3. LINEAR ALGEBRA BACKGROUND

3.1 INTRODUCTION
3.2 SUMMARY OF ESSENTIAL CONCEPTS
- 3.2.1 Essentials of Linear Systems
- 3.2.2 Essentials of Eigenvalue Problems

3.3 NORMS
- 3.3.1 Vector Norms
- 3.3.2 Matrix Norms

3.4 SPECIAL CLASSES OF MATRICES
- 3.4.1 Positive Definite Symmetric Matrices
- 3.4.2 Orthogonal Matrices

3.5 APPENDICES
- A.1 Transpose
- A.2 Inverses
- A.3 Permutations of Matrices
- A.4 Determinants

3.6 REFERENCES
3.7 QUESTIONS

3.1 INTRODUCTION

> *"The problem of solving a linear system Ax = b is central in scientific computation"* ([p. 87], Golub and van Loan 1997).
>
> *"The solution of systems of linear equations forms the cornerstone of many numerical methods for solving a wide variety of practical computational problems, and, consequently, it is imperative that we be able to solve such systems accurately and efficiently*" ([p. 50], Heath 2002).

Solving systems of *n* linear algebraic equations with *n* unknowns is a *core* topic in computational physics and engineering. Linear systems are found in all areas of computational science. Discrete systems, such as static structures or electric circuits, can be investigated *directly* by systems of linear equations; continuous systems, represented by differential equations, which in turn can modeled by finite-difference or finite-element techniques which themselves can be broken down into systems of linear equations ([p. 31] Kiusalaas 2005).

Another linear-algebra technique used frequently in science and engineering is linear transforms. Usually linear transforms, a common tool in linear algebra, are a complicated mix of many simpler actions: expanding or shrinking of vectors, vector rotations, vector reflections, permutation of individual components, and so forth (e.g., [p. 116] Strang 1988; [p. 157] Heath 2002). Eigenvalues and eigen-vectors can be employed to decompose complicated linear transformations into far simpler behavior (e.g. [p. 157] Heath 2002).

Preliminaries:

The matrix ***A*** is an m-by-n ***rectangular table*** of $m \cdot n$ numbers enclosed in square brackets; the numbers are arranged in *m* horizontal rows and in *n* vertical columns. The *(i,j)* element is termed A_{ij} and equals the entry in the ith row and jth column. The numbering of the rows is from top to bottom and columns are numbered from left to right. Thus m-by-n matrix *A* is represented as

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & \dots & A_{1n} \\ A_{21} & A_{22} & A_{23} & \dots & A_{2n} \\ & & \dots & & \\ A_{m1} & A_{m2} & A_{m3} & \dots & A_{mn} \end{bmatrix}.$$

Matrix Properties:

Symmetric:	$A = A^T$
Hermitian:	$A = A^H$
Orthogonal:	$A \cdot A^T = A^T \cdot A = I$
Unitary:	$A \cdot A^H = A^H \cdot A = I$

Transpose: $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^{T} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$,

Conjugate Transpose: $\begin{bmatrix} 1+i & 1+2i \\ 2-i & 2-2i \end{bmatrix}^{H} = \begin{bmatrix} 1-i & 2+i \\ 1-2i & 2+2i \end{bmatrix}$,

Symmetric: $\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$, Nonsymmetric: $\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$,

Hermitian: $\begin{bmatrix} 1 & 1+i \\ 1-i & 2 \end{bmatrix}$, NonHermitian: $\begin{bmatrix} 1 & 1+i \\ 1+i & 2 \end{bmatrix}$

Orthogonal: $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \bullet \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^{T} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathrm{I}$,

Unitary: $\begin{bmatrix} +i\sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \bullet \begin{bmatrix} +i\sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}^{H} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathrm{I}.$

3.2 SUMMARY OF ESSENTIAL CONCEPTS

3.2.1 Essentials of Linear Systems

Systems of linear equations have the form

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n &= b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n &= b_2 \\ A_{31}x_1 + A_{32}x_2 + \cdots + A_{3n}x_n &= b_3 \\ \vdots \qquad\quad \vdots \qquad\quad & \qquad\quad \vdots \\ A_{n1}x_1 + A_{n2}x_2 + \cdots + A_{nn}x_n &= b_n, \end{aligned}$$

where the known coefficients of the n-by-n matrix, A, are represented as A_{jk} (jth row & kth column) and the known elements of the right-hand-side vector are signified by b_j and the elements of the unknown solution vector are represented by x_k. In matrix format these equations are written as

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \bullet \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

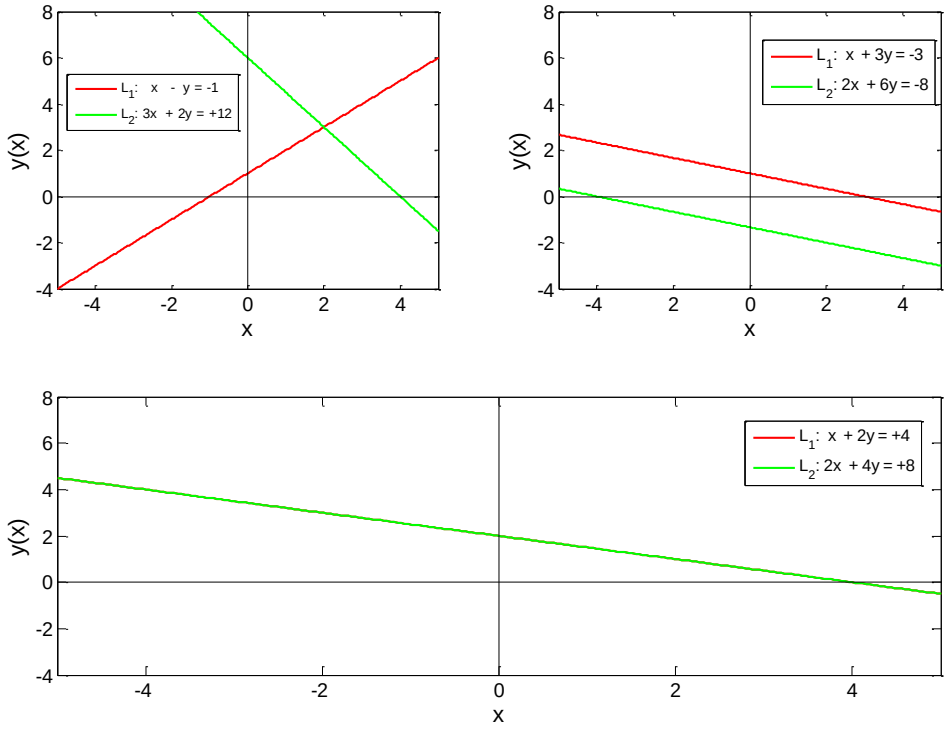
or expressed compactly in matrix-vector format as

$$A \bullet x = b.$$

An *arbitrary* system of linear equations does not necessarily possess a solution, much less a unique solution. A linear system can have no solution, an infinite number of solutions, or a single unique solution

([p. 50] Heath 2002). Linear systems that possess unique solutions are termed *nonsingular*; the existence of such unique solutions depends only on the structure of the coefficient matrix, *A* ([p. 116] Meyers 2000). Whether or not a *singular* system has no solution or has an infinite number of solutions depends on the structure of the vector *b* ([p. 50] Heath 2002).

One way to understand the concept of singularity is to inspect geometrically how small linear systems can have unique solutions. Consider the trivial example of two linear equations, each of which can be represented as a straight line in two-dimensional Cartesian space. In two dimensions unique solutions occur when the two lines intersect. When two lines are parallel with different intercepts, there is no solution. Yet when two lines lie on top of each other, there exist infinite numbers of solutions. The figure below illustrates these three cases. In (a) the system possesses a single, unique solution; the two lines intersect at a single point, which always occurs when the two slopes differ. In (b) the system has no solution; the two lines have identical slopes but different intercepts. In (c) the system has an infinite number of solutions, since the lines coincide - the two lines possess the same slopes and the same intercepts.



When a linear system of *n* equations with *n* unknowns possesses a *unique* solution, the square coefficient matrix *A* is said to be *nonsingular* or *invertible* when the following *equivalent* conditions (e.g. [p. 51] Heath 2002; [p. 8] Watkins 2002; [p. 68] Ascher & Greif 2011), are satisfied, that is if one condition holds, they all hold:

- *A* possesses an *inverse*, symbolized as A^{-1}, defined as

  $$A * A^{-1} \equiv A^{-1} * A \equiv I,$$

  where $I \equiv$ **`eye(n, n)`**, is the identity matrix . Thus, *symbolically*[1] a linear system can be solved by

[1] Practically, the inverse operation is never employed to solve systems of equations. Computationally, less expensive (~3 times faster), more stable methods, such as LU decomposition, are employed to solve linear systems.

$$A^{-1}\bullet(A\bullet x = b) = A^{-1}\bullet A\bullet x = (A^{-1}\bullet A)\bullet x = I\bullet x = x = A^{-1}\bullet b.$$

- The determinant of *A* is nonzero, $\det(A) \neq 0.0$,

  This approach is not very useful *computationally*. Solving a linear system of *n* equations requires ~$n^3/3$ floating-point multiplications ([p. 79] Heath 2002). Solving a linear system of *n* equations using determinants, *Cramer's rule*[2], requires the order of n! floating-point operations ([p. 188] Datta 1995). Additionally, Cramer rule is numerically unstable ([p. 13] Higham 2002)! Moreover, determinants are not useful in identifying singular systems, since $\det(\beta A) = \beta^n \det(A)$, where β is a scalar. Thus, depending on the value of β, $\det(\beta A)$ can be arbitrarily very small, but still represent a nonsingular matrix. Moreover, as noted by Trefethen and Bau ([p. 8] 1997) *"the determinant, though a convenient notation theoretically, rarely finds a useful role in numerical algorithms."*

- $A\cdot x = 0$, has only a *single* solution, $x = 0$.

  Thus if one creates an n-by-1 vector of zeros, *c*, then the solution vector *x* of $A\cdot x = c$, is found to be an n-by-1 vector of zeros. Consequently, for any non-zero vector, *b*, $A\cdot y = b$, has a single unique solution, $y = A^{-1}.b$.

- The columns of *A* are linearly independent[3].

  One can represent the linear system, $A\cdot x = b$, as

  $$A_1x_1 + A_2x_2 + \cdots + A_nx_n = b,$$

  where A_j is the jth column of *A* ([p. 67] Ascher & Greif 2011). *Visually* we express the linear system as –

  $$\begin{pmatrix}\vdots\\ \vdots\\ A_1\\ \vdots\\ \vdots\end{pmatrix} x_1 + \begin{pmatrix}\vdots\\ \vdots\\ A_2\\ \vdots\\ \vdots\end{pmatrix} x_2 + \cdots + \begin{pmatrix}\vdots\\ \vdots\\ A_n\\ \vdots\\ \vdots\end{pmatrix} x_n = \begin{pmatrix}\vdots\\ \vdots\\ b\\ \vdots\\ \vdots\end{pmatrix} \text{ and } x = \begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_{n-1}\\ x_n\end{pmatrix}.$$

  If A is nonsingular, then the A_j columns are linearly independent. Thus the vector, b, can be represented as a *unique* combination of the column vectors.

- The rows of *A* are also linearly independent.

---

[2] If *A* is a nonsingular n-by-n matrix, the solution x of $A\cdot x = b$ is found from Cramer's rule, $x_i = \det(A_i)/\det(A)$, for i = 1:n, and A_i is formed by replacing the ith column of *A* by the vector *b* ([p. 187] Datta 1995).

[3] The concept of l*inear independence* can be described in the following way. Given a set of vectors $\{A_1, A_2, \ldots, A_k\}$ inspect their linear combinations, $\beta_1A_1 + \beta_2A_2 + \beta_3A_3 + \ldots + \beta_kA_k$, where β_j represents the jth scalar weight. If the *only* way to create a *zero* combination vector of k length is that *all* the scalar coefficients, $\beta_1 = \beta_2 = \cdots = \beta_k = 0$, then the set of vectors is *linearly independent*. Thus for $c = \{0, 0, \ldots 0\}$ and solving $A\cdot\beta = c$ for β, one finds that β itself is a k-length vector of zeros. Otherwise the set is *linearly dependent*, and one of them is a linear combination of the other vectors ([p. 80], Strang 1988).

- rank(A) = n.

The rank of matrix A is the maximum of number of linearly independent rows or columns. Unfortunately, *none* of these definitions, based on the language of linear algebra, provide useful algorithms for deriving the solution vectors, Additional approaches are needed to identify solution algorithms.

Two types of algorithms exist for solving systems of linear equations, direct and iterative methods. Direct methods, in the absence of floating-point, round-off error, would supply an *exact* solution to A*x = b after a finite number of steps; these algorithms are all variants on Gaussian elimination ([p. 31], Demmel 1997). Iterative methods begin with a trial solution, x_0 and then compute a sequence of solution vectors, x_1, x_2, x_3 . of ever more accurate estimates to A*x = b, until the solution meets a predefined convergence criteria ([p. 31], Demmel 1997). Depending on the structure of the particular matrix *A* and the rate at which an iterative solution converges either a direct method or iterative approach may be faster or more accurate, consequently, in the absence of special knowledge of the matrix *A*, Demmel (1997) recommends that a direct method be the method of choice. Similarly, Golub and van Loan ([p. 87], 1997) observe that Gaussian elimination is the algorithm of choice when A is square, dense and unstructured. Consequently, direct methods, based on variants of Gaussian elimination, will be addressed next.

Creating an efficient Gaussian elimination algorithm comes from the following concept. A linear system can be solved by transforming it to a similar system that possesses a solution identical to that of the βoriginal, but the solution of the transformed system is much easier to calculate ([p. 63], Heath, 2002). Well what linear systems are easy to solve?

3.2.2 Essentials of Eigenvalue Problems

The key eigenvalue problem is following. Given a known n-by-n matrix *A* one must determine the scalar λ and the nonzero n-by-1 vector *x* that satisfy

$$A \bullet x = \lambda x,$$

where λ is called the *eigenvalue* and x is termed the *eigenvector*. Eigenvectors and eigenvalues exist only for *square* matrices. The matrix *A* expands or contracts any vector lying in the direction of any of its eigenvectors; the magnitude of such expansions or contractions is given by the corresponding eigenvalue λ ([p. 158] Heath 2002). Observe that eigenvectors are fixed only up to an arbitrary constant; if A·x = λx, then A·(ηx) ≡ λ(ηx) for any scalar, η. Together (λ, x) is termed an *eigenpair* and the total set of eigenvalues comprises the *spectrum* of *A, λ(A)*. The *spectral radius* of the matrix *A*, ρ(A), is defined as the maximum of |λ(A)|.

The above matrix-vector equation, A·x = λx, is equivalent to

$$(A - \lambda I) \bullet x = 0, \qquad \text{[I is the n-by-n identity matrix]}$$

which possesses a *nonzero* solution *x* only if this matrix equation is singular and, consequently,

$$\det(A - \lambda I) = 0,$$

and thus the eigenvalues, λ, are roots of characteristic polynomial[4] of degree, n. Observe that the *Fundamental Theorem of Algebra* requires that the n-by-n matrix *A* has *n* eigenvalues but these *n* eigenvalues do not have to be real or distinct. Note further that complex eigenvalues of real matrices occur in conjugate pairs. Thus if $\alpha + j\beta$ is an eigenvalue, then $\alpha - j\beta$ is also an eigenvalue, here $j = \sqrt{-1}$. Thus

$$\det(\lambda I - A) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) = 0,$$

where the roots (not always real or distinct) $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A. If *m* roots of these roots are identical, then such duplicate eigenvalues are described as having an *algebraic multiplicity m*.

Fundamental eigenvalue properties:

- If $A \cdot x = \lambda x$, and η is an arbitrary scalar, then

$$(A - \eta I) \bullet x = (\lambda - \eta)x.$$

- If A is invertible, and $A \cdot x = \lambda x$, and $x \neq 0$, $\lambda \neq 0$,

$$A^{-1} \bullet x = (1/\lambda)x,$$

  and eigenvalues of the inverse of A are reciprocals of the eigenvalues of A.

- For positive integers *k*

$$A^k \bullet x = \lambda^k x.$$

- The matrix *B* is termed *similar* to the matrix *A*, if a nonsingular matrix *S* exists such that

$$B = S^{-1} \bullet A \bullet S \quad \Leftrightarrow$$
$$B \bullet y = \lambda y \;\Rightarrow\; S^{-1} \bullet A \bullet S \bullet y = \lambda y \;\Rightarrow A \bullet (S \bullet y) = \lambda (S \bullet y),$$

  and the *A* and *B* have the same eigenvalues. If y is an eigenvector of *B*, then $S \cdot y$ is an eigenvector of *A*. Then matrix *B* is labeled *similar* to matrix *A* and $A \rightarrow S^{-1} \cdot A \cdot S$ is termed a *similarity transformation* of *A*. Note if the matrix *A* is real, then the matrix *S* is orthogonal.

- If the columns of the square, nonsingular matrix X are actually eigenvectors, *x*, of the matrix *A*, which has *n* linearly independent eigenvectors, then,

$$A \bullet X = [\lambda_1 x_1 \;\; \lambda_2 x_2 \cdots \lambda_n x_n] = X\Delta,$$

  where Δ is a diagonal matrix with λ_m along its diagonal, $\Delta = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Thus

$$A = X \cdot \Delta \cdot X^{-1} \text{ or, alternatively, } \Delta = X^{-1} \cdot A \cdot X.$$

  This process is termed the *spectral decomposition* of A and any such matrix A is called *diagonalizable*.

[4] Computations of roots of polynomials of order n > 4 do not possess analytic closed forms. Moreover, the numerical determinations of such roots are subject to severe conditioning problems, e.g. Wilkinson polynomial. Consequently, employing characteristic polynomials to ascertain eigenvalues is not recommended.

- Matrices which do not possess a complete set of *n* linearly independent eigenvectors, which are not diagonalizable, are termed *defective*.

- When the matrix *A* is either diagonal or triangular, the eigenvalues are its diagonal elements.

Examining a 2-by-2 matrix system one finds that

$$\det(A - \lambda I) = \det\begin{pmatrix} A_{11}-\lambda & A_{12} \\ A_{21} & A_{22}-\lambda \end{pmatrix} = (A_{11}-\lambda)(A_{22}-\lambda) - A_{12}A_{21} =$$
$$\lambda^2 - (A_{11} + A_{22})\lambda + (A_{11}A_{22} - A_{12}A_{21}) = 0.$$

Setting $\Gamma \equiv (A_{11} - A_{22})^2 + 4A_{12}A_{21}$, if $\Gamma > 0$, then there exists two real eigenvalues,

$$\lambda_{1,2} = 0.5(A_{11} + A_{22} \pm \sqrt{\Gamma}).$$

If $\Gamma = 0$, there is a double real root at $\lambda = 0.5(A_{11} + A_{22})$. Finally, if $\Gamma < 0$, there exists two complex (conjugate) roots at

$$\lambda_{1,2} = 0.5(A_{11} + A_{22} \pm j\sqrt{-\Gamma}).$$

Thus if

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \lambda_1 = 4 \,\&\, \lambda_2 = 2, \qquad B = \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix}, \lambda_1 = +j \,\&\, \lambda_2 = -j.$$

3.3. NORMS

3.3.1 Vector Norms

For a vector, *x*, of length, *n*, the Euclidean length is defined as the following

$$\begin{aligned} \|x\|_2 &= \sqrt{x^T \bullet x} \\ &= (x_1^2 + x_2^2 \cdots + x_n^2)^{1/2} \\ &= \left(\sum_{m=1}^{n} x_m^2 \right)^{1/2}. \end{aligned}$$

This is termed the ℓ_2 *-norm* the standard way to gauge a vector's length. Yet several other definitions of vector length exist:

- ℓ_∞ *-norm,* the magnitude of the largest element in x -

$$\|x\|_\infty = \max_{1 \le m \le n} |x_m|,$$

- ℓ_1 *-norm,* the sum of the magnitudes of the elements in x –

$$\|x\|_1 = \sum_{m=1}^{n} |x_m|,$$

The norm of a vector, represented by the symbols, $\|\bullet\|$, meets the following fundamental requirements:

- $\|x\| > 0$, if $x \neq 0$, $\|x\| = 0$, if and only if $x = 0$,
- $\|\beta x\| = |\beta| \bullet \|x\|$, for any scalar β,
- $\|x + y\| \le \|x\| + \|y\|$, (triangle inequality).

The above three norms are specific examples of the ℓ_p -norms –

$$\|x\|_p \equiv \left(\sum_{i=1}^{i=n} |x_i|^p \right)^{1/p}.$$

```
>> x1 = randi(9999, 1, 3)
x1 =    2785        5469        9575
>> x2 = randi(9999, 1, 3)
x2 =    9648        1576        9705
>> dx = x1 - x2
dx =   -6863        3893        -130
>> norm(dx, 1)
ans =  10886
>> norm(dx, 2)
ans =  7891.3
>> norm(dx, 3)
ans =  6863
```

As can be seen from these MATLAB calculations –

$$\|x\|_1 \ge \|x\|_2 \ge \|x\|_1$$
$$\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\,\|x\|_\infty$$
$$\|x\|_\infty \le \|x\|_1 \le n\,\|x\|_\infty .$$

3.3.2 Matrix Norms

Given these vector norms, one defines corresponding matrix norms of an m-by-n matrix, *A*,

$$\|A\| \equiv \max_{x \neq 0} \frac{\|A \bullet x\|}{\|x\|}.$$

A matrix norm is defined to be *induced* from its corresponding vector norm ([p. 55] Heath 2002; [p. 76] Ascher & Greif 2011]. Matrix norm measures the maximum stretching the matrix performs on any vector ([p. 55] Heath 2002).

Some matrix norms are simpler to compute that others -

- *1-norm* is the maximum absolute column sum of A -

$$\|A\|_1 = \max_j \sum_{i=1}^{m} |A_{ij}|,$$

- ∞*-norm,* is the maximum absolute row sum of A -

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |A_{ij}|,$$

- 2*-norm, or Euclidean norm,* is not as *simple* to calculate*;* it is related to the spectral radius of the matrix product[5] $A^T \cdot A$ ([p. 78] Ascher & Greif 2011) -

$$\|A\|_2 = \sqrt{\rho(A^T \bullet A)}.$$

Consequently, employing MATLAB,

```
>> A = randi(9999, 3, 3) - randi(9999, 3, 3)
A =    -2272        6372        1970
         989        2448        9078
         332        4828        -887
>> norm(A, 1)
ans =  13648
>> sum(abs(A), 1)         % Here 1 refers summing over columns
ans =   3593       13648       11935
>> max(sum(abs(A), 1))
ans =  13648
>> norm(A, inf)
ans =  12515
>> sum(abs(A), 2)
ans =  10614
       12515
        6047
>> max(sum(abs(A), 2)) % Here 2 refers to summing over rows
ans =  12515
>> norm(A, 2)          % Here 2 refers to the Euclidean norm
ans =  10500
>> C = transpose(A)*A
C =  6250329   -10453216     4207818
   -10453216    69904672    30493348
     4207818    30493348    87077753
>> e = eig(C)
e =  3.4422e+006
     4.9535e+007
     1.1026e+008
>> spectral = max(abs(e))
```

[5] If A is rectangular m-by-n matrix, A^T is an n-by-m matrix, and the product $A^T \cdot A$ is square, possessing a size of n-by-n. Hence, $A^T \cdot A$ possesses eigenvalue spectrum.

```
spectral =  1.1026e+008
>> sqrt(spectral)
ans = 10500
```

Thus such matrix norms satisfy the following properties, where ***A*** and *B* represent two different matrices of the same size for purposes of illustration:

- $\|A\| > 0$, if $A \neq 0$,
- $\|\gamma A\| = |\gamma| \bullet \|A\|$, for any scalar γ,
- $\|A + B\| \leq \|A\| + \|B\|$, (triangle inequality).
- $\|A \bullet B\| \leq \|A\| \bullet \|B\|$,
- $\|A \bullet x\| \leq \|A\| \bullet \|x\|$, for any vector x.

3.4. SPECIAL CLASSES OF MATRICES

3.4.1 Positive Definite Symmetric Matrices

A square matrix is defined to be symmetric if $A^T = A$. The matrix *A* is termed *positive definite* if

$$x^T \bullet A \bullet x > 0, \text{ for all } x \neq 0.$$

A symmetric matrix must have real eigenvalues. If *A* is also positive definite matrix, for any eigenpair (λ_m, x_m) , one has

$$0 < x^T \bullet A \bullet x = \lambda_m x_m \bullet x_m = \lambda_m \sum_{k=1}^{n} x_k^2,$$

since the summation is positive, then any λ_m must be also be positive .

3.4.2 Orthogonal Matrices

Two vectors, *u*, and, *w*, of identical lengths are termed *orthogonal* if their product equals zero -

$$u^T \cdot w = 0.$$

If both vectors are unit length, $||w||_2 = 1$ and $||u||_2 = 1$, these vectors are called *orthonormal*.

A real square matrix, *Q*, is called *orthogonal*, if its columns are orthogonal,

$$Q^T \bullet Q = I.$$

Thus for any vector, x, with ℓ_2- *norm*, $\|Q \bullet x\| \equiv \|x\|$ since

$$\|Q \bullet x\|^2 = (Q \bullet x)^T \bullet (Q \bullet x) = x^T \bullet Q^T \bullet Q \bullet x = x^T \bullet x = \|x\|^2.$$

This norm preserving property is critical in future matrix (least-squares) computations.

3.5 APPENDICES

A.1 Transpose

The transpose of an m-by-n A, termed, A^T, is the n-by-m matrix found by interchanging the rows and columns of A => $A^T_{ij} = A_{ji}$.

$$B = \begin{bmatrix} 4 & 7 & 3 \\ 3 & 9 & 7 \end{bmatrix}, \quad B^T = \begin{bmatrix} 4 & 3 \\ 7 & 9 \\ 3 & 7 \end{bmatrix}$$

```
>> F = B'   (transpose is a single quote)  MATLAB
```

Hermitian transpose of an m-by-n A matrix, is represented by , A^H, the n-by-m matrix found by taking the complex conjugate of the elements of A and interchanging the rows and columns of A $A^H_{ij} = A^*_{ji}$.

$$A = \begin{bmatrix} (3+2i) & (9-3i) \\ (7-8i) & (1-2i) \end{bmatrix}, \quad A^H = \begin{bmatrix} (3-2i) & (7+8i) \\ (9+3i) & (1+2i) \end{bmatrix}$$

Laws for transpose:

$$(A^T)^T = A \quad \text{and} \quad (A^H)^H = A.$$

$$(A \pm B)^T = A^T \pm B^T \quad \text{and} \quad (A \pm B)^H = A^H \pm B^H$$

$$(\beta \bullet A)^T = \beta \bullet A^T \quad \text{and} \quad (\beta \bullet A)^H = \beta \bullet A^H$$

$$(A \bullet B)^T = B^T \bullet A^T \quad \text{and} \quad (A \bullet B)^H = B^H \bullet A^H \quad \text{[Note: reverse order!]}$$

$$(A \bullet B \bullet C)^T = C^T \bullet B^T \bullet A^T$$

NOTE: a *product* of two symmetric or Hermitian matrices is ***not*** in general symmetric or Hermitian.

A.2 Inverses

Inverse is defined **ONLY** for square matrices!

```
>>  P = inv(A);     % MATLAB command
```

Any matrix, ***X,*** such that ***X*A*** = ***I*** and ***A*X*** = ***I*** is called the **inverse** of A.

A square matrix ***A*** is termed nonsingular if it has an inverse ***X***. The inverse ***X*** is called A^{-1}. A square matrix, which does not possess an inverse, is termed singular.

Rules for matrix inverse:

$A \bullet B$ is nonsingular, $(A \bullet B)^{-1} = B^{-1} \bullet A^{-1}$ [NOTE: reverse order!]

A^{-1} is nonsingular & $(A^{-1})^{-1} = A$

A^{T} is nonsingular with $(A^{T})^{-1} = (A^{-1})^{T}$

A^{H} is nonsingular with $(A^{H})^{-1} = (A^{-1})^{H}$

NOTE: product of nonsingular matrices is nonsingular, and order of the product is reversed –

$$(A \bullet B \bullet C)^{-1} = C^{-1} \bullet B^{-1} \bullet A^{-1}$$

A.3 Permutations of Matrices

A permutation matrix is a square matrix that only one nonzero entry (value of 1) in each column and rows. The following are some examples:

$$P_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Define a general 3-by-3 matrix:

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix}$$

Action of the permutation matrix:

$\Rightarrow$ P·A permutes rows of A

$\Rightarrow$ A·P permutes columns of A

$$P_1 \bullet A = \begin{bmatrix} A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \\ A_{11} & A_{12} & A_{13} \end{bmatrix} = \begin{bmatrix} 2^{nd}\text{row of A} \\ 3^{rd}\text{row of A} \\ 1^{st}\text{row of A} \end{bmatrix}$$

$$A \bullet P_1 = \begin{bmatrix} A_{13} & A_{11} & A_{12} \\ A_{23} & A_{21} & A_{22} \\ A_{33} & A_{31} & A_{32} \end{bmatrix} = \left(3^{rd}\text{column}, 1^{st}\text{column}, 2^{nd}\text{column}\right)$$

Inverse of permutation matrix is its transpose, $P^{-1} = P^T$. `P' == inv(P) % MATLAB`

Product of two permutation matrices is also a permutation matrix.

A.4 Determinants ⇐ Not used for any calculations in this course!

```
>> det(B) % MATLAB
```

A *determinant* is a *single* number associated with an n-by-n matrix, ***A***, and it is denoted by det(***A***). The determinant is defined *recursively* in terms of a determinant of an (n - 1)-by-(n - 1) matrix, which in turn is determined by the determinant of an (n - 2)-by-(n - 2) matrix, and so on down to a determinant of 1-by-1 matrix [β], which is defined as det([β]) = β.

Definitions:

The (I,j)-*minor* of ***A***, termed $\boldsymbol{M}_{ij}$, is the determinant of the (n - 1)-by-(n - 1) matrix created from ***A*** by deleting the ith row and jth column of ***A***.

The (I,j)-*cofactor* of ***B***, denoted $\boldsymbol{B}_{ij}$, is $(-1)^{i+j} M_{ij}$.

Determinant of ***A*** is defined to be the sum of the products of the elements of the first row and the cofactors of the first row.

1-by-1 Array: $A = [\beta], \quad \det(A) = \beta$

2-by-2 array $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$

$$\det(A) = A_{11}A_{22} - A_{12}A_{21}$$

3-by3 Array $A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix},$

$$\det(A)=A_{11}\det(A_{11})+A_{12}\det(A_{12})+A_{13}\det(A_{13})=$$

$$A_{11}\det\begin{bmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{bmatrix}-A_{12}\det\begin{bmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{bmatrix}+A_{13}\det\begin{bmatrix} A_{21} & A_{22} \\ A_{31} & A_{32} \end{bmatrix} =$$

$$\det(A)=A_{11}(A_{22}A_{33}-A_{23}A_{32})-A_{12}(A_{21}A_{33}-A_{23}A_{31})+A_{13}(A_{21}A_{32}-A_{22}A_{31})$$

Properties of determinants:

1) $\det(A) = \det(A^T)$.
2) If any row or column of A = 0, then $\det(A) = 0$.
3) For any scalar β and for an n-by-n array A, $\det(\beta A) = \beta^n \det(A)$.
4) If A has two rows (columns) equal, $\det(A) = 0$.
5) If A' is created from A by interchanging two rows, $\det(A') = -\det(A)$.
6) If A' is created from A by replacing a row (column) of A by the number α times that row(column) itself, $\det(A') = \alpha\det(A)$.
7) If A' is created from A by replacing a row (column) of A by that row(column) plus some multiple of a different row, $\det(A') = \det(A)$.

It can be shown that determinant for triangular array equals solely product of diagonal entries.

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{bmatrix},\ A_{21}= 0,\ A_{31} = 0,\ A_{32} =0.$$

General solution for the determinant of a 3-by-3 matrix:

$$\det(A) = A_{11}(A_{22}A_{33}-A_{23}A_{32}) - A_{12}(A_{21}A_{33}-A_{23}A_{31}) + A_{13}(A_{21}A_{32}-A_{22}A_{31})$$
$$= A_{11}(A_{22}A_{33}-A_{23}*0) - A_{12}(0*A_{33}-A_{23}*0) + A_{13}(0*A_{32}-A_{22}*0)$$
$$= A_{11}A_{22}A_{33}$$

Determinants and Inverses:

1) If A is nonsingular, $\det(A) \neq 0$. If A is singular, $\det(A) = 0$.

2) If A is nonsingular, $\det(A^{-1}) = 1/\det(A)$.

3) $\det(I) = 1$.

4) $\det(AB) = \det(A)\det(B)$.

3.6 REFERENCES

Ascher, U. M. and Greif, C. 2011, *A First Course in NUMERICAL METHODS* (SIAM: Philadelphia, PA).

Datta, B. N. 1995, *Numerical Linear Algebra and Applications* (Brooks/Cole: Pacific Grove CA).

Demmel, J. W. 1997, *Applied Numerical Linear Algebra* (SIAM: Philadelphia, PA).

Golub, G. H. and van Loan, C. H. 1996, *Matrix Computations, 3rd Ed.* (Johns Hopkins: Baltimore)

Heath, M. T. 2002, *Scientific Computing: An Introductory Survey*, 2nd Ed. (McGraw-Hill: NY).

Higham, N. J. 2002, *Accuracy and Stability of Numerical Algorithms 2nd Ed,* (SIAM: Philadelphia, PA).

Kiusalaas, J, 2005, *Numerical Methods in Engineering with MATLAB®* (Cambridge University Press, Cambridge, UK).

Meyers, C. 2001, *Matrix Analysis and Applied Linear Algebra* (SIAM: Philadelphia).

Moler, C. B.2004, *Numerical Computing with MATLAB* (SIAM: Philadelphia).

Trefethen, L. N,, and Bau, D. 1997, *Numerical Linear Algebra* (SIAM: Philadelphia).

Strang, G. 1988, *Linear Algebra and its Applications 3rd Ed.* (Cengage Learning: Florence KY).

Watkins, D. S. 2002, *Fundamentals of Matrix Computations 2nd Ed.* (John Wiley: NY)

3.7 QUESTIONS

1. True or false: If matrix A is nonsingular, then the number of solutions to the linear system $A \cdot x = b$ depends on the particular choice of right-hand-side vector b.

2. True or false: If a matrix has a very small determinant, then the matrix is nearly singular.

3. True or false: If $||A|| = 0$, then $A = 0$.

4. Can a system of linear equations $Ax\ b$ have exactly two distinct solutions?

5. *(a)* Is the magnitude of the determinant of a matrix a good indicator of whether the matrix is nearly singular? *(b)* If so, why? If not, what is a better indicator of near singularity?

6. Which of the following are good indicators that a matrix is nearly singular?
(a) Its determinant is small.
(*b)* Its norm is small.
(c) Its norm is large.

7. What is a singular matrix?