Chapter 3

Numerical Linear Algebra

Numerical linear algebra is primarily concerned with two important subjects:

1. efficient solution of linear systems, and

2. computation of eigenvalues and eigenvectors of matrices.

Numerical solution of nonlinear systems, partial differential equations, integral equations, etc., generally involve the solution of linear systems. Eigenvalue-eigenvector computation occurs in physical and biological applications. Also, for example, solution of systems of differential equations can involve eigenvalue-eigenvector computation.

This chapter deals primarily with efficient solution of linear systems, while eigenvalues and eigenvectors are treated in Chapter 5. Some good reference texts are for numerical linear algebra are [34, 37, 40, 68, 78, 85, 97, 103].

3.1 Basic Results from Linear Algebra

Here, we give a brief review of matrix algebra.[1]

DEFINITION 3.1 *Let* $\mathbb{R}^n(\mathbb{C}^n)$ *be the n-dimensional space of n-tuples of real (complex) numbers, i.e., if* $x \in \mathbb{C}^n$ *then* $x = [x_1, x_2, \ldots, x_n]^T$ *where* $x_i \in \mathbb{C}$ *for* $i = 1, 2, \cdots, n$.

DEFINITION 3.2 $x^T = [x_1, x_2, \cdots, x_n]$, $x^H = [\overline{x}_1, \overline{x}_2, \cdots, \overline{x}_n]$, *where T and H refer to transpose and conjugate transpose. (If* $x = a + ib$, *then* $\overline{x} = a - ib$.*)*

REMARK 3.1 $x^H x$ is a nonnegative number whereas xx^H is an $n \times n$ matrix. ▯

[1] This is not meant to be a self-contained introduction, but introduces our notation, and serves as a quick reference. It is assumed that the reader already knows the basic concepts.

DEFINITION 3.3 *The set of real (complex) $n \times m$ matrices is denoted by $L(\mathbb{R}^m, \mathbb{R}^n)$ $(L(\mathbb{C}^m, \mathbb{C}^n))$ or $L(\mathbb{R}^n)$ $(L(\mathbb{C}^n))$ if $m = n$. The elements of matrix A will be written a_{ij}, and we sometimes write $A = (a_{ij})$. We also may sometimes say $A \in \mathbb{R}^{m\times n}$ to mean that A is a real m by n matrix, and $A \in \mathbb{C}^{m\times n}$ to mean that A is a complex m by n matrix.*

DEFINITION 3.4 *If $A = (a_{ij})$, then $A^T = (a_{ji})$ and $A^H = (\overline{a}_{ji})$ denote the transpose and conjugate transpose of A, respectively.*

REMARK 3.2 $(AB)^H = B^H A^H$, $(AB)^T = B^T A^T$ and if A is $m \times n$ then A^T or A^H is $n \times m$. $\square$

DEFINITION 3.5 *If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then $C = AB$ where*

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

for $i = 1, \cdots, m$, $j = 1, \cdots, p$. Thus, C is an $m \times p$ matrix.

DEFINITION 3.6 *If A is an $n \times n$ matrix $(A \in L(\mathbb{C}^n))$, then a scalar λ and a nonzero x are an* eigenvalue *and* eigenvector *of A if $Ax = \lambda x$.*

DEFINITION 3.7 *Suppose $A \in L(\mathbb{R}^n)$ or $A \in L(\mathbb{C}^n)$. A^{-1} is the* inverse *of A if $A^{-1}A = AA^{-1} = I$, where I is the n by n* identity matrix, *consisting of 1's on the diagonal and 0's in all off-diagonal elements. If A has an inverse, then A is said to be* nonsingular *or* invertible.

DEFINITION 3.8 *The* rank *of a matrix A,* $\mathrm{rank}(A)$, *is the maximum number of linearly independent rows it possesses. It can be shown that this is the same as the maximum number of linearly independent columns. If A is an m by n matrix and* $\mathrm{rank}(A) = \min\{m, n\}$, *then A is said to be of* full rank. *For example, if $m < n$ and the rows of A are linearly independent, then A is of full rank.*

The following theorem deals with rank, nonsingularity, and solutions to systems of equations.

THEOREM 3.1
Let A be an $n \times n$ matrix $(A \in L(\mathbb{C}^n))$. Then the following are equivalent:

1. *A is nonsingular.*
2. $\det(A) \neq 0$, *where* $\det(A)$ *is the determinant of the matrix A.*

3. *The linear system $Ax = 0$ has only the solution $x = 0$.*

4. *For any $b \in \mathbb{C}^n$, the linear system $Ax = b$ has a unique solution.*

5. *The columns (and rows) of A are linearly independent, that is, if a_1, a_2, ..., a_n are the columns of A and $\sum_{i=1}^{n} \beta_i a_i = 0$, then $\beta_i = 0$ for $i = 1, 2, \cdots, n$. (That is,* $\text{rank}(A) = n$.*)*

REMARK 3.3 By Definition 3.6 and Theorem 3.1, λ is an eigenvalue of A if and only if $\det(A - \lambda I) = 0$. The equation $\det(A - \lambda I) = 0$ is called the *characteristic equation* of A. ▯

DEFINITION 3.9 $\rho(A) = \max_{1 \le i \le n} |\lambda_i|$, *where $\{\lambda_i\}_{i=1}^{n}$ is the set of eigenvalues of A, is called the* spectral radius *of A.*

DEFINITION 3.10 *If $A^T = A$, then A is said to be* symmetric. *If $A^H = A$, then A is said to be* Hermitian.

Example 3.1

$$\text{If } A = \begin{pmatrix} 1 & 2-i \\ 2+i & 3 \end{pmatrix}, \text{ then } A^H = \begin{pmatrix} 1 & 2-i \\ 2+i & 3 \end{pmatrix}, \quad \text{so } A \text{ is Hermitian.}$$

▯

DEFINITION 3.11 *For $A \in L(\mathbb{R}^n)$, if $A^T = A$ and $x^T Ax > 0$ for any $x \in \mathbb{R}^n$ except $x = 0$, then A is said to be* symmetric positive definite. *For $A \in L(\mathbb{C}^n)$, if $A^H = A$ and $x^H Ax > 0$ for $x \in \mathbb{C}^n$, $x \neq 0$, then A is said to be* Hermitian positive definite.[2] *Similarly, if $x^T Ax \ge 0$ (for a real matrix A) or $x^H Ax \ge 0$ (for a complex matrix A) for every $x \neq 0$, we say that A is* symmetric positive semi-definite *or* Hermitian positive semi-definite, *respectively.*

Example 3.2

$$\text{If } A = \begin{pmatrix} 4 & 1 \\ 1 & 3 \end{pmatrix}, \text{ then } A^T = A, \text{ so } A \text{ is symmetric.}$$

---

[2]A matrix need not be symmetric or Hermitian to be positive definite or positive semi-definite, although that is the usual context. That is, A is positive definite provided $x^T Ax > 0$ for every $x \neq 0$, and positive semi-definite provided $x^T Ax \ge 0$ for every $x \neq 0$.

Also, $x^T Ax = 4x_1^2 + 2x_1x_2 + 3x_2^2 = 3x_1^2 + (x_1 + x_2)^2 + 2x_2^2 > 0$ for $x \neq 0$. Thus, A is symmetric positive definite. ∎

PROPOSITION 3.1
If A is symmetric positive definite or Hermitian positive definite, then its eigenvalues are real and positive.

PROOF Suppose that $Ax = \lambda x$. Then $x^H Ax = \lambda x^H x$. Also, $(x^H Ax)^H = (\lambda x^H x)^H$, which yields $x^H Ax = \overline{\lambda}\, x^H x$. Thus, $\overline{\lambda}\, x^H x = \lambda x^H x$. Hence $\overline{\lambda} = \lambda$, so λ is real. Also, $\lambda = x^H Ax / x^H x > 0$. ∎

Now consider a linear system of equations $Ax = b$, where A is $n \times n$, and $b, x \in \mathbb{R}^n$.

DEFINITION 3.12 *Elementary row operations on a system of linear equations are of the following three types:*

1. *interchanging two equations,*
2. *multiplying an equation by a nonzero number,*
3. *adding to one equation a scalar multiple of another equation.*

THEOREM 3.2
If system $Bx = d$ is obtained from system $Ax = b$ by a finite sequence of elementary operations, then the two systems have the same solutions.

A proof of Theorem 3.2 can be found in elementary texts on linear algebra and can be done, for example, with Theorem 3.1 and using elementary properties of determinants.

3.2 Normed Linear Spaces

We will use these concepts both in this chapter and in subsequent chapters.

DEFINITION 3.13 *Let V be a vector space (recall that a vector space is closed under addition and scalar multiplication) over the field of real or complex numbers. V is called a* normed vector space *if to each $u \in V$ a nonnegative number $\|u\|$, called the* norm *of u, is assigned with the following properties:*

1. $\|u\| \geq 0$.
2. $\|u\| = 0$ if and only if $u = 0$.
3. $\|\lambda u\| = |\lambda| \|u\|$ for $\lambda \in \mathbb{R}$ (or $\lambda \in \mathbb{C}$ if V is a complex vector space).
4. $\|u + v\| \leq \|u\| + \|v\|$ (triangle inequality).

DEFINITION 3.14 *W is a* subspace *of a real vector space V if $u \in W$, $v \in W$ imply that $\alpha u + \beta v \in W$ for all $\alpha, \beta \in \mathbb{R}$.*

DEFINITION 3.15 *Let V be a vector space. Then $u_1, u_2, \ldots, u_n \in V$ are* linearly independent *if $\sum_{i=1}^{n} \alpha_i u_i = 0$ implies that $\alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$.*

Example 3.3

Let $V = C[a, b]$, the space of continuous functions on an interval $[a, b]$. Then $u_1 = 1$, $u_2 = x$ are linearly independent, while $u_1 = 1$, $u_2 = x$, $u_3 = 2 - x$ are linearly dependent. ▯

DEFINITION 3.16 *Let $u_1, u_2, \ldots, u_n \in V$. The set of all linear combinations of $u_1, u_2, \ldots, u_n$ is called the* span *of $u_1, u_2, \ldots, u_n$.*

REMARK 3.4 Let $W = \text{span}(u_1, u_2, \ldots, u_n)$. It is easy to show that W is a subspace of V. ($w \in W$ has the form $w = \sum_{i=1}^{n} c_i u_i$.) ▯

DEFINITION 3.17 *If $V =$ span $(u_1, u_2, \ldots, u_n)$ and $u_1, u_2, \ldots, u_n$ are linearly independent, then $u_1, u_2, \ldots, u_n$ forms a* basis *for V.*

REMARK 3.5 Suppose that V has a basis of n elements. Then, every basis of V has n elements. Any collection of $n + 1$ elements is linearly dependent. ▯

DEFINITION 3.18 *If a vector space V has a basis with a finite number n of elements, then n is called the* dimension *of the vector space.*

Example 3.4

Let P^2 represent the set of polynomials of degree 2 or less. Then P^2 is a subspace of $V = C[a, b]$. $P^2 = \text{span}(1, x, x^2)$, and, since 1, x, and x^2 are linearly independent, the dimension of P^2 is 3. ▯

Consider $V = \mathbb{C}^n$, the vector space of n-tuples of complex numbers. Note that $x \in \mathbb{C}^n$ has the form $x = (x_1, x_2, \cdots, x_n)^T$. Also,

$$x + y = (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n)^T$$

and

$$\lambda x = (\lambda x_1, \lambda x_2, \cdots, \lambda x_n)^T.$$

Important norms on $\mathbb{C}^n$ are:

(a) $\|x\|_\infty = \max_{1 \le i \le n} |x_i|$: the ℓ_∞ or max norm (for $z = a + ib$, $|z| = \sqrt{a^2 + b^2} = \sqrt{z\overline{z}}$.)

(b) $\|x\|_1 = \sum_{i=1}^{n} |x_i|$: the ℓ_1 norm

(c) $\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}$: the ℓ_2 norm (*Euclidean norm*)

(d) $\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$: the ℓ_p norm

REMARK 3.6 (b) and (c) are special cases of (d), and it can be shown that $\|x\|_\infty = \lim_{p \to \infty} \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$. ☐

REMARK 3.7 It is not hard to see that (1), (2), and (3) of Definition 3.13 are satisfied for the above norms. The triangle inequality (4) of Definition 3.13 for (a) and (b) follow from $|x_i + y_i| \le |x_i| + |y_i|$, e.g.,

$$\|x + y\|_1 = \sum_{i=1}^{n} |x_i + y_i| \le \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| = \|x\|_1 + \|y\|_1.$$

☐

The triangle inequality, or Minkowski's inequality, is not proved here for p-norms since we won't use general p-norms. The triangle inequality for the Euclidean norm follows from:

THEOREM 3.3

(the Cauchy–Schwarz inequality)

$$\left| \sum_{i=1}^{n} x_i \overline{y}_i \right| \le \|x\|_2 \|y\|_2.$$

PROOF Let $|x|$ and $|y|$ be vectors with components $|x_i|$ and $|y_i|$. Then for $\theta \in \mathbb{R}$,

$$0 \le \|\theta|x| + |y|\,\|_2^2 = \theta^2 \sum_{j=1}^n |x_j|^2 + 2\,\theta \sum_{j=1}^n |x_j||y_j| + \sum_{j=1}^n |y_j|^2.$$

The quadratic polynomial in θ on the right-hand side does not change sign. Note that if $p(\theta) = a\theta^2 + b\theta + c \ge 0$ then the discriminant $b^2 - 4ac \le 0$. Thus,

$$\left(\sum_{j=1}^n |x_j||y_j|\right)^2 \le \sum_{j=1}^n |x_j|^2 \sum_{j=1}^n |y_j|^2.$$

But

$$\left|\sum_{j=1}^n x_j \overline{y}_j\right|^2 \le \left(\sum_{j=1}^n |x_j||y_j|\right)^2 \le \sum_{j=1}^n |x_j|^2 \sum_{j=1}^n |y_j|^2,$$

that is,

$$\left|\sum_{j=1}^n x_j \overline{y}_j\right| \le \|x\|_2 \|y\|_2.$$

$\square$

The triangle inequality for $\|\cdot\|_2$ now follows from the Cauchy–Schwarz inequality as follows:

$$\begin{aligned}
\|x+y\|_2 &= \left(\sum_{j=1}^n (x_j + y_j)(\overline{x}_j + \overline{y}_j)\right)^{1/2} \\
&= \left(\sum_{j=1}^n |x_j|^2 + \sum_{j=1}^n (x_j \overline{y}_j + \overline{x}_j y_j) + \sum_{j=1}^n |y_j|^2\right)^{1/2} \\
&\le \left(\|x\|_2^2 + 2\left|\sum_{j=1}^n x_j \overline{y}_j\right| + \|y\|_2^2\right)^{1/2} \\
&\le \left(\|x\|_2^2 + 2\|x\|_2\|y\|_2 + \|y\|_2^2\right)^{1/2} = \|x\|_2 + \|y\|_2.
\end{aligned}$$

An important type of normed space is an inner product space:

DEFINITION 3.19 *An* inner product *on* $\mathbb{C}^n \times \mathbb{C}^n$ *is a complex-valued function, denoted by* $(\cdot,\cdot)$ *defined on all pairs* $x, y \in \mathbb{C}^n$ *such that*

(a) $(x,x) \ge 0$ and $(x,x) = 0$ if and only if $x = 0$,

(b) $(x, y) = \overline{(y, x)}$,

(c) $(x, \lambda y + \mu z) = \overline{\lambda}(x, y) + \overline{\mu}(x, z)$ for $\lambda, \mu \in \mathbb{C}$.

REMARK 3.8 $(\lambda x, y) = \overline{(y, \lambda x)} = \lambda(x, y)$. ▯

REMARK 3.9 If $\|x\| = (x, x)^{1/2}$, then $\|\cdot\|$ is a norm. Thus, any inner product space is also a normed vector space. ▯

Example 3.5

The following define inner products.

1. $(x, y) = \langle x, y\rangle = \displaystyle\sum_{i=1}^{n} x_i \overline{y}_i$. Clearly, $\langle x, x\rangle = \|x\|_2^2$. Also, $\langle x, y\rangle = y^H x$ where $y^H = (\overline{y}_1, \overline{y}_2, \cdots, \overline{y}_n)$.

2. $(x, y) = \displaystyle\sum_{j=1}^{n}\sum_{i=1}^{n} x_j h_{ji} \overline{y}_i = \langle Hx, y\rangle = y^H H x$ where H is an $n \times n$ Hermitian positive definite matrix.

▯

We now introduce a concept and notation for describing errors in computations involving vectors.

DEFINITION 3.20 *The* distance *from u to v is defined as $\|u - v\|$.*

The following concept is also fundamental to inner product spaces. In particular, we will use it heavily in Section 3.3.8 and Section 4.2.2.

DEFINITION 3.21 *Let $(.,.)$ represent an inner product. Two vectors u and v are called* orthogonal *with respect to $(.,.)$ provided $(u, v) = 0$. A set of vectors $v^{(i)}$ is said to be orthonormal, provided $(v^{(i)}, v^{(j)}) = \delta_{ij}$, where δ_{ij} is the* Kronecker delta *function*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

To analyze iterative techniques involving vectors and matrices, we use:

DEFINITION 3.22 *A sequence of vectors $\{x^k\}_{k=1}^{\infty}$ is said to* converge *to a vector $x \in \mathbb{C}^n$ if and only if $\|x^k - x\| \to 0$ as $k \to \infty$ for some norm $\|\cdot\|$.*

REMARK 3.10 Definition 3.22 implies that a sequence of vectors $\{x^k\}$ converges to x if and only if[3] $x_i^k \to x_i$ as $k \to \infty$ for all i. (We will discuss this further later.) ▯

Example 3.6

Let $x^k = [1^k, (\frac{1}{2})^k, (\frac{1}{3})^k, (\frac{1}{4})^k]^T$. Then $x^k \to x = [1, 0, 0, 0]^T$ in the ℓ_∞- norm, since

$$\|x^k - x\|_\infty = \max_{1 \le i \le n} |x_i^k - x_i| = \left(\frac{1}{2}\right)^k \to 0 \text{ as } k \to \infty.$$

▯

DEFINITION 3.23 *Two norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are called* equivalent *if there exist positive constants c_1 and c_2 and such that*

$$c_1\|x\|_\alpha \le \|x\|_\beta \le c_2\|x\|_\alpha.$$

Hence, also,

$$\frac{1}{c_2}\|x\|_\beta \le \|x\|_\alpha \le \frac{1}{c_1}\|x\|_\beta.$$

REMARK 3.11 If the norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are equivalent, then $x^k \to x$ in norm $\|\cdot\|_\alpha$ if and only if $x^k \to x$ in norm $\|\cdot\|_\beta$. ▯

THEOREM 3.4

Any two norms on $\mathbb{C}^n$ are equivalent.

REMARK 3.12 By Theorem 3.4, convergence in any norm in $\mathbb{C}^n$ thus implies convergence in any other norm. ▯

PROOF (of Theorem 3.4) We will prove that any norm is equivalent to the ℓ_2 norm. Then any two norms are equivalent. That is, if

$$c_1\|x\|_2 \le \|x\|_\beta \le c_2\|x\|_2$$

and

$$c_1^*\|x\|_2 \le \|x\|_\alpha \le c_2^*\|x\|_2,$$

then

$$\frac{c_1}{c_2^*}\|x\|_\alpha \le \|x\|_\beta \le \frac{c_2}{c_1^*}\|x\|_\alpha.$$

[3] Here, we are considering only finite-dimensional vector spaces.

Let $\|\cdot\|$ be any vector norm. Let $e_1, e_2, \cdots e_n$ be the usual basis vectors for $\mathbb{C}^n$, that is, $e_j = (\delta_{ij}, \delta_{2j}, \cdots, \delta_{nj})^T$, $j = 1, 2, \cdots, n$, where

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

Then

$$x = (x_1, x_2, \cdots, x_n)^T = \sum_{j=1}^{n} x_j e_j.$$

Thus,

$$\|x\| \leq \sum_{j=1}^{n} |x_j| \, \|e_j\| \leq \left(\sum_{j=1}^{n} |x_j|^2 \right)^{\frac{1}{2}} \left(\sum_{j=1}^{n} \|e_j\|^2 \right)^{\frac{1}{2}}$$

by the Cauchy–Schwarz inequality. Thus,

$$\|x\| \leq \gamma_1 \|x\|_2, \text{ where } \gamma_1 = \left(\sum_{j=1}^{n} \|e_j\|^2 \right)^{\frac{1}{2}}.$$

Let $h(x) = \|x\|$, i.e., $h : \mathbb{C}^n \to \mathbb{R}$. Notice that

$$|h(x) - h(y)| = \big| \|x\| - \|y\| \big| \leq \|x - y\| \leq \gamma_1 \|x - y\|_2.$$

Let $S = \{x \in \mathbb{C}^n : \|x\|_2 = 1\}$, i.e., S is the surface of the unit ball in $\mathbb{C}^n$. The unit spherical surface S is closed and bounded, and h is a continuous function on S. By a classical theorem of topology or analysis,[4] h has a minimum at some z in S. Let $\gamma_0 = h(z) \leq h(y)$ for all $y \in S$. (Notice that $\gamma_0 > 0$ since if $\gamma_0 = 0$ then $h(z) = \|z\| = 0$ implies $z = 0$, which is impossible since $\|z\|_2 = 1$.) Now $y = x/\|x\|_2 \in S$. Thus, $0 < \gamma_0 \leq h\left(x/\|x\|_2\right) = \|x\|/\|x\|_2$. Hence, $\gamma_0 \|x\|_2 \leq \|x\|$. Combining this with our earlier result $\|x\| \leq \gamma_1 \|x\|_2$ gives $\gamma_0 \|x\|_2 \leq \|x\| \leq \gamma_1 \|x\|_2$. ∎

REMARK 3.13 Recall that earlier we claimed that $\|x - x^k\| \to 0$ as $k \to \infty$ if and only if $x_i^k \to x_i$, $1 \leq i \leq n$. This is now obvious, since there exist constants c_1 and c_2 such that $c_1 \|x - x^k\|_\infty \leq \|x - x^k\| \leq c_2 \|x - x^k\|_\infty$, keeping in mind that $\|x - x^k\|_\infty = \max_{1 \leq i \leq n} |x_i - x_i^k|$. ∎

Consider a specific case of Theorem 3.4:

PROPOSITION 3.2
For each $x \in \mathbb{C}^n$, $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\, \|x\|_\infty$.

[4] The theorem states that a continuous function on a compact set attains its minimum and a maximum at points on that set.

PROOF Let x_j be the element of x such that

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i| = |x_j|.$$

Then

$$\|x\|_\infty^2 = |x_j|^2 \le \sum_{i=1}^{n} |x_i|^2 = \|x\|_2^2.$$

Also,

$$\|x\|_2^2 = \sum_{i=1}^{n} |x_i|^2 \le \sum_{i=1}^{n} |x_j|^2 = n|x_j|^2 = n\|x\|_\infty^2.$$

Thus, $\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\,\|x\|_\infty$. ▯

We now consider norms for $n \times n$ complex matrices. In the following, A and B are arbitrary square matrices and λ is a complex number.

DEFINITION 3.24 *A* matrix norm *is a real-valued function of A, denoted by $\|\cdot\|$ satisfying:*

1. $\|A\| \ge 0$.
2. $\|A\| = 0$ *if and only if* $A = 0$.
3. $\|\lambda A\| = |\lambda|\,\|A\|$.
4. $\|A + B\| \le \|A\| + \|B\|$.
5. $\|AB\| \le \|A\|\,\|B\|$.

REMARK 3.14 In contrast to vector norms, we have an additional fifth property, referred to as a submultiplicative property, dealing with the norm of the product of two matrices. ▯

The following will be used in our analysis of iterative methods for solving systems of equations, and elsewhere.

PROPOSITION 3.3
If $\|\cdot\|$ is any matrix norm as in Definition 3.24,

$$\rho(A) \le \|A\|.$$

PROOF Let

$$B = (x, x, x, \ldots, x),$$

that is, let B be the n by n matrix, each of whose columns is the nonzero vector x, where x is such that $Ax = \lambda x$. Then

$$AB = (\lambda x, \dots \lambda x) = \lambda B.$$

Thus, by the third and fifth properties of matrix norms, $|\lambda| \, \|B\| \le \|A\| \, \|B\|$, so $|\lambda| \le \|A\|$ for any eigenvalue of A. Since the spectral radius is the maximum eigenvalue of A, the result follows. ▯

REMARK 3.15 By choosing a particular ordering of the elements, $n \times n$ matrices may be viewed as vectors in $\mathbb{C}^{n^2}$, e.g., $\tilde{a}_\ell = a_{jk}$, $\ell = j + n(k-1)$ for $j = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, n$. Thus, if we define a matrix norm using a vector norm on the n^2 vector, the first four conditions in the above definition are automatically satisfied. However, in general, condition (5) will not hold. For example, consider $\|A\|_\infty = \max_{i,j} |a_{ij}|$, and take

$$A = B = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Then $\|AB\|_\infty \not\le \|A\|_\infty \, \|B\|_\infty$. One norm for which this procedure holds is the Euclidean norm

$$\|A\|_E = \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}},$$

which is also called the *Frobenius norm*. One sees that (1), (2), (3), and (4) of Definition 3.24 are satisfied by considering the corresponding ℓ_2-norm $\| \cdot \|_2$ of the corresponding vector in $\mathbb{C}^{n^2}$. To prove (5), we have ▯

$$\|AB\|_E^2 = \sum_{i,j=1}^{n} \left(\sum_{k=1}^{n} a_{ik} b_{kj} \right)^2 \le \sum_{i,j=1}^{n} \sum_{k=1}^{n} |a_{ik}|^2 \sum_{k=1}^{n} |b_{kj}|^2 = \|A\|_E^2 \|B\|_E^2.$$

DEFINITION 3.25 *A matrix norm $\|A\|$ and a vector norm $\|x\|$ are called* compatible *if for all vectors x and matrices A we have $\|Ax\| \le \|A\| \, \|x\|$.*

REMARK 3.16 A consequence of the Cauchy–Schwarz inequality is that $\|Ax\|_2 \le \|A\|_E \|x\|_2$, i.e., the Euclidean norm $\| \cdot \|_E$ for matrices is compatible with the ℓ_2-norm $\| \cdot \|_2$ for vectors. ▯

We now define some commonly used matrix norms.

DEFINITION 3.26 *Given a vector norm* $\|\cdot\|$*, we define a* natural *or* induced *matrix norm associated with it as*

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}. \tag{3.1}$$

REMARK 3.17 If an induced matrix norm is compatible with the given vector norm, then

$$\|A\|\,\|x\| \geq \|Ax\| \text{ for all } x \in \mathbb{C}^n. \tag{3.2}$$

▯

REMARK 3.18 It is straightforward to show that an induced matrix norm satisfies the five properties required of a matrix norm so is indeed a matrix norm. In particular,

$$\|AB\| = \sup_{x\neq 0} \frac{\|ABx\|}{\|x\|} \leq \sup_{x\neq 0} \frac{\|A\|\,\|Bx\|}{\|x\|} = \|A\|\,\|B\|,$$

since $\|ABx\| = \|Az\| \leq \|A\|\,\|z\|$. ▯

REMARK 3.19 Definition 3.26 is equivalent to

$$\|A\| = \sup_{\|y\|=1} \|Ay\|,$$

since

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{x\neq 0} \left\| A\frac{x}{\|x\|} \right\| = \sup_{\|y\|=1} \|Ay\|$$

(letting $y = x/\|x\|$). ▯

REMARK 3.20 We shall use matrix norms $\|\cdot\|_\infty$, $\|\cdot\|_1$, and $\|\cdot\|_2$ induced by the corresponding vector norms. That is,

$$\|A\|_\infty = \sup_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty}, \quad \|A\|_1 = \sup_{x\neq 0} \frac{\|Ax\|_1}{\|x\|_1}, \quad \text{and} \quad \|A\|_2 = \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2}.$$

▯

REMARK 3.21 The Frobenius norm $\|\cdot\|_E$ is not a natural norm, because $\|I\| = 1$ for any natural norm but $\|I\|_E = \sqrt{n}$. Consider briefly the relation between $\|A\|_2$ and $\|A\|_E$. First note that, in general, $\|A\|_2 \neq \|A\|_E$. (For example, let $A = I$.) Also,

$$\|A\|_2 = \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} \leq \sup_{x\neq 0} \frac{\|A\|_E\|x\|_2}{\|x\|_2} = \|A\|_E.$$

Thus, $\|A\|_2 \le \|A\|_E$. ∎

We now develop explicit expressions for $\|A\|_\infty$, $\|A\|_1$, and $\|A\|_2$.

PROPOSITION 3.4
(Formulas for the induced ℓ_1 and ℓ_∞ matrix norms)

(a) $\|A\|_\infty = \max\limits_{1\le i\le n} \sum\limits_{j=1}^{n} |a_{ij}| = \{\text{maximum absolute row sum}\}.$

(b) $\|A\|_1 = \max\limits_{1\le j\le n} \sum\limits_{i=1}^{n} |a_{ij}| = \{\text{maximum absolute column sum}\}.$

PROOF

$$\|Ax\|_\infty = \max_{1\le i\le n} \Big| \sum_{j=1}^{n} a_{ij}x_j \Big| \le \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}||x_j| \le \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|\|x\|_\infty.$$

Thus,

$$\|A\|_\infty \le \max_{1\le i\le n} \left(\sum_{j=1}^{n} |a_{ij}| \right), \tag{3.3}$$

since

$$\frac{\|Ax\|_\infty}{\|x\|_\infty} \le \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|.$$

Now let k be the index of the row which has the maximum absolute row sum, i.e.,

$$\sum_{j=1}^{n} |a_{kj}| = \max_{1\le i\le n} (\sum_{j=1}^{n} |a_{ij}|),$$

and let y be defined by

$$y_j = \begin{cases} \overline{a}_{kj}/|a_{kj}| & \text{if} \quad a_{kj} \ne 0, \\ 0 & \text{if} \quad a_{kj} = 0. \end{cases}$$

Hence, $\|y\|_\infty = 1$ (if $A \ne 0$). Also,

$$\begin{aligned} \|Ay\|_\infty &= \max_i \Big| \sum_{j=1}^{n} a_{ij}y_i \Big| \qquad (3.4) \\ &\ge \Big| \sum_{j=1}^{n} a_{kj}y_j \Big| = \sum_{j=1}^{n} |a_{kj}| = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}|. \end{aligned}$$

Since, from the definition of $\|A\|_\infty$, $\|A\|_\infty \geq \|Ay\|_\infty > 0$, (3.4) implies

$$\|A\|_\infty \geq \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|. \tag{3.5}$$

Combining (3.3) and (3.5) gives (a).

To prove (b), consider

$$\|Ax\|_1 = \sum_{i=1}^{n} \Big| \sum_{j=1}^{n} a_{ij}x_j \Big| \leq \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}||x_j| \leq \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|\|x\|_1.$$

(Recall $\|x\|_1 = \sum_{j=1}^{n} |x_j|$.) Thus,

$$\|A\|_1 \leq \max_{1\leq j\leq n} \left(\sum_{i=1}^{n} |a_{ij}| \right). \tag{3.6}$$

Let k be such that $\sum_i |a_{ik}| = \max_j (\sum_i |a_{ij}|)$ and let $y = e_k$, where $(e_k)_j = \delta_{kj}$. Then $\|y\|_1 = 1$, and

$$\|Ay\|_1 = \sum_{i=1}^{n} \Big| \sum_{j=1}^{n} a_{ij}y_j \Big| = \sum_{i=1}^{n} \Big| \sum_{j=1}^{n} a_{ij}\delta_{jk} \Big| = \sum_{i=1}^{N} |a_{ik}| = \max_{1\leq j\leq n} \left(\sum_{i=1}^{n} |a_{ij}| \right).$$

However, from the definition of $\|A\|_1$, $\|A\|_1 \geq \|Ay\|_1$. Thus,

$$\|A\|_1 \geq \max_j (\sum_{i=1}^{n} |a_{ij}|). \tag{3.7}$$

Hence, (b) follows from (3.7) and (3.6). ∎

Now consider $\|A\|_2$. Recall that $\|A\|_2 \neq \|A\|_E$. Also, recall the following.

(i) $\rho(A) = \max_i |\lambda_i(A)| = \{$spectral radius of matrix $A\}$ (the eigenvalue of A with the largest magnitude).

(ii) A Hermitian matrix B ($B^H = B$) has real eigenvalues and a linearly independent set of eigenvectors which are orthonormal with respect to the inner product $(\cdot,\cdot)$, where $(x, y) = y^H x$.

With this, we have:

PROPOSITION 3.5
$\|A\|_2 = \sqrt{\rho(A^H A)}$.

PROOF Let $x \neq 0$ in $\mathbb{C}^n$. Then

$$\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{(Ax, Ax)}{(x, x)} = \frac{(Ax)^H Ax}{(x, x)} = \frac{x^H A^H Ax}{(x, x)} = \frac{(A^H Ax, x)}{(x, x)}.$$

Thus,

$$\frac{\|Ax\|_2}{\|x\|_2} = \left(\frac{(A^H Ax, x)}{(x, x)}\right)^{\frac{1}{2}}.$$

Now $A^H A$ is Hermitian. Let ν_i, $i = 1, 2, \cdots, n$, be the orthonormal eigenvectors of $A^H A$ and let

$$x = \sum_{i=1}^{n} c_i \nu_i.$$

Since $(\nu_i, \nu_j) = \delta_{ij}$ and since

$$A^H Ax = \sum_{i=1}^{n} \lambda_i c_i \nu_i,$$

it follows that

$$(x, x) = \sum_{i=1}^{n} |c_i|^2$$

and

$$(Ax, Ax) \ = \ (A^H Ax, x) = \sum_{i=1}^{n} \lambda_i |c_i|^2,$$

where λ_i, $1 \leq i \leq n$, are the eigenvalues of $A^H A$. Now $\lambda_i = (A\nu_i, A\nu_i) \geq 0$, so

$$\frac{\|Ax\|_2}{\|x\|_2} = \left(\frac{\sum_{i=1}^{n} \lambda_i |c_i|^2}{\sum_{i=1}^{n} |c_i|^2}\right)^{\frac{1}{2}} \leq (\max_i \lambda_i)^{\frac{1}{2}} = \sqrt{\rho(A^H A)}. \tag{3.8}$$

Thus, $\|A\|_2 \leq \sqrt{\rho(A^H A)}$. Also, letting $x = \nu_k$ where $\lambda_k = \max_i \lambda_i = \rho(A^H A)$, it follows that

$$\|A\|_2 \geq \frac{\|A\nu_k\|_2}{\|\nu_k\|_2} = (A^H A\nu_k, \nu_k)^{\frac{1}{2}} = \lambda_k^{\frac{1}{2}} = \sqrt{\rho(A^H A)}. \tag{3.9}$$

Combining the inequalities (3.8) and (3.9), we obtain $\|A\|_2 = \sqrt{\rho(A^H A)}$. ∎

It is interesting in view of Proposition 3.5 to revisit Proposition 3.3, namely, that $\rho(A) \leq \|A\|$ for any square matrix A and any matrix norm. An alternate proof of this proposition for induced matrix norms can be based on quotients $\|Ax\|/\|x\|$, where x is an arbitrary eigenvector of A. It is also interesting to note that the difference between $\|A\|$ and $\rho(A)$ may be arbitrarily large. Consider

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}.$$

Then $\rho(A) = 0$ but $\|A\|_1 = 2$. However, the following result shows that there is a matrix norm arbitrarily close to $\rho(A)$.

PROPOSITION 3.6

Let A be a given $n \times n$ matrix and let $\epsilon > 0$. Then there exists an induced matrix norm $\|\cdot\|$ such that $\|A\| \leq \rho(A) + \epsilon$.

PROOF The proof rests on the result in linear algebra that any matrix is similar to an upper triangular matrix, i.e., given any matrix A, there exists a nonsingular matrix P such that $PAP^{-1} = T = \Lambda + U$ where T is upper triangular. Furthermore, the diagonal entries of T are the eigenvalues of A, i.e., if $\Lambda = \text{diag}\,[\lambda_1, \lambda_2, \cdots, \lambda_n]$ then U has zero's on and below the diagonal. (This result is known as Schur's Theorem, and the decomposition $A = P^{-1}TP$ is known as the *Schur decomposition*.) With $\delta > 0$, define

$$D^{-1} = \text{diag}[1, \delta, \delta^2, \cdots, \delta^{n-1}].$$

Then $C = DTD^{-1} = \Lambda + E$ where $E = (e_{ij}) = DUD^{-1}$ has elements

$$e_{ij} = \begin{cases} 0 & \text{if} \quad j \leq i, \\ u_{ij}\delta^{j-i} & \text{if} \quad j > i. \end{cases}$$

Hence, the elements of E can be made arbitrarily small by choosing δ small enough. Also, note that $A = P^{-1}D^{-1}CDP$. Since DP is nonsingular, a vector norm can be defined by

$$\|x\| = \|DPx\|_2 = (x^H P^H D^H DPx)^{\frac{1}{2}}.$$

The matrix norm induced by this vector norm is

$$\|A\| = \sup_{\|y\|=1} \|Ay\|.$$

For the particular A above,

$$\|Ay\| = \|DPAy\|_2 = \|CDPy\|_2.$$

Letting $z = DPy$, we have $\|Ay\| = \|Cz\|_2 = (z^H C^H Cz)^{\frac{1}{2}}$. Put

$$C^H C = (\Lambda^H + E^H)(\Lambda + E) = \Lambda^H \Lambda + M(\delta),$$

where $M(\delta)$ is an $n \times n$ matrix whose elements are order δ, i.e., $|m_{ij}/\delta| \leq k$ for δ sufficiently small. Thus,

$$\|Ay\| = z^H C^H Cz \leq \max_i |\lambda_i^2(A)| z^H z + |z^H M(\delta) z| \leq (\rho^2(A) + \mathcal{O}(\delta)) z^H z.$$

Since $z = DPy$ and $\|y\| = 1$, we have $\|y\| = \|DPy\|_2 = \|z\|_2 = 1$, that is, $z^H z = 1$. Therefore,

$$\|A\| \leq [\rho^2(A) + \mathcal{O}(\delta)]^{\frac{1}{2}} \leq \rho(A) + (\mathcal{O}(\delta))^{\frac{1}{2}}.$$

For δ sufficiently small, we can make $(\mathcal{O}(\delta))^{\frac{1}{2}} < \epsilon$. ∎

The following is now worth noting.

PROPOSITION 3.7
All matrix norms are equivalent, that is, given any two matrix norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ there exist positive constants c_1 and c_2 such that

$$c_1\|A\|_\alpha \leq \|A\|_\beta \leq c_2\|A\|_\alpha.$$

PROOF By particular ordering of the elements of A, A can be viewed as a vector in $\mathbb{C}^{n^2}$. Thus, a matrix norm of A can be viewed as a vector norm in $\mathbb{C}^{n^2}$. Since any two vector norms are equivalent, any two matrix norms are equivalent.
∎

Note: Propositions 3.3, 3.6, and 3.7 give $\rho(A) \leq \|A\| \leq c\rho(A) + \epsilon$ where $\|\cdot\|$ is any matrix norm, c is a positive constant that depends on the norm $\|\cdot\|$ and matrix A, and $\rho(A)$ is the spectral radius of A. However, given A and $\epsilon > 0$, there is a norm $\|\cdot\|'$ such that $\rho(A) \leq \|A\|' \leq \rho(A) + \epsilon$.

We now consider sequences of matrices.

DEFINITION 3.27 $\{A_k\}_{k=1}^\infty$ converges *to matrix A if and only if $\|A_k - A\| \to 0$ as $k \to \infty$ for some matrix norm.*

Note: The choice of norm is not important from a theoretical point of view, since all matrix norms are equivalent. Thus, one should use the norm most suitable for the particular problem.
Note: It is clear that if $\|A_k - A\| \to 0$ as $k \to \infty$, then $a_{ij}^{(k)} \to a_{ij}$ as $k \to \infty$ for each i, j.

We now consider the sequence $\{A^k\}_{k=0}^\infty$, i.e., elements of the sequence are powers of A. We have the following convergence result:

THEOREM 3.5
Given an $n \times n$ matrix A, the following are equivalent.

(a) $\lim\limits_{k\to\infty} A^k = 0$ *(i.e.,* $\lim\limits_{k\to\infty} \|A^k\| = 0$*).*

(b) $\lim_{k\to\infty} A^k x = 0$, $\forall x \in \mathbb{C}^n$ *(i.e.,* $\lim_{k\to\infty} \|A^k x\| = 0$*).*

(c) $\rho(A) < 1$.

(d) there exists a norm $\|\cdot\|$ *such that* $\|A\| < 1$.

REMARK 3.22 A matrix for which (a) holds is called a convergent matrix. ▯

PROOF We will prove that (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d) $\Rightarrow$ (a).

(a) $\Rightarrow$ (b): We have $\|A^k x\| \le \|A^k\|\|x\|$. Thus, $\lim_{k\to\infty} \|A^k x\| = 0$.

(b) $\Rightarrow$ (c): Let v be an eigenvector of A with associated eigenvalue λ. By (b), $\lim_{k\to\infty} \|A^k v\| = \lim_{k\to\infty} |\lambda|^k \|v\| = 0$. This implies that $|\lambda| < 1$. Since this holds for any eigenvalue of A, $\rho(A) < 1$.

(c) $\Rightarrow$ (d): we choose for $\rho(A) < 1, \epsilon$ such that $0 < \epsilon < 1 - \rho(A)$. Then, by Proposition 3.6, $\|A\| \le \rho(A) + \epsilon < 1$.

(d) $\Rightarrow$ (a): we have $\|A^k\| \le \|A\|^k$. Thus, $\lim_{k\to\infty} \|A^k\| \le \lim_{k\to\infty} \|A\|^k = 0$, since $\|A\| < 1$.

▯

REMARK 3.23 In condition (d), <u>there exists</u> should be stressed. Consider

$$A = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}.$$

$\|A\|_\infty = \|A\|_1 = \|A\|_2 = \|A\|_E = 2 > 1$, but $\lambda_1(A) = \lambda_2(A) = 0$, so $\rho(A) = 0$. Thus, $\lim_{k\to\infty} A^k = 0$. In fact, $A^k = 0$ for $k \ge 2$. ▯

Two more useful results are now given.

PROPOSITION 3.8
The geometric series

$$\sum_{k=0}^{\infty} A^k = I + A + A^2 + A^3 + \cdots$$

converges to a certain matrix if and only if $A^k \to 0$ *as* $k \to \infty$. *Furthermore, if* $A^k \to 0$ *as* $k \to \infty$, *then* $I - A$ *is nonsingular and*

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

PROOF Let $A^k \to 0$. By Theorem 3.5, $\rho(A) < 1$. Consider $B = I - A$. Then $\lambda(B) = 1 - \lambda(A)$, where $\lambda(B)$ is any eigenvalue of B and $\lambda(A)$ is the corresponding eigenvalue of A. Thus, $\lambda(B) \neq 0$ since $|\lambda(A)| < 1$. Hence, $I - A$ is nonsingular because no eigenvalues are zero. Now consider the identity

$$(I - A)(I + A + \cdots + A^k) = I - A^{k+1}.$$

This implies

$$(I + A + \cdots + A^k) = (I - A)^{-1}(I - A^{k+1}),$$

that is,

$$\lim_{k\to\infty} \sum_{j=0}^{k} A^k = (I - A)^{-1} \lim_{k\to\infty} (I - A^{k+1}) = (I - A)^{-1}.$$

Thus,

$$(I - A)^{-1} = \sum_{j=0}^{\infty} A^j.$$

Conversely, let $\sum_{k=0}^{\infty} A^k$ be a convergent series. Then

$$A^k = \sum_{j=0}^{k} A^j - \sum_{j=0}^{k-1} A^j \to 0$$

as $k \to \infty$. That is, a necessary condition for convergence of the series is $\lim_{k\to\infty} A^k = 0$. ∎

PROPOSITION 3.9

Let $\|\cdot\|$ be a vector norm and $\|A\|$ its induced matrix norm, i.e.,

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|}.$$

If $\|A\| < 1$, then $(I - A)$ is nonsingular and

$$\frac{1}{1 + \|A\|} \leq \|(I - A)^{-1}\| \leq \frac{1}{1 - \|A\|}.$$

PROOF By Theorem 3.5 and Proposition 3.8, $(I - A)$ is nonsingular. We also have $\|I\| = 1$, since $\|\cdot\|$ is an induced matrix norm. Thus,

$$1 = \|I\| = \|(I - A)^{-1}(I - A)\| \leq \|(I - A)^{-1}\|\|I - A\| \leq \|(I - A)^{-1}\|(1 + \|A\|).$$

Hence, $1/(1 + \|A\|) \leq \|(I - A)^{-1}\|$.

Also, $(I - A)^{-1} = I + A(I - A)^{-1}$. Thus,

$$\|(I - A)^{-1}\| \leq 1 + \|A\|\|(I - A)^{-1}\|,$$

that is,

$$(1 - \|A\|)\|(I - A)^{-1}\| \leq 1.$$

Hence, $\|(I - A)^{-1}\| \leq 1/(1 - \|A\|)$. ∎

Note: This concludes, for now, the review of linear algebra. More results, such as for eigenvalues and eigenvectors and for irreducible matrices, will be given later in this chapter or in Chapter 5.

3.3 Direct Methods for Solving Linear Systems

We discuss elimination and factorization methods for solving linear systems in this section. We begin with Gaussian elimination.

3.3.1 Gaussian Elimination

We first present our notation.

3.3.1.1 Statement of the Problem

Given a nonsingular complex $n \times n$ matrix A and $b \in \mathbb{C}^n$, find $x \in \mathbb{C}^n$ such that $Ax = b$. That is, if $A = (a_{ij})$ and $b = (b_1, \cdots, b_n)^T$ find $x = (x_1, x_2, \cdots, x_n)^T$ such that

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2, \\ \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n = b_n. \end{array}$$

For a general matrix A, the goal of Gaussian elimination is to reduce A to an upper triangular matrix through a sequence of elementary row operations.[5] We will see that this is equivalent to finding an invertible matrix M such that $MA = U$, where U is upper triangular, so $MAx = Mb$, that is, $Ux = Mb$.

[5]Recall that row operations involve interchanging two rows and replacement of a row by the sum of that row and a scalar multiple of another row; see Definition 3.12 on page 88.