3-27-2012

LINEAR ALGEBRA & MATRICES

Chaouki T. Abdallah

M. Ariola

F. Amato

LINEAR ALGEBRA & MATRICES

C.T. Abdallah, M. Ariola, F. Amato

January 21, 2006

Contents

LINEAR ALGEBRA & MATRICES

These notes deal with the study of linear Algebra and matrices. Linear Algebra plays an important role in the subareas of signal processing, control systems, communications, and more broadly in the studies of systems. The notes rely heavily on [1, 2, 3, 4].

1 Review of Matrix Algebra

We start this chapter by introducing matrices and their algebra. Notationally, matrices are denoted by capital letters A, M, Γ and are rectangular arrays of elements . Such elements are referred to as scalars and denoted by lowercase letters, $a, b, \alpha, etc.$. Note that the scalars are not necessarily real or complex constants: they maybe real or complex numbers, polynomials, or general functions.

Example 1 Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & -3 & 4 & 5 & 6 \\ 2\times 10^{-5} & -1 & 0.9 & -7.2\times 10^{-4} & -0.17 & -4.96\times 10^{-3} \\ 0.012338 & 11.72 & -2.6 & 8.7\times 10^{-4} & -31 & 22 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -30 \end{bmatrix}$$

This is a 5×6 matrix with real entries. On the other hand,

$$P(s) = \begin{bmatrix} s+1 & s+3 \\ s^2+3s+2 & s^2+5s+4 \end{bmatrix}$$

is a 2×2 matrix of polynomials in the variable s. Finally,

$$\Gamma = \begin{bmatrix} 1+j & -j & 0 \\ -1+3j & 5 & 2-j \end{bmatrix}$$

is a 2×3 complex matrix.

△

A matrix which has m rows and n columns, is said to be $m \times n$. A matrix may be denoted by $A = [a_{ij}]$ where $i = 1, \cdots m$ and $j = 1, \cdots n$.

Example 2 The matrix A in Example 1 has 5 rows and 6 columns. The element $a_{11} = 1$ while $a_{35} = -31$.

△

A scalar is a 1×1 matrix. If $m = n$ the matrix is said to be square. If $m = 1$ the matrix is a row matrix (or vector). If $n = 1$ the matrix is an m column matrix (or vector). If A is square then we define the *Trace* of A by the sum of its diagonal elements, i.e.

$$Trace(A) = \sum_{i=1}^{n} a_{ii} \tag{1}$$

Example 3 Consider $P(s)$ of Example 1, then $Trace(P(s)) = s^2 + 6s + 5$.

$\triangle$

Two matrices A and B are equal, written $A = B$, if and only if A has the same number of rows and columns as B and if $a_{ij} = b_{ij}$ for all i, j. Two matrices A and B that have the same numbers of rows and columns may be added or subtracted element by element, i.e.

$$C = [c_{ij}] = A \pm B \Longleftrightarrow c_{ij} = a_{ij} \pm b_{ij} \ \forall i, j \tag{2}$$

the multiplication of 2 matrices A and B to obtain $C = AB$ may be performed if and only if A has the same number of columns as B has rows. In fact,

$$C = AB \Longleftrightarrow c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{3}$$

the matrix C is $m \times q$ if A is $m \times n$ and B is $n \times q$. Note that BA may not even be defined and in general, BA is not equal to AB even when both products are defined. The matrix of all zeros is the Null matrix, and the square matrix A with $a_{ii} = 1$ and $a_{ij} = 0$ for $i \neq j$ is the identity matrix. The identity matrix is denoted by I. Note that $AI = IA = A$ assuming A is $n \times n$ as is I.

The following properties of matrix Algebra are easily verified

1. $A \pm B = B \pm A$
2. $A + (B + C) = (A + B) + C$
3. $\alpha(A + B) = \alpha A + \alpha B$ for all scalars α.
4. $\alpha A = A \alpha$
5. $A(BC) = (AB)C$
6. $A(B + C) = AB + AC$

7. $(B+C)A = BA + CA$

Let $A = [a_{ij}]$ be an $m \times n$ matrix. The transpose of A is denoted by $A^T = [a_{ji}]$ and is the $n \times m$ matrix obtained by interchanging columns and rows. The matrix A is *symmetric* if $A = A^T$, *skew-symmetric* if $A = -A^T$. Also, it can be seen that $(AB)^T = B^T A^T$. In the case where A contains complex elements, we let $\bar{A}$ be the *conjugate* of A whose elements are the conjugates of those of A. Matrices satisfying $A = \bar{A}^T$ are *Hermitian* and those satisfying $A = -\bar{A}^T$ are *skew-Hermitian*.

1.1 Determinants, Minors, and Cofactors

In this section we will consider square matrices only. The determinant of a square matrix A denoted by $det(A)$ or $\mid A \mid$ is a scalar-valued function of A and is given as follows for $n = 1, 2, 3$.

1. $n = 1$, then $\mid A \mid = a_{11}$
2. $n = 2$, then $\mid A \mid = a_{11}a_{22} - a_{12}a_{21}$
3. $n = 3$, then $\mid A \mid = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(a_{23}a_{31} - a_{21}a_{33}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$

Standard methods apply for calculating determinants. The determinant of a square matrix is 0, that of an identity matrix is 1, and that of a triangular or diagonal matrix is the product of all diagonal elements. Each element a_{ij} in an $n \times n$ square matrix A has associated with it a minor M_{ij} obtained as the determinant of the $(n-1) \times (n-1)$ matrix resulting form deleting the ith row and the jth column. From these minors, we can obtain the cofactors given by $C_{ij} = (-1)^{i+j} M_{ij}$.

Example 4 Given

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

Then, $M_{12} = 5$, $C_{12} = -5$, $M_{32} = 1 = -C_{32}$.

△

Some useful properties of determinants are listed below:

1. Let A and B be $n \times n$ matrices, then $\mid AB \mid = \mid A \mid . \mid B \mid$.
2. $\mid A \mid = \mid A^T \mid$

3. If A contains a row or a column of zeros, then $\mid A \mid= 0$.
4. If any row (or column) of A is a linear combination of other rows (or columns), then $\mid A \mid= 0$.
5. If we interchange any 2 rows (or columns) of a matrix A, the determinant of the resulting matrix is $- \mid A \mid$.
6. If we multiply a row (or a column) of a matrix A by a scalar α, the determinant of the resulting matrix is $\alpha \mid A \mid$.
7. Any multiple of a row (or a column) may be added to any other row (or column) of a matrix A without changing the value of the determinant.

1.2 Rank, Trace, and Inverse

The rank of an $m \times n$ matrix A denoted by $r_A = rank(A)$ is the size of the largest nonzero determinant that can be formed from A. Note that $r_A \leq min\{m,n\}$. If A is square and if $r_A = n$, then A is nonsingular. If r_A is the rank of A and r_B is the rank of B, and if $C = AB$, then $0 \leq r_C \leq min\{r_A, r_B\}$. The trace of a square matrix A was defined earlier as $Tr(A) = \sum_{i=1}^{n} a_{ii}$.

Example 5 Let us look at the rank of:

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

It is easy to see that $r_A = 3$. Now consider the matrix

$$B = \begin{bmatrix} 2 & 0 \\ 3 & 0 \\ 1 & 1 \end{bmatrix}$$

whose rank is $r_B = 2$ and form

$$A.B = \begin{bmatrix} 17 & 1 \\ 8 & 2 \\ 1 & 1 \end{bmatrix}$$

and the $rank(C) = 2 \leq \{r_A, r_B\}$.

$\triangle$

If A and B are conformable square matrices, $Tr(A+B) = Tr(A) + Tr(B)$, and $Tr(AB) = Tr(BA)$. Also, $Tr(A) = Tr(A^T)$.

Example 6 Show that $Tr(ABC) = Tr(B^T A^T C^T)$. First, write $Tr(ABC) = Tr(CAB)$, then $Tr([CAB]^T) = Tr(CAB)$, thus proven.

△

Note that $rank(A+B) \neq rank(A) + rank(B)$ and that $Tr(AB) \neq Tr(A) \times Tr(B)$.

Example 7 Let

$$A = \begin{bmatrix} 2 & 4 & 1 \\ 3 & 0 & 2 \\ 2 & 0 & 3 \end{bmatrix}; \; B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

and note that $r_A = 3$, $r_B = 1$, while $r_{A+B} = 3$. Also, note that $Tr(AB) = 2$ while $Tr(A) \times Tr(B) = 5 \times 2 = 10$.

△

Next, we define the inverse of a square, nonsingular matrix A as the square matrix B of the same dimensions such that

$$AB = BA = I$$

the inverse is denoted by A^{-1} and may be found by

$$A^{-1} = \frac{C^T}{|A|}$$

where C is the matrix of cofactors C_{ij} of A. Note that for the inverse to exist, $|A|$ must be nonzero, which is equivalent to saying that A is nonsingular. We also write $C^T = Adjoint(A)$.

The following property holds:

$$(AB)^{-1} = B^{-1}A^{-1}$$

assuming of course that A and B are compatible and both invertible.

1.3 Elementary Operations and Matrices

The basic operations called *elementary operations* are as follows

1. Interchange any 2 rows (or columns)
2. Multiply any row (or column) by a scalar α
3. Multiply any row (or column) by a scalar α and add the resulting row (or column) to any other row (or column)

Note that each of these elementary operations may be represented by a postmultiplication for column operations (or premultiplication for row operations) by elementary matrices which are nonsingular.

Example 8 Let

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 1 & 1 & -1 \end{bmatrix}$$

Then, let use the following row operations:

1. Exchange rows 1 and 3
2. Multiply row 1 by -2 and add it to row 3
3. Multiply row 2 by 1/2 and add it to row 3
4. Multiply row 2 by 1/2
5. Multiply row 3 by 2/5
6. Multiply column 1 by -1 and add it to column 2
7. Multiply column 1 by 1 and add it to column 3
8. Multiply column 2 by -1/2 and add it to column 3

The end result is

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$\triangle$

Example 9 Given the following polynomial matrix

$$P(s) = \begin{bmatrix} s^2 & 0 \\ 0 & s^2 \\ 1 & s+1 \end{bmatrix}$$

then the corresponding row operations are performed:

1. interchange rows 3 and 1.
2. multiply row 1 by $-s^2$ and add it to row 3
3. multiply row 2 by $s+1$ and add it to row 3.

This corresponds to the following multiplications by the matrices given:

1.

$$P_1(s) = \begin{bmatrix} 1 & s+1 \\ 0 & s^2 \\ s^2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} P(s)$$

2.

$$P_2(s) = \begin{bmatrix} 1 & s+1 \\ 0 & s^2 \\ 0 & -s^2(s+1) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -s^2 & 0 & 0 \end{bmatrix} P_1(s)$$

3.

$$P_3(s) = \begin{bmatrix} 1 & s+1 \\ 0 & s^2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & s+1 & 0 \end{bmatrix} P(s)$$

$\triangle$

In general, for any matrix of rank r we can reduce it via column and row operations to one of the following normal forms

$$I_r,\ [I_r \mid 0],\ [I_r \mid 0]^T, or \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] \tag{4}$$

Next, we discuss vector spaces, or as they are sometimes called, linear space..

2 Vectors and Linear (Vector) Spaces

In most of our applications, we need to deal with linear real and complex vector spaces which are defined subsequently.

Definition 1 A *real linear vector space* (resp. *complex linear vector space* is a set V, equipped with 2 binary operations: the addition (+) and the scalar multiplication (.) such that

1. $x+y=y+x,\ \forall x,y\in V$
2. $x+(y+z)=(x+y)+z,\ \forall x,y,z\in V$
3. There is an element 0_V in V such that $x+0_V=0_V+x=x,\ \forall x\in V$
4. For each $x\in V$, there exists an element $-x\in V$ such that $x+(-x)=(-x)+x=0_V$
5. For all scalars $r_1,r_2\in\mathbb{R}$ (resp. $c_1,c_2\in\mathbb{C}$), and each $x\in V$, we have $r_1.(r_2.x)=(r_1r_2).x$ (resp. $c_1.(c_2.x)=(c_1c_2).x$
6. For each $r\in\mathbb{R}$ (resp. $c\in\mathbb{C}$), and each $x_1,x_2\in V$, $r.(x_1+x_2)=r.x_1+r.x_2$ (resp. $c.(x_1+x_2)=c.x_1+c.x_2$)
7. For all scalars $r_1,r_2\in R$ (resp. $c_1,c_2\in C$), and each $x\in V$, we have $(r_1+r_2).x=r_1.x+r_2.x$ (resp. $(c_1+c_2).x=c_1.x+c_2.x$)
8. For each $x\in V$, we have $1.x=x$ where 1 is the unity in $\mathbb{R}$ (resp. in $\mathbb{C}$).

Example 10 The following are linear vector spaces with the associated scalar fields: $\mathbb{R}^n$ with $\mathbb{R}$, $\mathbb{C}^n$ with $\mathbb{C}$.

$\triangle$

Definition 2 A subset M of a vector space V is a subspace if it is a linear vector space in its own right. One necessary condition for M to be a subspace is that it contains the zero vector.

Let $x_1,\dots x_n$ be some vectors in a linear space $\mathscr{X}$ defined over a field $\mathbb{F}$. In these notes we shall consider either the real field $\mathbb{R}$ or the complex field $\mathbb{C}$. The *Span* of the vectors $x_1,\dots x_n$ over the field $\mathbb{F}$ is defined as

$$\mathrm{Span}_{\mathbb{F}}\{x_1,\dots x_n\}:=\{x\in\mathscr{X}:x=\sum_{i=1}^{n}a_ix_i,\ a_i\in\mathbb{F}\}.$$

The vectors $x_1, \ldots x_n$ are said to be *linearly independent* if

$$\sum_{i=1}^{n} a_i x_i = 0 \Rightarrow a_i = 0,\ i = 1, \ldots n$$

otherwise they are *linearly dependent*. A set of vectors $x_1, \ldots x_m$ is a *basis* of a linear space $\mathscr{X}$ if the vectors are linearly independent and their Span is equal to $\mathscr{X}$. In this case the linear space $\mathscr{X}$ is said to have finite *dimension* m.

Note that any set of vectors containing the zero vector is linearly dependent. Also, if $\{x_i\}$ $i = 1, \cdots, n$ is linearly dependent, adding a new vector x_{n+1} will not make the new set linearly independent. Finally, if the set $\{x_i\}$ is linearly dependent, then, at least one of the vectors in the set may be written as a linear combination of the others.

How do we test for linear dependence/independence? Given a set of n vectors $\{x_i\}$ each having m components. Form the $m \times n$ matrix A which has the vectors x_i as its columns. Note that if $n > m$ then the set of vectors has to be linearly independent. Therefore, let us consider the case where $n \leq m$. Form the $n \times n$ matrix $G = A^T A$ and check if it is nonsingular, i.e. if $\mid G \mid \neq 0$, then $\{x_i\}$ is linearly independent, otherwise $\{x_i\}$ is linearly dependent.

Geometrically, linear independence may be explained in $\mathbb{R}^n$. Consider the plane $\mathbb{R}^2$ and suppose we have 2 vectors x_1 and x_2. If the 2 vectors are linearly dependent, then $x_2 = a x_1$ or both vectors lie along the same direction. If they were linearly independent, then they form the 2 sides of a parallelogram. Therefore, in the case of linear dependency, the parallelogram degenerate to a single line. We can equip a vector space with many functions. One of which is the inner product which takes two vectors in V to a scalar either in $\mathbb{R}$ or in $\mathbb{C}$, the other one is the norm of a vector which takes a vector in V to a positive value in $\mathbb{R}$.

Given two linear spaces $\mathscr{X}$ and $\mathscr{Y}$ over the same field $\mathbb{F}$, a function $\mathscr{A} : \mathscr{X} \mapsto \mathscr{Y}$ is a linear transformation if $\mathscr{A}(ax + by) = a\mathscr{A}(x) + b\mathscr{A}(y)$, for all $a,\ b\ \in\ \mathbb{F}$. Let $\mathscr{A}$ be a linear transformation $\mathscr{A} : \mathscr{X} \mapsto \mathscr{X}$ and $\mathscr{Y}$ a linear subspace $\mathscr{Y} \subseteq \mathscr{X}$. The subspace $\mathscr{Y}$ is said to be $\mathscr{A}-$*invariant* if

$$\mathscr{A}(y) \in \mathscr{Y},\ \forall y \in \mathscr{Y}.$$

Given a linear transformation, we define the *Range* or *Image* of $\mathscr{A}$ as the subspace of $\mathscr{Y}$

$$\text{Range}(\mathscr{A}) := \{y \in \mathscr{Y} : y = \mathscr{A}(x),\ x \in \mathscr{X}\},$$

and the *Kernel* or *Null-Space* of $\mathscr{A}$ as the subspace of $\mathscr{X}$

$$\mathscr{N}(\mathscr{A}) := \text{Ker}(\mathscr{A}) := \{x \in \mathscr{X} : \mathscr{A}(x) = 0\}.$$

Given $\mathscr{X} = \mathbb{C}^m$ and $\mathscr{Y} = \mathbb{C}^n$, a matrix $A \in \mathbb{C}^{m \times n}$, denoted by $A = [a_{ij};\ 1 \leq i \leq m,\ 1 \leq j \leq n$, is an example of linear transformation.

Example 11 The set $\{e_i\}$ $i = 1, \cdots, n$ is linearly independent in $\mathbb{R}^n$. On the other hand, the 2 vectors, $x_1^T = [1\ -2]$; $x_2^T = [2\ -4]$ are linearly dependent in $\mathbb{R}^2$.

△

Example 12 Let X be the vector space of all vectors $x = [x_1\ \cdots\ x_n]^T$ such that all components are equal. Then X is spanned by the vector of all 1's. Therefore dim (X) = 1. On the other hand, if X is the vector space of all polynomials of degree $n-1$ or less, a basis is $\{1, t, t^2, \cdots, t^{n-1}\}$ which makes dim (X) = n.

△

2.1 Linear Equations

In many control problems, we have to deal with a set of simultaneous linear algebraic equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= y_2 \\ \vdots &= \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= y_m \end{aligned} \tag{5}$$

or in matrix notation

$$Ax = y \tag{6}$$

for an $m \times n$ matrix A, an $n \times 1$ vector x and an m vector y. The problem is to find the solution vector x, given A and y. Three cases might take place:

1. No solution exist
2. A unique solution exists
3. An infinite number of solutions exist

Now, looking back at the linear equation $Ax = y$, it is obvious that y should be in $\mathscr{R}(A)$ for a solution x to exist, in other words, if we form

$$W = [A \mid y] \tag{7}$$

then $rank(W) = rank(A)$ is a necessary condition for the existence of at least one solution. Now, if $z \in \mathscr{N}(A)$, and if x is any solution to $Ax = y$, then $x + z$ is also a solution. Therefore, for a unique solution to exist, we need that $\mathscr{N}(A) = 0$. That will require that the columns of A form a basis of $\mathscr{R}(A)$, i.e. that there will be n of them and that they will be linearly independent, and of dimension n. Then, the A matrix is invertible and $x = A^{-1}y$ is the unique solution.

3 Eigenvalues and Eigenvectors

Let A be an $n \times m$ matrix, and denote the corresponding identity matrix by I. Then, let x_i be a nonzero vector in $\mathbb{R}^n$ and λ_i be scalar such that

$$Ax_i = \lambda x_i \tag{8}$$

Then, λ_i is an eigenvalue of A and x_i is the corresponding eigenvector. There will be n eigenvalues of A (some of which redundant). In order to find the eigenvalues of A we rewrite the previous equation

$$(A - \lambda_i I)x_i = 0 \tag{9}$$

Noting that x_i can not be the zero vector, and recalling the conditions on the existence of solutions of linear equations, we see that we have to require

$$det(A - \lambda_i I) = \mid (A - \lambda_i I) \mid = 0 \tag{10}$$

We thus obtain an nth degree polynomial, which when set to zero, gives the characteristic equation

$$\begin{aligned} \mid (A - \lambda I) \mid &= \Delta(\lambda) = 0 \\ &= (-\lambda)^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 \\ &= (-1)^n(\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_p)^{m_p} \end{aligned} \tag{11}$$

Therefore, λ_i is an eigenvalue of A of algebraic multiplicity m_i. One can then shows,

$$\begin{aligned} Tr(A) &= \sum_{i=1}^{n} = (-1)^{n+1}c_{n-1} \\ \mid A \mid &= \prod_{i=1}^{n} \lambda_i = c_0 \end{aligned} \tag{12}$$

Also, if λ_i is an eigenvalue, then so is λ_i^*.

How do we determine eigenvectors? We distinguish 2 cases:

1. All eigenvalues are distinct: In this case, we first find $Adj(A - \lambda I)$ with λ as a parameter. Then, successively substituting each eigenvalue λ_i and selecting any nonzero column gives all eigenvectors.

   Example 13 Let

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -18 & -27 & -10 \end{bmatrix}$$

then, solve for

$$|A-\lambda I| = -\lambda^3 - 10\lambda^2 - 27\lambda - 18 = 0$$

which has 3 solutions, $\lambda_1 = -1$, $\lambda_2 = -3$, $\lambda_3 = -6$. Now, let us solve for the eigenvectors using the suggested method

$$Adj(A-\lambda I) = \begin{bmatrix} \lambda^2+10\lambda+27 & \lambda+10 & 1 \\ -18 & \lambda^2+10\lambda & \lambda \\ -18\lambda & -27\lambda-18 & \lambda^2 \end{bmatrix}$$

so that for $\lambda_1 = -1$ we can see that column 1 is

$$x_1 = [18 \quad -18 \quad 18]^T$$

Similarly, $x_2 = [7 \quad -21 \quad 63]^T$ and $x_3 = [1 \quad -6 \quad 36]^T$. There is another method of obtaining the eigenvectors from the definition by actually solving $Ax_i = \lambda_i x_i$

△

Note that once all eigenvectors are obtained, we can arrange them in an $n \times n$ modal matrix $M = [x_1 \quad x_2 \cdots x_n]$. Note that the eigenvectors are not unique since if x_i is an eigenvector, then so is any αx_i for any scalar α.

2. Some eigenvalues are repeated: In this case, a full set of independent eigenvectors may or may not exist. Suppose λ_i is an eigenvalue with an algebraic multiplicity m_i. Then, the dimension of the Null space of $A-\lambda_i I$ which is also the number of linearly independent eigenvectors associated with λ_i is the geometric multiplicity q_i of the eigenvalue λ_i. We can distinguish 3 cases

   (a) Fully degenerate case $q_i = m_i$: In this case there will be q_i independent solutions to $(A-\lambda_i I)x_i$ rather than just one.

   Example 14 Given the matrix

$$A = \begin{bmatrix} 10/3 & 1 & -1 & -1/3 \\ 0 & 4 & 0 & 0 \\ -2/3 & 1 & 3 & -1/3 \\ -2/3 & 1 & -1 & 11/3 \end{bmatrix}$$

   Then, its characteristic equation is

$$\Delta(\lambda) = \lambda^4 - 14\lambda^3 + 72\lambda^2 - 160\lambda + 128 = 0$$

There are 4 roots, $\lambda_1 = 2$, $\lambda_2 = \lambda_3 = \lambda_4 = 4$. We can find the eigenvector associated with λ_1 as $x_1 = [-1\ 0\ -1\ -1]^T$. Then, we find

$$(4I - A)x = \begin{bmatrix} 2/3 & -1 & 1 & 1/3 \\ 0 & 0 & 0 & 0 \\ 2/3 & -1 & 1 & 1/3 \\ 2/3 & -1 & 1 & 1/3 \end{bmatrix} x$$

There are an infinite number of solutions, 3 of which are $x_2 = [1\ 0\ 0\ -2]^T$; $x_3 = [0\ 1\ 1\ 0]^T$; $x_4 = [0\ 1\ 0\ 3]^T$

$\triangle$

Note that now if $x_1, \cdots x_m$ are eigenvectors for eigenvalue λ_i, then so is any $y = \sum_{i=1}^{m} \alpha_i x_i$.

(b) Simple degenerate case $q_i = 1$: Here we can find the first eigenvector in the usual means, then we have $m_i - 1$ *generalized eigenvectors*. These are obtained as

$$\begin{aligned} Ax_1 &= \lambda_i x_1 \\ (A - \lambda_i)x_2 &= x_1 \\ \vdots &= \vdots \\ (A - \lambda_i)x_{m_i} &= x_{m_i - 1} \end{aligned} \tag{13}$$

Note that all x_i found this way are linearly independent.

Example 15

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 8 & -20 & -18 & -7 \end{bmatrix}$$

its characteristic polynomial is

$$\Delta(\lambda) = \lambda^4 + 7\lambda^3 + 18\lambda^2 + 20\lambda + 8 = 0$$

then, $\lambda_1 = -1$, $\lambda_2 = \lambda_3 = \lambda_4 = -2$. The eigenvector of λ_1 is easily found to be $x_1 = [-1\ 1\ -1\ 1]^T$. On the other hand, one eigenvector 0f -2 is found to be $x_2 = [0.125\ -0.25\ 0.5 - 1]^T$, then $x_2 = (A + 2I)x_3$ leading to $x_3 = [0.1875\ -0.25\ 0.25\ 0]^T$ also, $x_3 = (A + 2I)x_4$ so that $x_4 = [0.1875\ -0.1875\ 0.125\ 0]^T$

△

(c) General case $1 \leq q_i \leq m_i$: Here, a general top-down method should be used. Here we solve the problem by writing

$$\begin{aligned}
(A-\lambda_i I)x_1 &= 0 \\
(A-\lambda_i I)x_2 &= x_1 \Rightarrow (A-\lambda_i I)^2 x_2 = (A-\lambda_i I)x_1 = 0 \\
(A-\lambda_i I)x_3 &= x_2 \Rightarrow (A-\lambda_i I)^2 x_3 = (A-\lambda_i I)x_2 = x_1 \neq 0 \\
& \quad (A-\lambda_i I)^3 x_3 = (A-\lambda_i I)x_1 = 0
\end{aligned} \tag{14}$$

This approach continues until we reach the index k_i of λ_i. This index is found as the smallest integer such that

$$rank(A-\lambda_i I)^k = n - m_i \tag{15}$$

the index indicates the length of the longest chain of eigenvectors and generalized eigenvectors associated with λ_i.

Example 16

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

its characteristic polynomial is

$$\Delta(\lambda) = \lambda^4 = 0$$

so $\lambda_1 = 0$ with $m_1 = 4$. Next, form $A - \lambda_1 I = A$, and determine that $rank(A-\lambda_1 I) = 2$. There are then 2 eigenvectors and 2 generalized eigenvectors. The question is whether we have one eigenvector-generalized eigenvector chain of length 3, or 2 chains of length 2. To check that, note that $n-m=0$, and that $rank(A-\lambda_1 I)^2 = 0$, therefore, the index is $k_1 = 2$. this then guarantees that one chain has length 2, making the length of the other chain 2. First, consider $(A-\lambda_1 I)^2 x = 0$ Any vector satisfies this equation but only 4 vectors are linearly independent. Let us choose $x_1 = [1\ 0\ 0\ 0]^T$. Is this an eigenvector or a generalized eigenvector? It is an eigenvector since $(A-\lambda_1 I)x_1 = 0$. Similarly, $x_2 = [0\ 1\ 0\ 0]^T$ is an eigenvector. On the other hand, $x_3 = [0\ 0\ 1\ 0]^T$ is a generalized eigenvector since $(A-\lambda_i I)x_3 = x_1 \neq 0$. Similarly, $x_4 = [0\ 0\ 0\ 1]^T$ is a generalized eigenvector associated with x_2.

$\triangle$

Example 17 *Let*

$$A = \begin{pmatrix} 4 & -1 & 2 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

The eigenvalues of A are $\lambda_1 = 4$, $\lambda_{2,3} = 2$ and the two correspondent eigenvectors are $x_1 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$, $x_2 = \begin{pmatrix} -1 & 0 & 1 \end{pmatrix}^T$. Therefore we have three eigenvalues but only two linearly independent eigenvectors. We call generalized eigenvector of A the vector $x_3 = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}^T$ such that

$$Ax_3 = \lambda_2 x_3 + x_2.$$

The vector x_3 is special in the sense that, x_2 and x_3 together span a two-dimensional $A-$invariant subspace. $\triangle$

In summary each $n \times n$ matrix has n eigenvalues and n linearly independent vectors, either eigenvectors or generalized eigenvectors.

To summarize, given the characteristic polynomial

$$p(\lambda) := \det(\lambda I - A) = \lambda^n + a_1 \lambda^{n-1} + \ldots + a_n, \tag{16}$$

the roots $\lambda_1, \ldots \lambda_n$ of $p(\lambda) = 0$ are the *eigenvalues* of A. The set $\Lambda = \{\lambda_1, \ldots \lambda_n\}$ is called the *spectrum* of A. The *spectral radius* of A is defined as

$$\rho(A) := \max_{i=1,\ldots n} |\lambda_i|.$$

where $|.|$ is the modulus of the argument and thus $\rho(A)$ is the radius of the smallest circle in the complex plane, centered at the origin, that includes all the eigenvalues. As described earlier, the non-null vectors x_i such that $Ax_i = \lambda_i x_i$ are the *(right-)eigenvectors* of A. Similarly $y \neq 0$ is a *left-eigenvector* of A if $y^* A = \lambda y^*$. In general a matrix A has at least one eigenvector. It is easy to see that if x is an eigenvector of A, Span$\{x\}$ is an $A-$*invariant* subspace.

In general let us suppose that a matrix A has an eigenvalue λ of multiplicity r but with only one correspondent eigenvector. Then we can define $r-1$ generalized eigenvectors in the following way

$$\begin{aligned} Ax_1 &= \lambda x_1 \\ Ax_2 &= \lambda x_2 + x_1 \\ &\vdots \\ Ax_r &= \lambda x_r + x_{r-1}. \end{aligned}$$

The eigenvector and the $r-1$ generalized eigenvectors together span a r-dimensional $A - invariant$ subspace.

3.1 Jordan Forms

Based on the analysis above, we can produce the Jordan form of any $n \times n$ matrix. In fact, if we have n different eigenvalues, we can find all eigenvectors x_i, such that $Ax_i = \lambda_i x_i$, then let $M = [x_1\ x_2\ \cdots\ x_n]$, and $\Lambda = diag(\lambda_i)$. This leads to $AM = M\Lambda$. We know that M^{-1} exists since all eigenvectors are independent and therefore $\Lambda = M^{-1}AM$. The Jordan form of A is then Λ. If the eigenvectors are orthonormal, $M^{-1} = M^T$. In the case that $q_i = m_i$, we proceed the same way to obtain $J = \Lambda$. On the other hand, if $q_i = 1$, then using the eigenvectors and generalized eigenvectors as columns of M will lead to $J = diag[J_1\ J_2\ \cdots\ J_p]$ where each J_i is $m_i \times m_i$ and has the following form

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix}$$

Finally, the general case will have Jordan blocks each of size k_i as shown in the examples below.

Example 18 Let

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -18 & -27 & -10 \end{bmatrix}$$

the eigenvalues are $\lambda_1 = -1,\ \lambda_2 = -3,\ \lambda_3 = -6$. The eigenvectors are

$$x_1 = [1\ -1\ 1]^T$$

Similarly, $x_2 = [1\ -3\ 9]^T$ and $x_3 = [1\ -6\ 36]^T$. Then,

$$M = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -3 & -6 \\ 1 & 9 & 36 \end{bmatrix}$$

then

$$M^{-1} = \begin{bmatrix} 1.8 & 0.9 & 0.1 \\ -1 & -1.167 & -0.167 \\ 0.2 & 0.2667 & 0.067 \end{bmatrix}$$

then $\Lambda = M^{-1}AM$.

△

Example 19 Given the matrix

$$A = \begin{bmatrix} 10/3 & 1 & -1 & -1/3 \\ 0 & 4 & 0 & 0 \\ -2/3 & 1 & 3 & -1/3 \\ -2/3 & 1 & -1 & 11/3 \end{bmatrix}$$

Then, its characteristic equation is

$$\Delta(\lambda) = \lambda^4 - 14\lambda^3 + 72\lambda^2 - 160\lambda + 128 = 0$$

There are 4 roots, $\lambda_1 = 2$, $\lambda_2 = \lambda_3 = \lambda_4 = 4$. We can find the eigenvector associated with λ_1 as $x_1 = [-1\ 0\ -1\ -1]^T$. Then, we find $x_2 = [1\ 0\ 0\ -2]^T$; $x_3[0\ 1\ 1\ 0]^T$; $x_2 = [0\ 1\ 0\ 3]^T$ Therefore,

$$M = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ -1 & 0 & 1 & 0 \\ -1 & -2 & 0 & 3 \end{bmatrix}$$

then

$$M^{-1} = \begin{bmatrix} -1/3 & 1/2 & -1/2 & -1/6 \\ 1/6 & 1/2 & -1/2 & -1/6 \\ -1/3 & 1/2 & 1/2 & -1/6 \\ 1/3 & 1/2 & -1/2 & 1/6 \end{bmatrix}$$

then

$$J = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

$\triangle$

Example 20 Now consider

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 8 & -20 & -18 & -7 \end{bmatrix}$$

its characteristic polynomial is

$$\Delta(\lambda) = \lambda^4 + 7\lambda^3 + 18\lambda^2 + 20\lambda + 8 = 0$$

then, $\lambda_1 = -1$, $\lambda_2 = \lambda_3 = \lambda_4 = -2$. Find eigenvectors as before and form

$$M = \begin{bmatrix} -1 & 1/8 & 0.1875 & 0.1875 \\ 1 & -1/4 & -1/4 & -0.1875 \\ -1 & 1/2 & 1/4 & 1/8 \\ 1 & -1 & 0 & 0 \end{bmatrix}$$

Then,

$$J = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & -2 \end{bmatrix}$$

$\triangle$

Example 21

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

its characteristic polynomial is

$$\Delta(\lambda) = \lambda^4 = 0$$

so $\lambda_1 = 0$ with $m_1 = 4$. Then,

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

therefore making

$$J = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$\triangle$

4 Inner Products & Norms

4.1 Inner Products

An inner product is an operation between two vectors of a vector space which will allow us to define geometric concepts such as orthogonality and Fourier series, etc. The following defines an inner product.

Definition 3 An inner product defined over a vector space V is a function $< .,. >$ defined from V to F where F is either $\mathbb{R}$ or $\mathbb{C}$ such that $\forall x,y,z, \in V$

1. $< x,y >=< y,x >^*$ where the $< .,. >^*$ denotes the complex conjugate.
2. $< x,y+z >=< x,y > + < x,z >$
3. $< x,\alpha y >= \alpha < x,y >, \ \forall \alpha \in F$
4. $< x,x >\geq 0$ where the 0 occurs only for $x = 0_V$

Example 22 The usual dot product in $\mathbb{R}^n$ is an inner product, i.e. $< x,y >= x^T y = \sum_{i=1}^n x_i y_i$.

$\triangle$

4.2 Norms

A *norm* is a generalization of the ideas of distance and length. As stability theory is usually concerned with the size of some vectors and matrices, we give here a brief description of some norms that will be used in these notes. We will consider first the norms of vectors defined on a vector space X with the associated scalar field of real numbers $\mathbb{R}$.

Let $\mathscr{X}$ be a linear space on a field $\mathbb{F}$. A function $\|\cdot\| : \mathscr{X} \mapsto \mathbb{R}$ is called a *norm* if it satisfies the following proporties

1. $\|x\| \geq 0, \quad \forall x \in \mathscr{X}$
2. $\|x\| = 0 \Longleftrightarrow x = 0$
3. $\|ax\| = |a|\|x\|, \quad \forall a \in \mathbb{F},\ \forall x \in \mathscr{X}$
4. $\|x+y\| \leq \|x\| + \|y\|, \quad \forall x, y \in \mathscr{X}$

Let now $\mathscr{X} = \mathbb{C}^n$. For a vector $x = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}^T$ in $\mathbb{C}^n$ the p-norm is defined as

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}},\ p \geq 1. \tag{17}$$

When $p = 1, 2, \infty$, we have the three important norms

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|$$
$$\|x\|_2 := \sqrt{\sum_{i=1}^{n} |x_i|^2}$$
$$\|x\|_\infty := \max_i |x_i|.$$

In a Euclidean space $\mathbb{R}^n$, the 2-norm is the usual Euclidean norm.

Example 23 Consider the vector

$$x = \begin{bmatrix} 1 \\ -2 \\ 2 \end{bmatrix}$$

Then, $\| x \|_1 = 5$, $\| x \|_2 = 3$ and $\| x \|_\infty = 2$.

$\triangle$

Let us now consider the norms for a matrix $A \in \mathbb{C}^{m \times n}$. First let us define the *induced* p-norms

$$\|A\|_p := \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p},\ p \geq 1.$$

These norms are induced by the p-norms on vectors. For $p = 1,2,\infty$, there exists some explicit formulae for the induced $p-$norms

$$\|A\|_1 := \max_{1\le j\le n} \sum_{i=1}^{m} |a_{ij}| \quad \text{(maximum absolute column sum)}$$
$$\|A\|_2 := \sqrt{\lambda_{max}(A^*A)} \quad \text{(spectral norm)}$$
$$\|A\|_\infty := \max_{1\le i\le m} \sum_{j=1}^{n} |a_{ij}| \quad \text{(maximum absolute row sum).}$$

Unless otherwise specified, we shall adopt the convention of denoting the 2-norm without any subscript. Therefore by $\|x\|$ and $\|A\|$ we shall mean respectively $\|x\|_2$ and $\|A\|_2$.

Another often used matrix norm is the so-called *Frobenius* norm

$$\|A\|_F := \sqrt{\operatorname{tr}(A^*A)} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}. \tag{18}$$

It is possible to show that the Frobenius norm cannot be induced by any vector norm.

The following Lemma presents some useful results regarding matrix norms.

Lemma 1 *Let $A \in \mathbb{C}^{m\times n}$. Then*

1. $\|AB\| \le \|A\|\|B\|$, *for any induced norm (submultiplicative property)*
2. *Given two unitary matrices U and V of suitable dimensions*
$$\|UAV\| = \|A\| \qquad \|UAV\|_F = \|A\|_F$$
3. $\rho(A) \le \|A\|$, *for any induced norm and the Frobenius norm*

Given a square nonsingular matrix A, the quantity

$$\kappa(A) := \|A\|_p \|A^{-1}\|_p \tag{19}$$

is called the *condition number* of A with respect to the induced matrix norm $\|\cdot\|_p$. From Lemma 1, we have

$$\kappa(A) = \|A\|_p \|A^{-1}\|_p \ge \|AA^{-1}\|_p = \|I\|_p = 1.$$

If $\kappa(A)$ is large, we say that A is *ill conditioned*; if $\kappa(A)$ is small (i.e. close to 1), we say that A is *well conditioned*. It is possible to prove that, given a matrix A, the reciprocal of $\kappa(A)$ gives a measure of how far A is from a singular matrix.

We now present an important property of norms of vectors in $\mathbb{R}^n$ which will be useful in the sequel.

Lemma 1 Let $\| x \|_a$ and $\| x \|_b$ be any two norms of a vector $x \in \mathbb{R}^n$. Then there exists finite positive constants k_1 and k_2 such that

$$k_1 \| x \|_a \leq \| x \|_b \leq k_2 \| x \|_a \ \ \forall x \in R^n$$

The two norms in the lemma are said to be equivalent and this particular property will hold for any two norms on $\mathbb{R}^n$.

Example 24 *Note the following*

1. It can be shown that for $x \in \mathbb{R}^n$

$$\begin{aligned} ||x||_1 &\leq \sqrt{n}||x||_2 \\ ||x||_\infty \leq ||x||_1 &\leq n||x||_\infty \\ ||x||_2 &\leq \sqrt{n}||x||_\infty \end{aligned}$$

2. Consider again the vector of ex 23. Then we can check that

$$\begin{aligned} ||x||_1 &\leq \sqrt{3}||x||_2 \\ ||x||_\infty \leq ||x||_1 &\leq 3||x||_\infty \\ ||x||_2 &\leq \sqrt{3}||x||_\infty \end{aligned}$$

△

Note that a norm may be defined independently from an inner product. Also, we can define the generalized angle between 2 vectors in $\mathbb{R}^n$ using

$$x.y = < x,y > = \| x \| . \| y \| \cos\theta \tag{20}$$

where θ is the angle between x and y. Using the inner product, we can define the orthogonality of 2 vectors by

$$x.y = < x,y > = 0 \tag{21}$$

which of course means that $\theta = (2i+1)\pi/2$.

5 Applications of Linear Algebra

5.1 The Cayley-Hamilton Theorem

Note that $A^i = A \cdots A$ i times. Then, we have the Cayley-Hamilton theorem

Theorem 1 Let A be an $n \times n$ matrix with a characteristic equation

$$\begin{aligned} |(A - \lambda I) &= \Delta(\lambda) \\ &= (-\lambda)^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 \\ &= 0 \end{aligned} \tag{22}$$

Then,

$$\begin{aligned} \Delta(A) &= (-1)^n A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I \\ &= 0 \end{aligned} \tag{23}$$

■

Example 25 Use the Cayley-Hamilton theorem to find the inverse of a matrix A

$$0 = (-1)^n A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I$$

then,

$$0 = (-1)^n A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1 I + c_0 A^{-1}$$

$$A^{-1} = \frac{-1}{c_0}[(-1)^n A^{n-1} + c_{n-1}A^{n-2} + \cdots + c_1 A]$$

△

5.2 Solving State Equations

Suppose we want to find a solution to the state-space equation

$$\dot{x} = Ax(t) + bu(t);\ x(0) = x_0$$

and assume initially that we are interested in finding $x(t)$ for $u(t) = 0; \forall t \geq 0$. Effectively, we then have

$$\dot{x} = Ax(t)$$

Let us then take the Laplace transform,

$$sX(s) - x_0 = AX(s)$$

which gives

$$X(s) = [sI - A]^{-1}x_0$$

which then leads to

$$x(t) = \mathscr{L}^{-1}[sI - A]^{-1}x_0$$

we define the matrix exponential

$$exp(At) = \mathscr{L}^{-1}[sI - A]^{-1}$$

so that

$$x(t) = exp(At)x_0$$

there are many reasons why that is justified. In fact, it can be shown that the matrix exponential has the following properties

1. $\frac{d}{dt}exp(At) = Aexp(At) = exp(At)A$
2. $expA(t_1 + t_2) = exp(At_1)exp(At_2)$
3. $exp(A0) = I$
4. $exp(At)$ is nonsingular and $[exp(At)]^{-1} = exp(-At)$
5. $exp(At) = \lim_{k\to\infty} \sum_{i=0}^{k} \frac{A^i t^i}{i}$

The matrix $exp(At)$ is known as the state transition matrix.

Now assume that $u(t)$ is no longer zero. Then, the solution due to both $x(0)$ and $u(t)$ is given by

$$x(t) = exp(At)x_0 + \int_0^t e^{A(t-\tau)}bu(\tau)d\tau$$

Example 26 Consider the system

$$\dot{x} = \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix} x$$
$$x(0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

To find $x(t)$ let us find $exp(At)$ using 2 methods. First, by evaluating the infinite series

$$
\begin{aligned}
exp(At) &= I + At + A^2\frac{t^2}{2} + \cdots \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2 & -1 \\ 4 & -2 \end{bmatrix} t + 0 \\
&= \begin{bmatrix} 1+2t & -t \\ 4t & 1-2t \end{bmatrix} \qquad (24)
\end{aligned}
$$

so that

$$
\begin{aligned}
x(t) &= \begin{bmatrix} 1+2t & -t \\ 4t & 1-2t \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\
&= \begin{bmatrix} 1+3t \\ 6t-1 \end{bmatrix}
\end{aligned}
$$

next, consider the Laplace transform approach, where we find $(sI-A)^{-1}$

$$(sI-A)^{-1} = \begin{bmatrix} \frac{s+2}{s^2} & \frac{-1}{s^2} \\ \frac{4}{s^2} & \frac{s-2}{s^2} \end{bmatrix} \qquad (25)$$

so that

$$exp(At) = \begin{bmatrix} 1+2t & -t \\ 4t & 1-2t \end{bmatrix} \qquad (26)$$

Now consider a third method whereby, we transform A to its Jordan normal form. The A matrix has a double eigenvalue at zero with an eigenvector $x_1 = [1 \;\; 2]^T$ and a generalized eigenvector at $x_2 = [0 \;\; -1]$. Therefore, the M matrix is given by

$$\begin{bmatrix} 1 & 0 \\ 2 & -11 \end{bmatrix} \qquad (27)$$

Using $T = M$ we obtain,

$$J = \bar{A} = T^{-1}AT = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad (28)$$

so that

$$\dot{\bar{x}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \bar{x}; \; \bar{x}(0) = [1 \;\; 3]^T$$

we can then solve 2 differential equations

$$\begin{aligned} \dot{\bar{x}}_1 &= x_2 \\ \dot{\bar{x}}_2 &= 0 \end{aligned}$$

so that $\bar{x}_2(t) = 3$ and $\bar{x}_1 = 3t + 1$. Therefore,

$$\begin{aligned} x(t) &= T^{-1}\bar{x} \\ &= \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 3t+1 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 3t+1 \\ 6t-1 \end{bmatrix} \end{aligned} \tag{29}$$

△

Note that in general

$$exp(At) = Mexp(Jt)M^{-1}$$

since

$$\begin{aligned} Mexp(Jt)M^{-1} &= Mexp(M^{-1}AM)M^{-1} \\ &= M[I + M^{-1}AMt + \frac{1}{2}(M^{-1}AMt)^2 + \cdots]M^{-1} \\ &= I + At + \frac{1}{2}A^2t^2 + \cdots \\ &= exp(At) \end{aligned}$$

It is extremely simple to calculate $exp(Jt) = exp(\Lambda t)$ for the case of diagonalizable matrix A as shown in the following example.

Example 27 Consider the system

$$\dot{x} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} x$$

$$x(0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

then there are 2 distinct eigenvalues, and we can find $exp(Jt)$,

$$\lambda_1 = -0.2361;\ \lambda_2 = 4.2361$$

$$M = \begin{bmatrix} 0.8507 & 0.5257 \\ -0.5257 & 0.8507 \end{bmatrix}$$

so that

$$M^{-1} = \begin{bmatrix} 0.8507 & -0.5257 \\ 0.5257 & 0.8507 \end{bmatrix}$$

Note that $M^{-1} = M^T$. Then,

$$exp(At) = \begin{bmatrix} 0.7237e^{-0.2361t} + 0.2764e^{4.2361t} & 0.4472(e^{4.2361t} - e^{-0.2361t}) \\ 0.4472(e^{4.2361t} - e^{-0.2361t}) & 0.2764e^{-0.2361t} + 0.7237e^{4.2361t} \end{bmatrix}$$

$\triangle$

Recall that A matrix $A \in \mathbb{C}^{n\times n}$ is said to be *Hermitian* if $A = A^*$, *unitary* if $A^{-1} = A^*$. A Hermitian matrix has the property that all its eigenvalues are real. Let $A = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}$. It is easy to verify that A is unitary if and only if $a_i^* a_j = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$, i.e. if and only if the columns of A form an orthonormal basis for $\mathbb{C}^n$.

Let us now consider a nonsingular A partitioned as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{30}$$

with A_{11} and A_{22} square matrices. Suppose that A_{11} is nonsingular. Then the matrix

$$\Lambda := A_{22} - A_{21}A_{11}^{-1}A_{12}$$

is called the *Schur complement* of A_{11} in A. Similarly, if A_{22} is nonsingular, the matrix

$$\hat{\Lambda} := A_{11} - A_{12}A_{22}^{-1}A_{21}$$

is the *Schur complement* of A_{22} in A. A useful expression for the inverse of A in terms of partitioned blocks is

$$A^{-1} = \begin{pmatrix} \hat{\Lambda}^{-1} & -A_{11}^{-1}A_{12}\Lambda^{-1} \\ -\Lambda^{-1}A_{21}A_{11}^{-1} & \Lambda^{-1} \end{pmatrix} \tag{31}$$

supposing that all the relevant inverses exist. The following well-established identities are also very useful

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} \tag{32}$$

$$(I + AB)^{-1}A = A(I + BA)^{-1} \tag{33}$$

$$I - AB(I + AB)^{-1} = (I + AB)^{-1} \tag{34}$$

supposing the existence of all the needed inverses.

If A_{11} is nonsingular then by multiplying A on the right by

$$\begin{pmatrix} I & -A_{11}^{-1}A_{12} \\ 0 & I \end{pmatrix}$$

it is easy to verify that

$$\det(A) = \det(A_{11})\det(\Lambda). \tag{35}$$

Similarly, if A_{22} is nonsingular, then by multiplying A on the right by

$$\begin{pmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{pmatrix}$$

we find that

$$\det(A) = \det(A_{22})\det(\hat{\Lambda}). \tag{36}$$

Consider $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$. Using identities (35) and (36), it is easy to prove that that

$$\det(I_m \pm AB) = \det(I_n \pm BA).$$

5.3 Matrix Decompositions

In this section we shall explore some matrix decompositions that use Unitary matrices.

Theorem 1 (Schur Decomposition) *Let $A \in \mathbb{C}^{n\times n}$. Then there exist a Unitary matrix $U \in \mathbb{C}^{n\times n}$ so that $A = UTU^*$, where T is a triangular matrix whose diagonal entries are the eigenvalues of A.*

Proof 1 *We shall prove the theorem by constructing the matrices U and T. Given a matrix A, we can always find an eigenvector u_1, associated with an eigenvalue λ_1, so that $Au_1 = \lambda_1 u_1$. Moreover we can assume that $\|u_1\| = 1$. Let $U_1 = \begin{pmatrix} u_1 & z_2 & \cdots & z_n \end{pmatrix}$ be a Unitary matrix. Since the columns of U_1 are a basis for $\mathbb{C}^n$, we have*

$$A\begin{pmatrix} u_1 & z_2 & \cdots & z_n \end{pmatrix} = \begin{pmatrix} u_1 & z_2 & \cdots & z_n \end{pmatrix}\begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix}.$$

Therefore

$$A = U_1\begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix}U_1^*. \tag{37}$$

Now let us consider $A_1 \in \mathbb{C}^{(n-1)\times(n-1)}$*. From* (37) *it is easy to see that the* $(n-1)$ *eigenvalues of* A_1 *are also eigenvalues of* A*. Given* λ_2 *an eigenvalue of* A_1 *with normalized eigenvector* u_2*, we can construct a Unitary matrix* $U_2 = \begin{pmatrix} u_2 & z_2' & \cdots & z_{n-1}' \end{pmatrix} \in \mathbb{C}^{(n-1)\times(n-1)}$ *so that*

$$A_1 U_2 = U_2 \begin{pmatrix} \lambda_2 & * \\ 0 & A_2 \end{pmatrix}.$$

Denoting by V_2 *the Unitary matrix*

$$V_2 = \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix},$$

we have

$$V_2^* U_1^* A U_1 V_2 = \begin{pmatrix} 1 & 0 \\ 0 & U_2^* \end{pmatrix} \begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & * & * \\ 0 & \lambda_2 & * \\ 0 & 0 & A_2 \end{pmatrix}.$$

The result can be proven iterating this procedure n *times.*

Using Theorem 1 it is possible to prove that a Hermitian matrix $A \in \mathbb{C}^{n\times n}$ has a set of n linearly independent eigenvectors. Indeed, since A is Hermitian, from Theorem 1 we have

$$A = UTU^* = UT^*U^* \Rightarrow T = T^*.$$

Since T is triangular, it must be diagonal. Therefore the columns of the matrix U are the eigenvectors of A.

In Theorem 1, we used the same basis both for the domain and the range of the matrix. In the following decomposition, we shall show what can be accomplished when we choose possibly different bases for domain and the range.

Theorem 2 (Singular Value Decomposition) *Let* $A \in \mathbb{C}^{m\times n}$*. Then* A *can always be written as* $A = U\Sigma V^*$*, where* $U \in \mathbb{C}^{m\times m}$ *and* $V \in \mathbb{C}^{n\times n}$ *are Unitary matrices and* $\Sigma = [\sigma_{ij}] \in \mathbb{R}^{m\times n}$ *is a diagonal matrix with* $\sigma_{ij} = 0$ *for* $i \neq j$ *and* $\sigma_{ii} = \sigma_i \geq 0$.

Proof 2 *We shall prove the theorem by induction. First of all, let us suppose, without any loss of generality that* $A \neq 0$ *and* $m > n$*. Let now* $m = p+1$ *and* $n = 1$*. In this case we can write*

$$A = \begin{pmatrix} \frac{A}{\sigma} & u_2 & \cdots & u_n \end{pmatrix} \sigma \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

with $\sigma = \|A\|$. *Therefore* $A = U\Sigma V^*$, *with* $U = \begin{pmatrix} \frac{A}{\sigma} & u_2 & \cdots & u_n \end{pmatrix}$, $\Sigma = \sigma \begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}^T$ *and* $V = 1$.

We suppose now that the theorem holds for $m = p+k$ *and* $n = k$ *and prove it for* $m = p+k+1$ *and* $n = k+1$. *Let* $A \in \mathbb{C}^{(p+k+1)\times(k+1)}$. *All the eigenvalues of* A^*A *are real (since* A^*A *is Hermitian), nonnegative and at least one is greater than zero since* $A \neq 0$. *Denote by* λ *the maximum eigenvalue of* A^*A *and by* v *the correspondent normalized eigenvector*

$$A^*Av = \lambda v.$$

Let now $\sigma = +\sqrt{\lambda}$ *and* $u = \frac{Av}{\sigma}$, *so that* $\|u\| = 1$. *Let moreover* $U_1 = \begin{pmatrix} u & U_0 \end{pmatrix} \in \mathbb{C}^{(p+k+1)\times(p+k+1)}$ *and* $V_1 = \begin{pmatrix} v & V_0 \end{pmatrix} \in \mathbb{C}^{(k+1)\times(k+1)}$ *be Unitary matrices. We have*

$$A_1 = U_1^* A V_1 = \begin{pmatrix} u^* \\ U_0^* \end{pmatrix} A \begin{pmatrix} v & V_0 \end{pmatrix} = \begin{pmatrix} \sigma & u^*AV_0 \\ 0 & U_0^*AV_0 \end{pmatrix}.$$

Since $u^*A = \frac{v^*A^*}{\sigma} A = \sigma v^*$, *it follows that* $u^*AV_0 = 0$. *Now an easy inductive argument completes the proof, noting that* $U_0^*AV_0 \in \mathbb{C}^{(p+k)\times k}$.

The scalars σ_i are called the *singular values* of A. They are usually ordered nonincreasingly $\sigma_1 \geq \sigma_2 \ldots \geq \sigma_r \geq 0$, with $r = \min\{m, n\}$. The largest singular value σ_1 and the smallest singular value σ_r are denoted by

$$\overline{\sigma}(A) := \sigma_1 \qquad \underline{\sigma}(A) := \sigma_r.$$

The columns v_i of V are called the *right singular vectors* of A, and the columns u_i of U the *left singular vectors*. They are related by

$$Av_i = \sigma_i u_i, \qquad i = 1, \ldots r.$$

The following Lemma shows some of the information we can get from the singular value decomposition of a matrix A.

Lemma 2 *Let* $A \in \mathbb{C}^{m\times n}$ *and consider its singular value decomposition* $A = U\Sigma V^*$. *Then*

1. *The rank* k *of* A *equals the number of singular values different from zero*
2. $\mathrm{Range}(A) = \mathrm{Span}\{u_1, \ldots u_k\}$
3. $\mathrm{Ker}(A) = \mathrm{Span}\{v_{k+1}, \ldots v_n\}$
4. $\sigma_1 = \|A\|$

5. $\|A\|_F^2 = \sum_{i=1}^r \sigma_i^2$

6. *Given a square nonsingular matrix A*

$$\overline{\sigma}(A) = \frac{1}{\underline{\sigma}(A^{-1})}$$

5.4 Bilinear Forms and Sign-definite Matrices

The expression $< y, Ax >= y^T Ax$ is a bilinear form. When $y = x$ we have a quadratic form $x^T Ax$. Every matrix A can be written as the sum of a symmetric and skew symmetric matrices if it is real, and of a Hermitian and skew-Hermitian if it is complex. In fact, define the symmetric and anti-symmetric parts of A as:

$$A_s = \frac{A + A^T}{2}; A_a = \frac{A - A^T}{2}$$

and the Hermitian and anti-Hermitian parts as:

$$A_H = \frac{A + (A^*)^T}{2}; A_{AH} = \frac{A - (A^*)^T}{2}$$

Then note that $< x, Ax >=< x, A_s x >$ if A is real, and $< x, Ax >=< x, A_H x >$ if A is complex.

5.4.1 Definite Matrices

Let us consider a Hermitian matrix A. A is said to be *positive (semi)definite* if $x^* Ax > (\geq)\ 0, \forall x \in \mathbb{C}^n$. We shall indicate a positive (semi)definite matrix by $A > (\geq)\ 0$. A Hermitian matrix A is said to *negative (semi)definite* if $(-A)$ is positive (semi)definite. The following Lemma gives a characterization of definite matrices.

Lemma 3 *Let A be a Hermitian matrix. Then*

1. *A is positive (negative) definite if and only if all its eigenvalues are positive (negative).*

2. *A is positive (negative) semidefinite if and only if all its eigenvalues are non-negative (nonpositive).*

Given a real $n \times n$ matrix Q, then

1. Q is positive-definite, if and only if $x^T Qx > 0$ for all $x \neq 0$.

2. Q is positive semidefinite if $x^T Qx \geq 0$ for all x.

3. Q is indefinite if $x^T Qx >$ for some x and $x^T Qx < 0$ for other x.

Q is negative definite (or semidefinite) if $-Q$ is positive definite (or semidefinite). If Q is symmetric then all its eigenvalues are real.

Example 28 Show that if A is symmetric, then A is positive-definite if and only if all of its eigenvalues (which we know are real) are positive. This then allows us to test for the sign-definiteness of a matrix A by looking at the eigenvalues of its symmetric part.

△

5.5 Matrix Inversion Lemmas

There are various important relations involving inverses of matrices, one of which is the *Matrix Inversion Lemma*

$$(A_1 - A_2A_4^{-1}A_3)^{-1} = A_1^{-1} + A_1^{-1}A_2(A_4 - A_3A_1^{-1}A_2)^{-1}A_3A_1^{-1} \tag{38}$$

Example 29 Prove the Matrix Inversion lemma
Solution: Let the LHS be A^{-1} and the RHS be B. The proof proceeds by showing that

$$AB = BA = I$$

so,

$$\begin{aligned}
AB &= (A_1 - A_2A_4^{-1}A_3)(A_1^{-1} + A^{-1}A_2(A_4 - A_3A_1^{-1}A_2)^{-1}A_3A_1^{-1}) \\
&= I - A_2A_4^{-1}A_3A_1^{-1} + A_2(A_4 - A_3A_1^{-1}A_2)^{-1}A_3A_1^{-1} \\
&\quad - A_2A_4^{-1}A_3A_1^{-1}A_2(A_4 - A_3A_1^{-1}A_2)^{-1}A_3A_1^{-1} \\
&= I - A_2A_4^{-1}[I - A_4(A_4 - A_3A_1^{-1}A_2)^{-1} + A_3A_1^{-1}A_2(A_4 - A_3A_1^{-1}A_2)^{-1}]A_3A_1^{-1} \\
&= I - A_2A_4^{-1}[I - (A_4 - A_3A_1^{-1}A_2)(A_4 - A_3A_1^{-1}A_2)^{-1}]A_3A_1^{-1} \\
&= I
\end{aligned}$$

On the other hand,

$$\begin{aligned}
BA &= (A_1^{-1} + A^{-1}A_2(A_4 - A_3A_1^{-1}A_2)^{-1}A_3A_1^{-1})(A_1 - A_2A_4^{-1}A_3) \\
&= I - A_1^{-1}A_2A_4^{-1}A_3 + A_1^{-1}A_2(A_4 - A_3A_1^{-1}A_2)^{-1}A_3 \\
&\quad - A_1^{-1}A_2(A_4 - A_3A_1^{-1}A_2)^{-1}A_3A_1^{-1}A_2A_4^{-1}A_3 \\
&= I - A_1^{-1}A_2[I - (A_4 - A_3A_1^{-1}A_2)^{-1}A_4 + (A_4 - A_3A_1^{-1}A_2)^{-1}A_3A_1^{-1}A_2)]A_4^{-1}A_3 \\
&= I - A_1^{-1}A_2[I - (A_4 - A_3A_1^{-1}A_2)^{-1}(A_4 - A_3A_1^{-1}A_2)]A_4^{-1}A_3 \\
&= I
\end{aligned}$$

△

Now let us consider partition (30) for a Hermitian matrix A

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{pmatrix} \tag{39}$$

with A_{11} and A_{22} square matrices. The following Theorem gives a characterization of a positive matrix in terms of its partition (39).

Theorem 3 *Let A be a Hermitian matrix partitioned as in* (39). *Then*

1. *A is positive definite if and only if A_{11} and $A_{22} - A_{12}^* A_{11}^{-1} A_{12}$ are positive definite*
2. *A is positive definite if and only if A_{22} and $A_{11} - A_{12} A_{22}^{-1} A_{12}^*$ are positive definite*

Proof 3

1. *The proof follows from the facts that*

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{pmatrix} = \begin{pmatrix} I & A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix}^* \begin{pmatrix} A_{11} & 0 \\ 0 & A_{22} - A_{12}^* A_{11}^{-1} A_{12} \end{pmatrix} \begin{pmatrix} I & A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix}$$

and that

$$\begin{pmatrix} I & A_{11}^{-1} A_{12} \\ 0 & I \end{pmatrix}$$

is a full rank matrix.

2. *As in* (1), *writing A as*

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{12}^* & A_{22} \end{pmatrix} = \begin{pmatrix} I & A_{12} A_{22}^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A_{11} - A_{12} A_{22}^{-1} A_{12}^* & 0 \\ 0 & A_{22} \end{pmatrix} \begin{pmatrix} I & A_{12} A_{22}^{-1} \\ 0 & I \end{pmatrix}^* .$$

6 MATLAB Commands

There are plenty of MATLAB commands dealing with matrices. In this section we shall give a brief overview of those related with the topics covered in this chapter. We recall that the (conjugate) transpose of a matrix A is evaluated simply typing `A'` and that a polynomial is represented by the row vector of its coefficients ordered in *descending* powers.

$\gg$ `Q=orth(A)` returns a matrix `Q` whose columns are an orthonormal basis for the range of a (rectangular) matrix `A`.

$\gg$ `Q=null(A)` returns a matrix `Q` whose columns are an orthonormal basis for the null space of a (rectangular) matrix `A`.

$\gg$ `det(A)` returns the determinant of a square matrix `A`.

$\gg$ `inv(A)` returns the inverse (if it exists) of a square matrix `A`. A warning is given if the matrix is ill conditioned.

$\gg$ `pinv(A)` returns the pseudo-inverse of a matrix `A`.

$\gg$ `trace(A)` returns the trace of a square matrix `A`.

$\gg$ `rank(A)` returns the rank of a matrix `A`.

$\gg$ `cond(A)` evaluates the condition number (19) of a square matrix `A`, using the matrix spectral norm. In this case (19) becomes

$$\kappa(A) = \frac{\overline{\sigma}(A)}{\underline{\sigma}(A)}.$$

The command `condest(A)` can be used to get an estimate of the 1-norm condition number.

$\gg$ `[V,D]=eig(A)` returns a diagonal matrix `D` whose entries are the eigenvalues of `A` and a matrix `V` whose columns are the normalized eigenvectors of `A`, such that `A*V=V*D`. In general `V` is singular. The command `eigs(A)` returns only some of the eigenvalues, by default the six largest in magnitude.

$\gg$ `poly(A)` returns a row vector with $(n+1)$ elements, whose entries are the coefficients of the characteristic polynomial (16) of `A`.

$\gg$ `norm(A)` returns the 2-norm for both vectors and matrices. Anyway there are some differences in the two cases

- If `A` is a vector, the p-norm defined in (17) can be evaluated typing `norm(A,p)`, where `p` can be either a real number or the string `inf`, to evaluate the ∞ norm.

- If `A` is a matrix, the argument `p` in the command `norm(A,p)` can be only `1,2,inf,'fro'`, where the returned norms are respectively the $1,2,\infty$ or Frobenius matrix norms defined in section 4.2.

≫ `[U,T]=schur(A)` returns an upper *quasi*-triangular matrix `T` and a Unitary matrix `U` such that `A = U*T*U'`. The matrix `T` presents the real eigenvalues of `A` on the diagonal and the complex eigenvalues in 2-by-2 blocks on the diagonal. The command `[U,T] = rsf2csf(U,T)` can be used to get an upper-triangular `T` from an upper quasi-triangular `T`.

≫ `[U,S,V]=svd(A)` returns a diagonal, in general rectangular, matrix `S`, whose entries are the singular values of `A`, and two Unitary matrices `U` and `V`, stacking up in columns the left and right singular vectors of `A`, such that `A=U*S*V'`. The command `svds(A)` returns only some of the singular values, by default the five largest.

References

[1] F. R. Gantmacher. *The Theory of Matrices*. Chelsea Publishing Company, New York, N.Y., 1959. Two volumes.

[2] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[3] B. Noble and J. W. Daniel. *Applied Linear Algebra*. Prentice-Hall, Third edition, 1988.

[4] G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, Cambridge, MA, 3rd edition, 2003.