Basic Concepts in Linear Algebra

Grady B. Wright
Department of Mathematics
Boise State University

Numerical linear algebra is one of the pillars of computational mathematics. In these notes we review (or introduce) some basic concepts in linear algebra that will be useful to the course.

1 Linear systems of equations

Linear systems of equations occur in almost every area of the applied science, engineering, and mathematics. A linear system of m equations and n unknowns can be expressed in the following general form:

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \cdots & + & a_{1n}x_n & = & b_1, \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \cdots & + & a_{2n}x_n & = & b_2, \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & + & \cdots & + & a_{3n}x_n & = & b_3, \\ \vdots & & \vdots & & \vdots & & \ddots & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & a_{m3}x_3 & + & \cdots & + & a_{mn}x_n & = & b_m. \end{array} \tag{1}$$

Here a_{ij} are the coefficients of the systems, b_i are the right hand sides (RHS), and x_j are the unknown values that must be determined. a_{ij} and b_i will be given by the problem.

Linear systems can be classified into the following three types:

1. **Square linear system**: If the number of equations equals the number of unknowns (i.e. $m = n$).
2. **Overdetermined system**: If the number of equations is greater than the number of unknowns (i.e. $m > n$).
3. **Underdetermined system**: If the number equations is less than the number of unknowns (i.e. $m < n$).

2 Matrices and vectors

A convenient notation to describe a linear system of equations is in terms of *matrices* and *vectors*.

2.1 Matrices

A matrix is just a table of numbers containing m rows and n columns. It can be expressed as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}.$$

We typically use capital letters to denote matrices. A common shorthand notation applied to the above matrix is

$$A = \left\{a_{ij}\right\}_{i=1,\ldots,m}^{j=1,\ldots,n},$$

or simply

$$A = \{a_{ij}\},$$

where the values for i and j are understood from the problem.

Finally, we use the following notation to specify that A is a matrix with m rows and n columns containing real numbers: $A \in \mathbb{R}^{m \times n}$.

2.2 Vectors

If the matrix only has one column then the matrix is called a **vector** (or **column vector**). A column vector with n entries can be expressed as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}.$$

A matrix with only one row is called a **row vector**. A row vector with n entries can be expressed as

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix}.$$

We typically use bold letters to denote matrices. A column vector with n real entries is denoted by $\mathbf{x} \in \mathbb{R}^n$, while a row vector with n real entries is denoted by $\mathbf{x} \in \mathbb{R}^{1 \times n}$.

2.3 Matrix-vector operations

Below is a brief summary of the operations that can be done with matrices and vectors.

1. **Matrix addition**: Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times n}$ then the sum of A and B is given by

   $$A + B = \left\{a_{ij} + b_{ij}\right\}_{i=1,\ldots,m}^{j=1,\ldots,n}.$$

   Note that this is just the sum of the corresponding entries of the elements of A and B. For this sum to make sense the matrices must be the same size.

2. **Scalar multiplication**: Let α be a real number and $A \in \mathbb{R}^{m \times n}$ then the product of α and A is given by

   $$\alpha A = \left\{\alpha\, a_{ij}\right\}_{i=1,\ldots,m}^{j=1,\ldots,n}.$$

   Note that this is just α times each entry of A.

3. **Matrix-vector product**: Let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{x} \in \mathbb{R}^n$ then the product of A and $\mathbf{x}$ is given by

   $$A\mathbf{x} = x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{m2} \end{bmatrix} + x_3 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ \vdots \\ a_{m3} \end{bmatrix} + \ldots + x_n \begin{bmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{mn} \end{bmatrix} \tag{2}$$

   Thus, the product $A\mathbf{x}$ is a *linear combination of the columns of* A. We note the following important observations:

   - The only way for this product to make sense is if A has the same number of columns as $\mathbf{x}$ does rows.

- $A\mathbf{x} \neq \mathbf{x}A$, i.e. the product does not commute.
- $A\mathbf{x} \in \mathbb{R}^m$, i.e. the product is a column vector containing m real numbers.
- If we let $\mathbf{b} = A\mathbf{x}$ then we can alternatively express the ith entry of $\mathbf{b}$ as

$$b_i = \sum_{j=1}^{n} a_{ij} x_j \,, \quad i = 1, \ldots, m.$$

- In general, computing $A\mathbf{x}$ using the above formulas requires mn multiplications and $m(n-1)$ additions.

4. **Matrix-Matrix product**: Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times p}$, and C denote the matrix-matrix product AB. Furthermore, let $\mathbf{b}_k$ represent the kth column of B and $\mathbf{c}_k$ denote the kth column of C. Then each column of C is given by

$$\mathbf{c}_k = A\mathbf{b}_k = b_{1k} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{bmatrix} + b_{2k} \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{m2} \end{bmatrix} + b_{3k} \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \\ \vdots \\ a_{m3} \end{bmatrix} + \ldots + b_{nk} \begin{bmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{mn} \end{bmatrix}, \quad k = 1, \ldots, p.$$

Here the final matrix C will be given by

$$C = \left[\begin{array}{c|c|c|c|c} \mathbf{c}_1 & \mathbf{c}_2 & \mathbf{c}_3 & \cdots & \mathbf{c}_p \end{array} \right].$$

Thus, the kth column of the product AB is a linear combination of the columns of A with the coefficients in the linear combinations being determined by entries in the kth column of B. We note the following important observations:

- The only way for the product AB to make sense is if A has the same number of columns as B does rows.
- In general, $AB \neq BA$, i.e. the product does not commute.
- $AB \in \mathbb{R}^{m \times p}$, i.e. the product is a matrix containing m rows and p columns.

$$b_i = \sum_{j=1}^{n} a_{ij} x_j \,, \quad i = 1, \ldots, m.$$

- In general, computing AB using the above formulas requires mnp multiplications and $m(n-1)p$ additions.

5. **Transpose**: Let $A \in \mathbb{R}^{m \times n}$ with entries $A = \left\{ a_{ij} \right\}_{i=1,\ldots,m}^{j=1,\ldots,n}$ then the transpose of A is given as $A^T = \left\{ a_{ji} \right\}_{j=1,\ldots,n}^{i=1,\ldots,n}$, or in expanded form:

$$A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{m1} \\ a_{12} & a_{22} & a_{32} & \cdots & a_{m2} \\ a_{13} & a_{23} & a_{33} & \cdots & a_{m3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{mn} \end{bmatrix}.$$

The transpose switches the columns of A with the rows of A so that $A^T \in \mathbb{R}^{n \times m}$. Note that the transpose of a column vector $\mathbf{x}$ is a row vector, and vice versa.

Let $B \in \mathbb{R}^{n \times p}$ then we have the following result on the transpose of the product AB:

$$(AB)^T = B^T A^T$$

6. **Vector-Vector products**: There are two types of vector-vector products that arise quite frequently. While these can be derived from the definition for matrix-matrix products, I feel it is worth stating them separately. First, let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then the *inner product* or *dot product* of $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = \sum_{j=1}^{n} x_j y_j.$$

Note that the inner product is a single number.

Now let $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ then the *outer product* of $\mathbf{x}$ with $\mathbf{y}$ is

$$\mathbf{x}\mathbf{y}^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_m \end{bmatrix} = \begin{bmatrix} x_1 y_1 & x_1 y_2 & \cdots & x_1 y_m \\ x_2 y_1 & x_2 y_2 & \cdots & x_2 y_m \\ \vdots & \vdots & \ddots & \vdots \\ x_n y_1 & x_n y_2 & \cdots & x_n y_m \end{bmatrix}$$

Note that the outer product is a matrix of size n-by-m.

3 Linear systems in matrix-vector notation

Letting $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{b} \in \mathbb{R}^m$, and $A \in \mathbb{R}^{m \times n}$, we can use the definitions from the previous section to express the linear system in (1) as $A\mathbf{x} = \mathbf{b}$ or

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_m \end{bmatrix}}_{\mathbf{b}}. \tag{3}$$

Using the very important fact that $A\mathbf{x}$ is a linear combination of the columns of A (as illustrated in (2)), we can infer that the only way there will be a solution to the above linear system is if $\mathbf{b}$ can be written as a linear combination of the columns of A. There are three possibilities for a linear system:

1. There are an *infinite number of solutions* to the linear system (i.e. an infinite number of linear combinations of the columns of A that equal $\mathbf{b}$).

2. There is *one unique solution* to the linear system (i.e. only one way to linearly combine the columns of A to equal $\mathbf{b}$).

3. There is *no solution* to the linear system (i.e. there is no possible linear combination of the columns of A to equal $\mathbf{b}$).

4 Some special types of matrices

4.1 Diagonal matrix

A diagonal matrix is an n-by-n square matrix with zeros on in every entry except possibly the main diagonal:

$$D = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & d_n \end{bmatrix},$$

where d_j, $j = 1, \ldots, n$ are some real numbers.

4.2 Identity matrix

The identity matrix is a diagonal matrix with every diagonal entry equal to 1:

$$I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

It has the property that for any matrix $A \in \mathbb{R}^{n\times n}$, $IA = AI = A$.

4.3 Upper triangular matrix

A matrix $U \in \mathbb{R}^{m\times n}$ is upper triangular if all the entries below its main diagonal are zero. Square n-by-n upper triangular matrices take the form

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ 0 & u_{22} & u_{23} & \cdots & u_{2n} \\ 0 & 0 & u_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & u_{nn} \end{bmatrix},$$

where $u_{i,j}$, $i = 1, \ldots, n$, $j = i, \ldots, n$ are some real numbers.

4.4 Lower triangular matrix

A matrix $L \in \mathbb{R}^{m\times n}$ is lower triangular if all the entries above its main diagonal are zero. Square n-by-n lower triangular matrices take the form

$$L = \begin{bmatrix} \ell_{11} & 0 & 0 & \cdots & 0 \\ \ell_{21} & \ell_{22} & 0 & \cdots & 0 \\ \ell_{31} & \ell_{32} & \ell_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell_{n1} & \ell_{n2} & \ell_{n3} & \cdots & \ell_{nn} \end{bmatrix},$$

where $\ell_{i,j}$, $i = 1, \ldots, n$, $j = i, \ldots, n$ are some real numbers.

4.5 Symmetric matrix

A matrix A is symmetric if $A = A^T$. Note that only square matrices can be symmetric.

5 Inverse of a matrix

Let A be an n-by-n square matrix (i.e. $A \in \mathbb{R}^{n \times n}$). If there exists a square matrix $B \in \mathbb{R}^{n \times n}$ such that

$$BA = AB = I,$$

where I is the n-by-n identity matrix, then B is called the inverse of A. We denote the inverse of A by A^{-1}. If A^{-1} exists then A is called *nonsingular*, otherwise it is *singular*.

Suppose A is nonsingular then the following statements are true

- A^{-1} is unique
- A^{-1} is nonsingular and its inverse is A
- A^T is nonsingular
- If $B \in \mathbb{R}^{n \times n}$ is nonsingular then AB is nonsingular and $(AB)^{-1} = B^{-1}A^{-1}$

5.1 Relationship to the solution of a square linear system

If A is a square, nonsingular matrix, then the solutions to the linear system $A\mathbf{x} = \mathbf{b}$ is given as $\mathbf{x} = A^{-1}\mathbf{b}$. Note, however, that when solving a linear system, one should never first compute A^{-1} and then compute the product $A^{-1}\mathbf{b}$. There are much better ways to solve the system using Gaussian *elimination* (when n is not too large).

6 Vector and matrix norms

6.1 Vector norms

A *vector norm* is a scalar quantity that reflects the "size" of a vector $\mathbf{x}$. We denote the norm of a vector $\mathbf{x}$ as $\|\mathbf{x}\|$. There are many ways to define the size of a vector. With $\mathbf{x} \in \mathbb{R}^n$, the three most popular are

$$\begin{aligned} \text{One-norm}: \quad & \|\mathbf{x}\|_1 = \sum_{k=1}^{n} |x_k|, \\ \text{Two-norm}: \quad & \|\mathbf{x}\|_2 = \sqrt{\sum_{k=1}^{n} |x_k|^2}, \\ \infty\text{-norm}: \quad & \|\mathbf{x}\|_\infty = \max_{1 \le k \le n} |x_k| \end{aligned}$$

However, a vector norm is defined, it must satisfy the following three properties to be called a norm:

1. $\|\mathbf{x}\| \ge 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = 0$ (i.e. $\mathbf{x}$ contains all zeros as its entries).
2. $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$, for any constant α.
3. $\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$, where $\mathbf{y} \in \mathbb{R}^n$. This is called the *triangle inequality*.

6.2 Matrix norms

Similarly, a *Matrix norm* is a scalar quantity that reflects the "size" of a matrix $A \in \mathbb{R}^{m \times n}$. We denote the norm of A as $\|A\|$. Any matrix norm must satisfy the following four properties:

1. $\|A\| \ge 0$ and $\|A\| = 0$ if and only if $A = 0$ (i.e. A contains all zeros as its entries).
2. $\|\alpha A\| = |\alpha|\|A\|$, for any constant α.
3. $\|A + B\| \le \|A\| + \|B\|$, where $B \in \mathbb{R}^{m \times n}$.

4. $\|AB\| \le \|A\|\|B\|$, where $B \in \mathbb{R}^{n\times p}$. This is called the submultiplicative inequality.

Each vector norm *induces* a matrix norm according to the following definition:

$$\|A\|_p = \max_{\|\mathbf{x}\|_p \neq 0} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p = 1} \|A\mathbf{x}\|_p,$$

where $\mathbf{x} \in \mathbb{R}^n$ and $p = 1, 2, \ldots$. Induced norms describe how the matrix stretches unit vectors with respect to that norm.

Two popular and easy to define induced matrix norms are

$$\text{One-norm}: \quad \|A\|_1 = \max_{1\le j\le m} \sum_{k=1}^{n} |a_{jk}|,$$

$$\infty\text{-norm}: \quad \|A\|_\infty = \max_{1\le k\le n} \sum_{j=1}^{m} |a_{jk}|.$$

The two-norm of A is defined as the *largest eigenvalue* of the matrix $A^T A$. This is computationally expensive to compute.

The most popular matrix norm that is not an induced norm is the *Frobenius* norm:

$$\|A\|_{\mathrm{F}} = \sqrt{\sum_{j=1}^{m} \sum_{k=1}^{n} |a_{jk}|^2}.$$

Some important results regarding norms:

1. $\|A\mathbf{x}\| \le \|A\|\|\mathbf{x}\|$
2. $\frac{1}{\sqrt{m}}\|A\|_1 \le \|A\|_2 \le \sqrt{n}\|A\|_1$
3. $\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty$
4. $\|A\|_2 \le \sqrt{\|A\|_1\|A\|_\infty}$