Chapter 3
Basics of Linear Algebra

Vector space, linear independence, and basis

- $\mathbf{x} = [x_1 \;\; x_2 \;\; \cdots \;\; x_n]^T$ is a vector in $\mathcal{R}^n$.
- Given a set of m vectors $\mathbf{x}_1, \; \mathbf{x}_2, \ldots, \; \mathbf{x}_m$ of the same dimension, $\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_m\mathbf{x}_m$ is called a linear combination of the vectors.
- If

$$\mathbf{x}_1 = \alpha_2\mathbf{x}_2 + \alpha_3\mathbf{x}_3 + \ldots + \alpha_m\mathbf{x}_m,$$

  x_1 is said to be linearly dependent on $\mathbf{x}_2, \; \mathbf{x}_3, \; \ldots, \; \mathbf{x}_m$.
- The set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m\}$ in $\mathcal{R}^n$ is said to be linearly independent if

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \cdots + \alpha_m\mathbf{x}_m = 0 \tag{1}$$

  implies that $\alpha_i = 0, \; i = 1, 2, \cdots, m$.
- If $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m$ are linearly dependent, there exist some non-zero α_i such that (1) holds. Suppose $\alpha_1 \neq 0$, then

$$\mathbf{x}_1 = -\frac{1}{\alpha_1}(\alpha_2\mathbf{x}_2 + \cdots + \alpha_m\mathbf{x}_m)$$

  i.e. $\mathbf{x}_1$ can be expressed as a linear combination of others.

Why is linear independence important? If a set of vectors is linearly dependent, then we can get rid of one or perhaps more of the vectors until we get a linearly independent set. This set is then the smallest "truly essential" set with which we can work.

- Consider a set of n linearly independent vectors, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, each with n components. All the possible linear combinations of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ form the vector space $\mathcal{R}^n$. This is the span of the n vectors.
- $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is a basis of $\mathcal{R}^n$.
- For example, $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ forms a basis of $\mathcal{R}^2$.
- The dimension of a linear space is defined as the maximum number of linearly independent vectors in the space.

Norms of Vectors

- A norm is a function $\|\cdot\| : \mathcal{C}^n \rightarrow \mathcal{R}$ that assigns a real-valued length to each vector.

  For all vectors $\mathbf{x}$ and $\mathbf{y}$ and all scalars $\alpha \in \mathcal{C}$, a norm must satisfy

  – $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ only if $\mathbf{x} = 0$
  – $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
  – $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$

- In general, 'p' norms are defined by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

  In particular, we have

  – 1-norm: $\quad \|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|.$

- Let $\mathbf{q}_1,\ \mathbf{q}_2,\ \cdots,\ \mathbf{q}_n$ be a basis of $\mathcal{R}^n$. Then, any $\mathbf{x} \in \mathcal{R}^n$ can be expressed as

$$\mathbf{x} = \alpha_1\mathbf{q}_1 + \alpha_2\mathbf{q}_2 + \cdots \alpha_n\mathbf{q}_n = Q\bar{\mathbf{x}}$$

where

$$Q = [\mathbf{q}_1\ \mathbf{q}_2\ \cdots\ \mathbf{q}_n], \quad \bar{\mathbf{x}} = [\alpha_1\ \ \alpha_2\ \cdots\ \alpha_n]^T$$

- $\bar{\mathbf{x}}$ is called *representation of* $\mathbf{x}$ *w.r.t. the basis* $\{\mathbf{q}_1, \cdots, \mathbf{q}_n\}$.

Norms of Vectors

- A norm is a function $\|\cdot\| : \mathcal{C}^n \rightarrow \mathcal{R}$ that assigns a real-valued length to each vector.

  For all vectors $\mathbf{x}$ and $\mathbf{y}$ and all scalars $\alpha \in \mathcal{C}$, a norm must satisfy

  - $\|\mathbf{x}\| \geq 0$, and $\|\mathbf{x}\| = 0$ only if $\mathbf{x} = 0$
  - $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$
  - $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$

- In general, 'p' norms are defined by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

  In particular, we have

  - 1-norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|.$

- 2-norm (Euclidean norm): $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^T\mathbf{x}} = \sqrt{\sum_{i=1}^{n} |x_i|^2}$ (Euclidean norm).
- infinite-norm: $\|\mathbf{x}\|_\infty = \max_i |x_i|$.

Matlab representations: norm(x,1); norm(x,2); norm(x,inf)

Orthogonal and orthonormal vectors

- Inner product: $< \mathbf{x}, \mathbf{y} >= \mathbf{x}^*\mathbf{y} = \sum_{i=1}^{m} \bar{x}_i y_i, \quad x, y \in \mathcal{C}^m$
- Schwarz inequality: $| < \mathbf{x}, \mathbf{y} > | \leq \|\mathbf{x}\|\|\mathbf{y}\|$
- Vector $\mathbf{x}$ is orthogonal to vector $\mathbf{y}$ if $\mathbf{x}^*\mathbf{y} = 0$. Further, if $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$, they are orthonormal vectors.

Matrix

A $m \times n$ matrix is written as

$$A = [a_{jk}] = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

- if $m = n$, A is a square matrix. The entries a_{11}, a_{22}, $\ddots$, a_{nn} are called the diagonal entries of A.
- Upper triangular matrix: square matrices with nonzero entries only on and above the main diagonal.
- Lower triangular matrices : nonzero entries only on and below the main diagonal.
- Diagonal matrices : nonzero entries only on the main diagonal.
- Identity matrices : diagonal and all diagonal entries are 1.

Matrix Transpose

Given

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

its transpose is

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \vdots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

Transposition has the following rules:

- $(A^T)^T = A$
- $(A+B)^T = A^T + B^T$
- $(cA)^T = cA^T$
- $(AB)^T = B^T A^T$

Range space, null space and rank of a matrix

- Range space is all possible linear combinations of all columns of matrix A, denoted as $R(A)$. That is, $R(A) = \{y|y = Ax, \ \forall x \in \mathcal{R}^n\}$.
- The rank of A is the dimension of $R(A)$ or equivalently the number of linearly independent columns or rows of A.
- For $A \in \mathcal{R}^{m\times n}$, $rank(A) \leq min(m, n)$.
- A vector $\mathbf{x}$ is called null vector of A if $A\mathbf{x} = 0$.
- Nullity is the maximum number of linearly independent null vectors of A, and
$$Nullity\ of\ A = n - rank(A)$$
where n is the number of columns of A.
- The null space of a matrix $A \in \mathcal{R}^{n\times n}$, denoted as N (A), is the vector space $\{x \in \mathcal{R}^n : \ Ax = 0\}$.
- Linear system of equations $A\mathbf{x} = \mathbf{y}$ admits a solution iff
$$rank(A) = rank([A \ \ \mathbf{y}])$$
- Given $A \in \mathcal{R}^{m\times n}$, for any $\mathbf{y}$, $A\mathbf{x} = y$ exists a solution iff $rank(A) = m$ (full row rank).

Parameterization of all solutions of $A\mathbf{x} = \mathbf{y}$

Given $A \in \mathcal{R}^{m\times n}$, let $\mathbf{x}_p$ be a solution of $A\mathbf{x} = \mathbf{y}$. Then, if $rank(A) = n$, the solution $\mathbf{x}_p$ is unique. If $rank(A) < n$, given any real $\alpha_i,\ i = 1, 2, \cdots, k = n - rank(A)$,

$$\mathbf{x} = \mathbf{x}_p + \alpha_1\mathbf{n}_1 + \alpha_2\mathbf{n}_2 + \cdots + \alpha_k\mathbf{n}_k$$

is a solution, where $\{\mathbf{n}_1, \cdots, \mathbf{n}_k\}$ is a basis of the null space of A.

Example 3.1

$$A\mathbf{x} = \begin{bmatrix} 0 & 1 & 1 & 2 \\ 1 & 2 & 3 & 4 \\ 2 & 0 & 2 & 0 \end{bmatrix}\mathbf{x} = \begin{bmatrix} -4 \\ -8 \\ 0 \end{bmatrix}$$

It is clear that $rank(A) = 2$ *and*

$$\mathbf{n}_1 = [1\ \ 1\ \ -1\ \ 0]^T, \quad \mathbf{n}_2 = [0\ \ 2\ \ 0\ \ -1]^T$$

form a basis of the null space of A. *Further,* $\mathbf{x}_p = [0\ \ -4\ \ 0\ \ 0]^T$ *is a solution. Hence, the general solution can be*

$$\mathbf{x} = \mathbf{x}_p + \alpha_1\mathbf{n}_1 + \alpha_2\mathbf{n}_2$$

Determinant

- 2×2 matrices:

$$det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

- 3×3 matrices:

$$\begin{aligned} det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & k \end{bmatrix} &= a \begin{vmatrix} e & f \\ h & k \end{vmatrix} - b \begin{vmatrix} d & f \\ g & k \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= aek + bfg + cdh - ahf - bdk - ceg \end{aligned}$$

- Determinate of A is non-zero iff it is full rank.
- Some properties:

$$\det(AB) = \det(BA) = \det(A)\det(B)$$
$$\det(A^T) = \det(A)$$
$$\det(cA) = c^n \det(A)$$

$$\det\left(\begin{bmatrix} X & Z \\ 0 & Y \end{bmatrix}\right) = \det(X)\det(Y)$$

Matrix inversion

- Suppose a square matrix A is of full rank. Then, it is invertible and

$$A^{-1} = \frac{1}{\det(A)} adj(A)^T$$

- 2×2 matrices:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

- 3×3 matrices:

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \frac{1}{\det(A)}\begin{bmatrix} ek - fh & -(bk - ch) & bf - ce \\ -(dk - fg) & ak - cg & -(af - cd) \\ dh - eg & -(ah - bg) & ae - bd \end{bmatrix}$$

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

Eigenvalue-Eigenvector Decomposition

- Let λ and x ($\bar{x}$) be an eigenvalue and right-eigenvector (left-eigenvector) of A. Then,
$$Ax = \lambda x, \quad \bar{x}^* A = \lambda \bar{x}^*$$
- Eigenvalues and eigenvectors can be computed as
$$\det(\lambda I - A) = 0, \quad (\lambda I - A)x = 0$$
- $\det(\lambda I - A) = \lambda^n + a_1\lambda^{n-1} + \ldots + a_{n-1}\lambda + a_n$ is a polynomial, called characteristic polynomial and $\det(\lambda I - A) = 0$ characteristic equation.
- $\det(\lambda I - A) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$, where $\lambda_i,\ i = 1, 2, \cdots, n$ are eigenvalues and correspondingly, $(\lambda_i I - A)x_i = 0$, with $x_i \neq 0$ being an eigenvector.

Example 3.2 *Consider*

$$A = \begin{bmatrix} -5 & 2 \\ 2 & -2 \end{bmatrix}$$

$$\det(\lambda I - A) = \begin{vmatrix} \lambda+5 & -2 \\ -2 & \lambda+2 \end{vmatrix} = (\lambda+5)(\lambda+2) - 4 = (\lambda+1)(\lambda+6) = 0$$

$$\Longrightarrow \lambda_1 = -1,\ \lambda_2 = -6$$

To obtain an eigenvector associated with $\lambda_1 = -1$*,*

$$(-1I - A)x_1 = \begin{bmatrix} 4 & -2 \\ -2 & 1 \end{bmatrix} x_1 = 0, \implies x_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

For $\lambda_2 = -6$*,*

$$(-6I - A)x_2 = \begin{bmatrix} -1 & -2 \\ -2 & -4 \end{bmatrix} x_2 = 0, \implies x_2 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

It can be easily checked that x_1 *and* x_2 *are linearly independent.*

Similarity Transformation

- Suppose A has n linearly independent eigenvectors $x_1, x_2, \cdots, x_n$ associated with eigenvalues $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$ (not necessarily distinct). Since $Ax_i = \lambda_i x_i$, $A\,[x_1\ x_2\ \cdots\ x_n] = [x_1\ x_2\ \cdots\ x_n]\,diag\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$. Then, with $T = [x_1\ x_2\ \ \cdots\ x_n]$, $AT = T diag\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$, i.e.

$$T^{-1}AT = diag\{\lambda_1, \lambda_2, \cdots, \lambda_n\} \implies A = T diag\{\lambda_1, \lambda_2, \cdots, \lambda_n\} T^{-1}$$

- If A has n distinct eigenvalues, then its eigenvectors $x_1,\ x_2,\ \cdots, x_n$ are linearly independent and A is diagonalizable.

- In general, there exists a nonsingular transformation T and an integer k such that

$$T^{-1}AT = diag\{J_1, J_2, \cdots, J_k\}$$

where

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ & & & \lambda_i \end{bmatrix}$$

is called a Jordan Block. J is known as the *Jordan Canonical form* of A.

Generalized eigenvectors

- An eigenvalue with multiplicity 2 or higher is called a repeated eigenvalue.
- If A has repeated eigenvalues, it may not have a diagonal form but contains Jordan blocks.
- Consider a $n \times n$ matrix with multiplicity n (only one distinct eigenvalue λ).
- Suppose $rank(A - \lambda I) = n - 1$. Then, $(A - \lambda I)\mathbf{v} = \mathbf{0}$ has only one independent solution. That is, A has only one eigenvector associated with λ.
- A vector $\mathbf{v}$ is said to be a generalized eigenvector of grade n if

$$(A - \lambda I)^n \mathbf{v} = 0, \;\; (A - \lambda I)^{n-1} \mathbf{v} \neq 0$$

  Apparently, if $n = 1$, $\mathbf{v}$ is an eigenvector.

- Define

$$\mathbf{v}_n = \mathbf{v}, \;\; \mathbf{v}_{n-1} = (A - \lambda I)\mathbf{v}_n, \quad \cdots, \; \mathbf{v}_1 = (A - \lambda I)\mathbf{v}_2$$

- It can be known that

$$(A - \lambda I)\mathbf{v}_1 = 0,\ (A - \lambda I)^2\mathbf{v}_2,\ \cdots,\ (A - \lambda I)^n\mathbf{v}_n = 0$$

- The set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ is called a chain of generalized eigenvectors of length n.
- For $n = 4$, we have

$$\begin{aligned} A\mathbf{v}_1 &= \lambda\mathbf{v}_1 \\ A\mathbf{v}_2 &= \mathbf{v}_1 + \lambda\mathbf{v}_2 \\ A\mathbf{v}_3 &= \mathbf{v}_2 + \lambda\mathbf{v}_3 \\ A\mathbf{v}_4 &= \mathbf{v}_3 + \lambda\mathbf{v}_4 \end{aligned}$$

i.e.

$$A[\mathbf{v}_1\ \ \mathbf{v}_2\ \ \mathbf{v}_3\ \ \mathbf{v}_4] = [\mathbf{v}_1\ \ \mathbf{v}_2\ \ \mathbf{v}_3\ \ \mathbf{v}_4]J$$

where

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

- If $rank(A - \lambda I) = n - 2$,

$$(A - \lambda I)\mathbf{q} = 0$$

will have two linearly independent eigenvectors $\mathbf{v}$ and $\mathbf{u}$. There exist two chains of generalized eigenvectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ and $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_l\}$ with $k + l = n$.

- Form $T = [\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{u}_1, \cdots, \mathbf{u}_l]$. Then,

$$T^{-1}AT = \hat{A} = diag\{J_k, J_l\}$$

where J_k and J_l are Jordan blocks of order k and l.

- For example,

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

occurs when $rank(A - \lambda I) = 4 - 3 = 1$.

- The fact that A can be decomposed as

$$A = T\hat{A}T^{-1}$$

can be used to prove a number of useful results.

- For example,

$$det(A) = det(T\hat{A}T^{-1}) = det(T)det(T^{-1})det(\hat{A}) = \lambda_1\lambda_2\cdots\lambda_n$$

- Hence, A is non-singular iff all its eigenvalues are non-zero.
- Observe for Jordan block of order 4,

$$(J-\lambda I) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (J-\lambda I)^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad (J-\lambda I)^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and $(J - \lambda I)^k = 0$ when $k \geq 4$. The matrix is called *nilpotent*.

- $trace(A) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$.

Example 3.3 *Consider* $A = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix}$. *Its eigenvalues are* $\lambda_1 = 1,\ \lambda_2 = 1,\ \lambda_3 = 2$. *The eigenvector* $\mathbf{v}_3 = [5\ \ 3\ \ 1]'$. *The rank of* $(A - I)$ *is 2=3-1, so only eigenvector associated with the eigenvalue 1. Hence, we need to find generalized eigenvectors.*

It is easy to compute $rank(A - I)^2 = 3 - 2$. *We search a* v *such that* $(A - I)^2\mathbf{v} = 0$ *but* $(A - I)\mathbf{v} \neq 0$. *Clearly,* $\mathbf{v} = [0\ \ 1\ \ 0]'$ *is a choice. Then,*

$$\mathbf{v}_2 = \mathbf{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_1 = (A - I)\mathbf{v} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Form $T = [\mathbf{v}_1\ \ \mathbf{v}_2\ \ \mathbf{v}_3]$. *Then,*

$$\hat{A} = T^{-1}AT = \left[\begin{array}{cc|c} 1 & 1 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 2 \end{array}\right]$$

Example 3.4 *Find the Jordan canonical form of the system*

$$\dot{x} = \begin{bmatrix} 0 & 1 \\ -9 & -6 \end{bmatrix} x + \begin{bmatrix} 4 \\ 2 \end{bmatrix} u$$

- *Find the eigenvalues of A:* $|A - \lambda I| = (\lambda + 3)^2 = 0, \ \lambda_{1,2} = -3$
- *Check the rank of* $A - \lambda_1 I$*:*

$$A + 3I = \begin{bmatrix} 3 & 1 \\ -9 & -3 \end{bmatrix}, \ rank(A + 3I) = 1 = 2 - 1,$$

implying that there exists one independent eigenvector and need to find one generalized eigenvector.

- *Find eigenvector and generalized eigenvector:*

$$(A + 3I)v_1 = 0, \ \ v_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}; \ (A + 3I)v_2 = v_1, \ \ v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- *Form* T*:* $T = [v_1 \ \ v_2] = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}, \ T^{-1} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$
- $T^{-1}AT = \begin{bmatrix} -3 & 1 \\ 0 & -3 \end{bmatrix}, \ T^{-1}B = \begin{bmatrix} 4 \\ 14 \end{bmatrix}.$

Functions of a Square Matrix

- Let $f(\lambda)$ be a polynomial such as $f(\lambda) = \lambda^3 + 2\lambda^2 - 6$. Then, $f(A) = A^3 + 2A^2 - 6I$.
- Introduce a similarity transformation $\hat{A} = T^{-1}AT$ or $A = T\hat{A}T^{-1}$. Since for any positive integer,

$$A^k = (T\hat{A}T^{-1})(T\hat{A}T^{-1})\cdots(T\hat{A}T^{-1}) = T\hat{A}^kT^{-1}$$

- Thus,

$$f(A) = Tf(\hat{A})T^{-1}$$

  implying that $f(A) = 0$ iff $f(\hat{A}) = 0$.
- A monic polynomial is a polynomial with 1 as its leading coefficient.
- The minimal polynomial of A is the monic polynomial $\psi(\lambda)$ of least degree such that $\psi(A) = 0$.
- The characteristic polynomial of A is

$$\Delta(\lambda) = det(\lambda I - A) = \Pi_i(\lambda - \lambda_i)^{n_i}$$

- Let $\bar{n}_i$ be the largest order of all Jordan blocks associated with λ_i. Clearly, $\bar{n}_i \leq n_i$.

- Then, the minimal polynomial can be expressed as

$$\psi(\lambda) = \Pi_i(\lambda - \lambda_i)^{\bar{n}_i}$$

- For example, for

$$\hat{A} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & 0 & \lambda_2 \end{bmatrix}$$

  we have

$$\psi(\lambda) = (\lambda - \lambda_1)^3(\lambda - \lambda_2)$$

- In fact, using the nilpotent property, it can be shown that $\psi(A) = 0$.
- *Cayley-Hamilton theorem*: Let

$$\Delta(\lambda) = \det(\lambda I - A) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n$$

  Then

$$\Delta(A) = A^n + a_1 A^{n-1} + \cdots + a_{n-1}A + a_n I = 0$$

  Proof The proof follows from the fact that $\psi(\lambda)$ is a factor of $\Delta(\lambda)$, i.e., $\Delta(\lambda) = \psi(\lambda)h(\lambda)$, and $\psi(A) = 0$.

- Cayley-Hamilton theorem implies that A^k with $k \geq n$ can be expressed as a linear combination of $\{I, A, A^2, \cdots, A^{n-1}\}$.
- Thus, for any polynomial $f(\lambda)$, $f(A)$ can always be expressed as

$$f(A) = \beta_0 I + \beta_1 A + \cdots + \beta_{n-1} A^{n-1} \tag{2}$$

for some β_i.

- If the minimal polynomial with degree $\bar{n} = \sum \bar{n}_i$ is available, then $f(A)$ can be written as a linear combination of $\{I, A, A^2, \cdots, A^{\bar{n}-1}\}$.
- To compute (2), using long division, we rewrite $f(\lambda)$ as

$$f(\lambda) = g(\lambda)\Delta(\lambda) + h(\lambda)$$

where the remainder $h(\lambda)$ is of degree less than n.

- Then,

$$f(A) = g(A)\Delta(A) + h(A) = h(A) \tag{3}$$

- Hence, instead of long division, let

$$h(\lambda) = \beta_0 + \beta_1\lambda + \cdots + \beta_{n-1}\lambda^{n-1}$$

If all eigenvalues λ_i of A are distinct, we can solve the n equations for $\beta_i,\ i = 1, 2, \cdots, n$:

$$f(\lambda_i) = g(\lambda_i)\Delta(\lambda_i) + h(\lambda_i) = h(\lambda_i)$$

- For repeated eigenvalues, the characteristic polynomial is

$$\Delta(\lambda) = \prod_{i=1}^{m}(\lambda - \lambda_i)^{n_i}$$

with $n = \sum_{i=1}^{m} n_i$. Then, β_i can be solved from

$$f^{(l)}(\lambda_i) = h^{(l)}(\lambda_i),\ \ l = 0, 1, \cdots, n_i - 1,\ \ i = 1, 2, \cdots, m$$

where

$$f^{(l)}(\lambda_i) = \frac{d^l f(\lambda)}{d\lambda^l}|_{\lambda=\lambda_i}$$

Example 3.5 *Compute* e^{At} *where* $A = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ 1 & 0 & 3 \end{bmatrix}$.

- $$\Delta(\lambda) = |\lambda I - A| = (\lambda - 1)^2(\lambda - 2).$$

  Let $f(\lambda) = e^{\lambda t}$ *and* $h(\lambda) = \beta_0 + \beta_1\lambda + \beta_2\lambda^2$. *Then,*

$$\begin{aligned} f(1) &= h(1) \rightarrow e^t = \beta_0 + \beta_1 + \beta_2 \\ f'(1) &= h'(1) \rightarrow te^t = \beta_1 + 2\beta_2 \\ f(2) &= h(2) \rightarrow e^{2t} = \beta_0 + 2\beta_1 + 4\beta_2 \end{aligned}$$

- *Solving, we get*

$$\beta_0 = -2te^t + e^{2t}, \;\; \beta_1 = 3te^t + 2e^t - 2e^{2t}, \;\; \beta_2 = e^{2t} - e^t - te^t$$

- *Hence,*

$$e^{At} = \beta_0 I + \beta_1 A + \beta_2 A^2 = \begin{bmatrix} 2e^t - e^{2t} & 0 & 2e^t - 2e^{2t} \\ 0 & e^t & 0 \\ e^{2t} - e^t & 0 & 2e^{2t} - e^t \end{bmatrix}$$

Example 3.6 *Consider the Jordan block*

$$\hat{A} = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_1 & 1 \\ 0 & 0 & 0 & \lambda_1 \end{bmatrix}$$

- *Instead of choosing* $h(\lambda) = \beta_0 + \beta_1\lambda + \beta_2\lambda^2 + \beta_3\lambda^3$, *it is simpler to choose*

$$h(\lambda) = \beta_0 + \beta_1(\lambda - \lambda_1) + \beta_2(\lambda - \lambda_1)^2 + \beta_3(\lambda - \lambda_1)^3$$

- *Then,* $f(\lambda) = h(\lambda)$ *on the spectrum of* $\hat{A}$ *yields*

$$\beta_0 = f(\lambda_1),\ \beta_1 = f'(\lambda_1),\ \beta_2 = \frac{f''(\lambda_1)}{2!},\ \beta_3 = \frac{f^{(3)}(\lambda_1)}{3!}$$

- $$f(\hat{A}) = f(\lambda_1)I + \frac{f'(\lambda_1)}{1!}(\hat{A} - \lambda_1 I) + \frac{f''(\lambda_1)}{2!}(\hat{A} - \lambda_1 I)^2 + \frac{f^{(3)}(\lambda_1)}{3!}(\hat{A} - \lambda_1 I)^3$$

  which gives

$$f(\hat{A}) = \begin{bmatrix} f(\lambda_1) & f'(\lambda_1)/1! & f''(\lambda_1)/2! & f^{(3)}\lambda_1)/3! \\ 0 & f(\lambda_1) & f'(\lambda_1)/1! & f''(\lambda_1)/2! \\ 0 & 0 & f(\lambda_1) & f'(\lambda_1)/1! \\ 0 & 0 & 0 & f(\lambda_1) \end{bmatrix}$$

- For $f(\lambda) = e^{\lambda t}$, we get

$$e^{\hat{A}t} = \begin{bmatrix} e^{\lambda_1 t} & te^{\lambda_1 t} & t^2 e^{\lambda_1 t}/2! & t^3 e^{\lambda_1 t}/3! \\ 0 & e^{\lambda_1 t} & te^{\lambda_1 t} & t^2 e^{\lambda_1 t}/2! \\ 0 & 0 & e^{\lambda_1 t} & te^{\lambda_1 t} \\ 0 & 0 & 0 & e^{\lambda_1 t} \end{bmatrix}$$

Matrix norms induced by vector norms

- An $m \times n$ matrix is represented as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & & \vdots & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

- For any $A \in \mathcal{C}^{m \times n}$, the induced matrix norm is

$$\|A\| = \sup_{0 \neq x \in \mathcal{C}^n} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1, x \in \mathcal{C}^n} \|Ax\|$$

General matrix norm

- For all matrices A and B and all scalars $\alpha \in \mathcal{C}$, a matrix norm must satisfy
  - $\|A\| \geq 0$, and $\|A\| = 0$ only if $A = 0$
  - $\|A + B\| \leq \|A\| + \|B\|$
  - $\|\alpha A\| = |\alpha| \|A\|$
- Frobenius norm

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}} = \sqrt{tr(A^* A)}, \ \ \forall A \in \mathcal{C}^{m \times n}$$

- $\|A\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}|$ (maximum 'column sum').
- $\|A\|_2 = \bar{\sigma}(A) = \sqrt{\lambda_{\max}(A^* A)}$ (maximum singular value).
- $\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$ (maximum 'row sum').

Note: $\|A\|_1, \ \|A\|_2$ and $\|A\|_\infty$ are induced matrix norms but $\|A\|_F$ is not.

Positive definite matrix

- A real matrix A is symmetric if $A = A^T$.
- All eigenvalues of real symmetric matrix must be real.
- A real symmetric matrix A is positive definite $(A > 0)$ if all of its eigenvalues are positive.
- A real symmetric matrix A is positive semidefinite $(A \geq 0)$ if all of its eigenvalues are greater than or equal to 0.
- *Rayleigh-Ritz inequality:* For a real symmetric matrix A, its quadratic form satisfies

$$\lambda_{\min} x^T x \leq x^T A x \leq \lambda_{\max} x^T x$$

  where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of A.
- Matrices $A > B (A \geq B)$ if $A - B > 0 (A - B \geq 0)$.

Some useful determinant properties:

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det(B - CA^{-1}D) \quad (if\ A\ is\ nonsingular)$$

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(B)\det(A - DB^{-1}C) \quad (if\ A\ is\ nonsingular)$$

$$\det(I_m + AB) = \det(I_n + BA) \quad (A \in \mathcal{C}^{m\times n},\ \ B \in \mathcal{C}^{n\times m})$$

Some useful matrix inversion formulae

- Given a nonsingular $A \in \mathcal{C}^{m\times n}$, and $u, v \in \mathcal{C}^n$ satisfying $v^* A^{-1} u \neq -1$, then

$$(A + uv^*)^{-1} = A^{-1} - \frac{A^{-1}uv^*A^{-1}}{1 + v^*A^{-1}u}$$

- Sherman-Morrison-Woodbury formula (matrix inversion lemma)

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(DA^{-1}B + C^{-1})^{-1}DA^{-1}$$

In particular, $(I + AB)^{-1}A = A(I + BA)^{-1}$.

Matrix differentiation

- Let $A(t)$ and $B(t)$ be matrices with differentiable entries. Then,

$$\frac{d}{dt}[A(t)B(t)] = \left[\frac{d}{dt}A(t)\right]B(t) + A(t)\left[\frac{d}{dt}B(t)\right]$$

- Further, if $A(t)$ is nonsingular,

$$\frac{d}{dt}\left[A^{-1}(t)\right] = -A^{-1}(t)\left[\frac{d}{dt}A(t)\right]A^{-1}(t)$$

Proof. Note that $A(t)A^{-1}(t) = I$. The result follows by differentiating both sides of the equation.

Singular Value Decomposition:

- For any real or complex $m \times n$ matrix A, there exist $U \in \mathcal{R}^{m\times m}(\mathcal{C}^{m\times m})$, $V \in \mathcal{R}^{n\times n}(\mathcal{C}^{n\times n})$ and $\Sigma \in \mathcal{R}^{m\times n}$ such that

$$A = U\Sigma V^*, \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

  where $U^*U = I$, $V^*V = I$ (unitary or orthogonal matrices) and

$$\Sigma_1 = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \cdots & \\ & & & \sigma_p \end{bmatrix}_{p\times p}$$

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$$

- $\sigma_{\max}(AB) \leq \sigma_{\max}(A)\sigma_{\max}(B)$.

Laplace Transform

- Given an $m \times n$ matrix function defined for $t \in [0, \infty)$, Laplace transform is

$$F(s) = L[F(t)] = \int_0^\infty F(t)e^{-st}dt$$

- Basic properties

$$L[\dot{F}(t)] = sL[F(t)] - F(0)$$

$$L\left[\int_0^\infty F(\sigma)d\sigma\right] = \frac{1}{s}L[F(t)]$$

$$L\left[\int_0^\infty F(t-\sigma)G(\sigma)d\sigma\right] = L[F(t)]L[G(t)]$$

$$\lim_{t\to 0} F(t) = \lim_{s\to\infty} sF(s)$$

$$\lim_{t\to\infty} F(t) = \lim_{s\to 0} sF(s)$$

Z-Transform

- Given an $m \times n$ matrix sequence defined for $k \in \{0, 1, 2, \cdots, \infty)$, Z-transform is

$$F(z) = Z[F(k)] = \sum_{k=0}^{\infty} F(k) z^{-k}$$

- Basic properties

$$Z[F(k-1)] = z^{-1} Z[F(k)]$$

$$Z[F(k+1)] = zZ[F(k)] - zF(0)$$

$$Z\left[\sum_{j=0}^{k} F(k-j)G(j)\right] = Z[F(k)]Z[G(k)]$$

$$F(0) = \lim_{z \to \infty} F(z)$$

$$\lim_{k \to \infty} F(k) = \lim_{z \to 1} (z-1)F(z)$$

Exercise 3.1 Consider the equation

$$\mathbf{x}(n) = A^n\mathbf{x}(0) + A^{n-1}bu(0) + A^{n-2}bu(1) + \cdots + Abu(n-2) + bu(n-1)$$

where A is an $n \times n$ matrix and b is an $n \times 1$ column vector. Under what conditions on A and b will there exist $u(0), u(1), \cdots, u(n-1)$ to meet the equation for any $\mathbf{x}(n)$ and $\mathbf{x}(0)$?

3.2 Find the Jordan forms of the following matrices:

$$A_1 = \begin{bmatrix} 1 & 4 & 10 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

3.3 Consider the companion-form matrix:

$$A = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Show that the characteristic polynomial is given by

$$\Delta(\lambda) = \lambda^4 + \alpha_1\lambda^3 + \alpha_2\lambda^2 + \alpha_3\lambda + \alpha_4$$

Show also that if λ_i is an eigenvalue of A, then $[\lambda_i^3 \;\; \lambda_i^2 \;\; \lambda_i \;\; 1]^T$ is an eigenvector of A associated with λ_i.

3.4 Compute A^{100}, where

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