0). Then note that

$$(x + \alpha y, x + \alpha y) \ge 0$$

for any choice of real number α. Using the properties of the inner product, this transforms to

$$(x, x) + 2\alpha(x, y) + \alpha^2 (y, y) \ge 0$$

This is a quadratic polynomial with respect to the variable α; and it is never negative. Therefore, its discriminant must be non-positive:

$$b^2 - 4ac = 4(x, y)^2 - 4\,(x, x)\,(y, y) \le 0$$

This completes the proof.

The triangle inequality: For $x, y \in \mathcal{V}$ ($\mathbb{R}^n$ or $\mathbb{C}^n$),

$$\|x+y\|_2 \leq \|x\|_2 + \|y\|_2$$

PROOF: Prove this in its squared form, again for the case of $\mathbb{R}^n$.

$$\begin{aligned} \|x+y\|_2^2 &= (x+y, x+y) \\ &= (x,x) + 2\,(x,y) + (y,y) \\ &\leq \|x\|_2^2 + 2\,\|x\|_2\,\|y\|_2 + \|y\|_2^2 \\ &= (\|x\|_2 + \|y\|_2)^2 \end{aligned}$$

Geometrically, interpret $x+y$ as the diagonal from the origin of the parallelogram determined by the sides x and y from the origin.

ANGLES

We can introduce a sense of angle in real inner product spaces. In particular, we say θ is the angle for which

$$\cos\theta = \frac{(x, y)}{\|x\|_2 \, \|y\|_2}, \quad 0 \le \theta \le \pi$$

This is justified by the Cauchy-Schwartz inequality, since then the fraction is in $[-1, 1]$. This is a generalization of the well-known property for dot products,

$$v \cdot w = |v| \, |w| \cos\theta$$

with θ the angle between v and w.

We say two vectors $x, y \in \mathcal{V}$ are *orthogonal* if

$$(x, y) = 0$$

We use this for both real and complex vector spaces.

ORTHOGONAL MATRICES

We define a square real matrix U to be *orthogonal* if it satisfies

$$U^T U = U U^T = I$$

EXAMPLE: For any angle θ, define

$$U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

This is an orthogonal matrix. For a general vector $x = [x_1, x_2]^T$, the matrix Ux is the rotation of x thru an angle of θ. The matrix U^T corresponds to a rotation thru an angle of $-\theta$, and thus $U^T U x = x$ for all $x \in \mathbb{R}^2$.

Write U in partitioned form:

$$U = [u_1, ..., u_n]$$

in terms of the columns of U. Then combining

$$U^T U = I$$

and

$$\begin{bmatrix} u_1^T \\ \vdots \\ u_n^T \end{bmatrix} \begin{bmatrix} u_1, & \cdots, & u_n \end{bmatrix} = \begin{bmatrix} u_1^T u_1 & \cdots & u_1^T u_n \\ \vdots & & \vdots \\ u_n^T u_1 & \cdots & u_n^T u_n \end{bmatrix}$$

implies

$$(u_i, u_j) = u_i^T u_j = \delta_{i,j}, \quad 1 \le i, j \le n$$

This means the columns of U are orthogonal vectors in $\mathbb{R}^n$. Since there are n of them, and since $\mathbb{R}^n$ has dimension n, the columns of U form an orthogonal basis for $\mathbb{R}^n$. An analogous result is true for the rows of U.

For complex matrices, we call the matrix U *unitary* if it satisfies

$$U^*U = UU^* = I$$

The columns (and the rows) of such a U form an orthogonal basis in $\mathbb{C}^n$.

ORTHOGONAL BASIS

Let $\{u_1, ..., u_n\}$ be an orthogonal basis for $\mathbb{R}^n$. Why is this important?

For a general vector $x \in \mathbb{R}^n$, write

$$x = c_1 u_1 + \cdots + c_n u_n$$

Form the inner product of x with u_k. Then

$$\begin{aligned}(x, u_k) &= (c_1 u_1 + \cdots + c_n u_n, u_k) \\ &= c_1 (u_1, u_k) + \cdots + c_n (u_n, u_k) \\ &= c_k\end{aligned}$$

for any $k = 1, ..., n$. Thus

$$x = (x, u_1) u_1 + \cdots + (x, u_n) u_n$$

This is a decomposition of x into orthogonal components $\{(x, u_k) u_k\}$. When dealing with functions $x(t)$, this is called an *orthogonal expansion*. Note that

$$(x, u_k) = |x| \, |u_k| \cos \theta_k = |x| \cos \theta_k$$

with θ_k the angle between x and u_k.