Linear Algebra Review

- Vector space and subspaces
- Independence, basis and dimension
- Range, nullspace, inverse and rank
- Similarity transform, norms and inner product
- Eigenvectors and eigenvalues
- Jordan canonical form and Cayley-Hamilton theorem

Vector Spaces

- Concept of a vector space.

  Consider the familiar spaces $\boldsymbol{R}^1$, $\boldsymbol{R}^2$, $\ldots$ $\boldsymbol{R}^n$.

  – $\boldsymbol{R}^1$ is a 1-dimensional space, i.e., a line.

  – $\boldsymbol{R}^2$ is a 2-dimensional or a plane.

  Can we generalize operations performed in this spaces?

- A linear (vector) space consists of

  – a set $\boldsymbol{\mathcal{V}}$.

  – a vector sum $+ : \boldsymbol{\mathcal{V}} \times \boldsymbol{\mathcal{V}} \rightarrow \boldsymbol{\mathcal{V}}$.

  – scalar multiplication : $\boldsymbol{R} \times \boldsymbol{\mathcal{V}} \rightarrow \boldsymbol{\mathcal{V}}$.

  – a zero element : $0 \in \boldsymbol{\mathcal{V}}$.

Vector Spaces

- The following properties hold for a vector space.

  – $x + y = y + x, \forall\, x, y \in \boldsymbol{\mathcal{V}}$ (commutative).

  – $(x + y) + z = x + (y + z), \forall\, x, y \in \boldsymbol{\mathcal{V}}$ (associative).

  – $0 + x = x, \forall\, x \in \boldsymbol{\mathcal{V}}$.

  – $\forall\, x \in \boldsymbol{\mathcal{V}}\ \exists\, (-x) \in \boldsymbol{\mathcal{V}}$ such that $x + (-x) = 0$.

  – $(\alpha\beta)x = \alpha(\beta x), \forall\, \alpha, \beta \in \boldsymbol{R}, \forall\, x \in \boldsymbol{\mathcal{V}}$.

  – $\alpha(x + y) = \alpha x + \alpha y, \forall\, i\alpha \in \boldsymbol{R}, \forall\, x, y \in \boldsymbol{\mathcal{V}}$.

  – $(\alpha + \beta)x = \alpha x + \beta x, \forall\, \alpha, \beta \in \boldsymbol{R}, \forall\, x \in \boldsymbol{\mathcal{V}}$.

  – $1 \cdot x = x, \forall\, x \in \boldsymbol{\mathcal{V}}$.

Vector Spaces

- Examples.

  – $\boldsymbol{\mathcal{V}}_1 = \boldsymbol{R}^n$ with standard vector addition and scalar multiplication.

  – $\boldsymbol{\mathcal{V}}_2 = \{0\}$ (where $0 \in \boldsymbol{R}^n$).

  – $\boldsymbol{\mathcal{V}}_3 = \text{span}(v_1, v_2, \ldots, v_k)$ where $\text{span}(v_1, v_2, \ldots, v_k) = \{\alpha_1 v_1 + \ldots + \alpha_k v_k \mid \alpha_i \in \boldsymbol{R}\}$.

  – $\boldsymbol{\mathcal{V}}_4 = \{x : \boldsymbol{R}_+ \rightarrow \boldsymbol{R}^n \mid x \text{ is differentiable}\}$.
  vector sum is the sum of the functions, i.e., $(x + z)(t) = x(t) + z(t)$ and scalar multiplication is defined by $(\alpha x)(t) = \alpha x(t)$.

  – $\boldsymbol{\mathcal{V}}_5 = \{x \in \boldsymbol{\mathcal{V}}_4 : \dot{x} = Ax\}$
  (points in $\boldsymbol{\mathcal{V}}_5$ are trajectories of the linear system $\dot{x} = Ax$).

Vector Spaces

- A subspace of a vector space is a subset of a vector space which is in itself a vector space.
  - $\mathcal{V}_1$, $\mathcal{V}_2$ and $\mathcal{V}_3$ are subspaces of R^n.
  - $\mathcal{V}_5$ is a subspace of $\mathcal{V}_4$.

- For a subspace, we only need to check that
  - vector addition holds and
  - scalar multiplication holds.

  Other properties follow and are automatically satisfied.

Independent Vectors, Basis and Dimension

- Vectors $v_1, v_2, \ldots, v_k \in R^n$ are independent iff

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \; + \alpha_k v_k = 0$$
$$\Rightarrow \alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$$

- Saying vectors are independent is equivalent to

  coefficients of $\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k = v$ are uniquely determined, i.e.,

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k$$
$$= \beta_1 v_1 + \beta_2 v_2 + \ldots + \beta_k v_k$$

  implies that $\alpha_1 = \beta_1,\ \alpha_2 = \beta_2,\ \ldots, \alpha_k = \beta_k$.

Independent Vectors, Basis and Dimension

- Independence is also equivalent to saying

  vector v_i, $i \in \{1, \ldots, k\}$ cannot be expressed as a linear combination of the other vectors $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k$.

- Linearly dependent vectors. One vector may be expressed as a linear combination of the other vectors.

- A set of vectors $\{v_1, v_2, \ldots, v_k\}$ is a basis for a vector space $\mathcal{V}$ if
  - $v_1, v_2, \ldots, v_k$ span $\mathcal{V}$, i.e., $\mathcal{V} = \operatorname{span}(v_1, v_2, \ldots, v_k)$.
  - $v_1, v_2, \ldots, v_k$ are linearly independent.

Independent Vectors, Basis and Dimension

- If you have a set of basis vectors, then every $v \in \mathcal{V}$ can be uniquely expressed as

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_k v_k$$

- You can have many sets of basis vectors for a given vector space $\mathcal{V}$.

  For a given vector space $\mathcal{V}$, the number of vectors in any basis is the same.

- Definition. The dimension of a vector space, $\dim \mathcal{V}$, is the number of vectors in any basis.

- The nullspace of $A \in R^{m \times n}$ is defined as

$$\mathcal{N}(A) = \{x \in R^n \mid Ax = 0\}$$

  - $\mathcal{N}(A)$ is a set of vectors mapped to zero by $y = Ax$.
  - $\mathcal{N}(A)$ is a set of vectors othogonal to all rows of A.

- Given $y = Ax$
  - if $z \in \mathcal{N}(A)$, then $y = A(x + z)$.
  - conversely, if $y = A\tilde{x}$, then $\tilde{x} = x + z$ for some $z \in \mathcal{N}(A)$.

- A zero nullspace. $\mathcal{N}(A) = \{0\} \Leftrightarrow$
  - x can always be uniquely determined from $y = Ax$ (i.e., the linear map $y = Ax$ does not lose information).
  - columns of A are independent (hence, a basis for their span).
  - A has a left inverse, i.e., the is a matrix $B \in R^{n \times m}$ such that $BA = I$.
  - $\det(A^T A) \neq 0$.

- Suppose $z \in \mathcal{N}(A)$.

- The linear equation $y = Ax$ can represent the output due to input x.
  - z is an input with no output result.
  - x and $x + z$ have the same output.

- $\mathcal{N}(A)$ characterizes the freedom of input choice that will result in a given output.

- The linear equation $y = Ax$ can represent measurement of x.
  - z will not be detected by sensors, i.e., you get zero sensor readings.
  - x and $z + z$ are indistinguishable from sensors : $Ax = A(x + z)$.

- $\mathcal{N}(A)$ characterizes ambiguity in x from $y = Ax$.

Nullspace and Range

- The range of $A \in R^{m\times n}$ is defined as

$$\mathcal{R}(A) = \{Ax \mid x \in R^n\} \subseteq R^m$$

- This can be interpreted as
  - $\mathcal{R}(A)$ is the set of vectors that can be 'reached' by the mapping $y = Ax$.
  - $\mathcal{R}(A)$ is equivalent to the span of columns of A.
  - $\mathcal{R}(A)$ is the set of vectors y such that $Ax = y$ has a solution.

- Also called as the column space of A.

Nullspace and Range

- A is called onto if $\mathcal{R}(A) = R^m \Leftrightarrow$
  - $Ax = y$ can be solved in x for any y.
  - the columns of A span R^m.
  - A has a right inverse, i.e., there is a matrix $B \in R^{n\times m}$ such that $AB = I$.
  - the rows of A are independent.
  - $\det(AA^T) \neq 0$.

Nullspace and Range

- Suppose $v \in \mathcal{R}(A)$ and $w \notin \mathcal{R}(A)$.

- Let $y = Ax$ represent a measurement of x.
  - $y = v$ is a possible or consistent sensor signal.
  - $y = w$ is impossible or inconsistent; sensors have failed or model is wrong.

- Let $y = Ax$ represent an output resulting from input x.
  - $y = v$ is a possible output.
  - $y = w$ cannot be an output or result.

  $\mathcal{R}(A)$ characterizes the possible results or outputs.

Inverse and Rank of a Matrix

- $A \in R^{n\times n}$ is invertible or nonsingular if $\det A \neq 0$.

- Equivalently.
  - the columns of A are a basis for R^n.
  - the rows of A are a basis for R^n.
  - $y = Ax$ has a unique solution x for every $y \in R^n$.
  - A has a (left and right) inverse denoted by $A^{-1} \in R^{n\times n}$, with $A^{-1}A = AA^{-1} = I$.
  - $\mathcal{N}(A) = \{0\}$.
  - $\mathcal{R}(A) = R^n$.
  - $\det(A^TA) = \det(AA^T) \neq 0$.

- Definition. Rank of $A \in R^{m \times n}$ as

$$\text{rank}(A) = \dim \mathcal{R}(A)$$

- Useful facts.
  - rank(A) = rank(A^T).
  - rank(A) is the maximum number of independent columns of A.
  - rank(A) is the maximum number of independent rows of A.
  - rank$(A) \leq \min(m, n)$.
  - rank$(A) + \dim \mathcal{N}(A) = n$.

- Interpretation of rank$(A) + \dim \mathcal{N}(A) = n$.

- The rank(A) is the dimension of the set that can be reached by the mapping $y = Ax$.

- The $\dim \mathcal{N}(A)$ is the dimension of the subset of all possible x that gets mapped to zero by $y = Ax$.

- Conservation of dimension. Each dimension of input either appears at the output or gets mapped to zero.

- Definition. Full rank. For $A \in R^{m \times n}$ we always have rank$(A) \leq \min(m, n)$.

  A is full rank if rank$(A) = \min(m, n)$.

- Full rank square matrices are nonsingular.

- For fat matrices ($n > m$), full rank means that the rows are independent.

- For tall matrices ($m > n$), full rank means that the columns are independent.

- Interpretation of rank in coding.

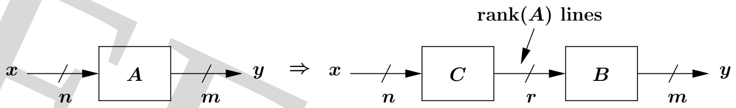
  Rank product. rank$(BC) \leq \min\{\text{rank}(B), \text{rank}(C)\}$.

- Thus, if $A = BC$ with $B \in R^{m \times r}$, $C \in R^{r \times n}$, then rank$(A) \leq r$.

- Conversely, if rank$(A) = r$ then $A \in R^{m \times n}$ can be factored as $A = BC$ with $B \in R^{m \times r}$, $C \in R^{r \times n}$.

Inverse and Rank of a Matrix

- **Transmission channel.**



rank(A) $=$ r is the minimum number of lines needed to faithfully reconstruct y from x.

- **Impact in computing $y = Ax$.**
  - **directly : you need to perform mn operations**
  - **compute $z = Cx$ first, then $y = Bz$: $rn + mr = (m + n)r$ operations.**

Significant savings if $r \ll \min\{m, n\}$.

Coordinate Transformation

- **Standard basis vectors in R^n : $(e_1, e_2, \ldots, e_n)$**

e_i has a 1 in the ith component and zeros in the rest.

$$e_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- **Thus, we can write any $x \in R^n$ as**

$$x = x_1 e_1 + x_2 e_2 + \ldots + x_n e_n$$

The coefficients x_i are the elements (coordinates) of x in the standard basis.

Coordinate Transformation

- **If $(t_1, t_2, \ldots, t_n)$ is another basis for R^n, we may write**

$$x = \tilde{x}_1 t_1 + \tilde{x}_2 t_2 + \ldots + \tilde{x}_n t_n$$

where $\tilde{x}_i$ are the coordinates of x in the new basis.

- **Define $T = [t_1\ t_2\ \ldots\ t_n]$ such that $x = T\tilde{x}$. We have**

$$\tilde{x} = T^{-1} x$$

  - **T is nonsingular since t_i are a basis.**
  - **T^{-1} transforms the standard basis coordinates of x into the t_i coordinates.**

Coordinate Transformation

- **Consider the linear transformation**

$$y = Ax, \qquad A \in R^{n \times n}$$

- **Express y and x in terms of $t_1, t_2, \ldots, t_n$.**

$$x = T\tilde{x}, \qquad y = T\tilde{y}$$

Thus,

$$\tilde{y} = (T^{-1}AT)\tilde{x}$$

  - **$A \rightarrow T^{-1}AT$ is called a similarity transformation.**
  - **a similarity transformation T expresses the linear transformation $y = Ax$ in $t_1, t_2, \ldots, t_n$ coordinates.**

Euclidean Norm

- For $x \in R^n$ we define the (Euclidean) norm as

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} = \sqrt{x^T x}$$

$\|x\|$ is the length of the vector.

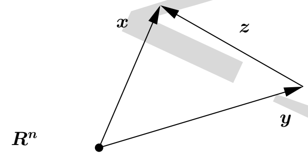
- Properties of the norm.
  - homogeneity : $\|\alpha x\| = |\alpha| \|x\|$
  - triangle inequality : $\|x + y\| \le \|x\| + \|y\|$
  - nonnegativity : $\|x\| \ge 0$
  - definiteness : $\|x\| = 0 \Leftrightarrow x = 0$.

Euclidean Norm

- The distance of the vector from the origin is also given by $\|x\|$.

- The distance between vectors is

$$\text{dist}(x, y) = \|x - y\|$$



Inner Product

- The inner product of two vectors $x, y \in R^n$ is defined as

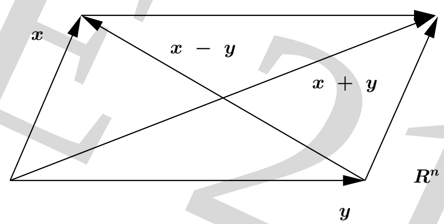
$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = x^T y$$

- Properties of the inner product.
  - $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$
  - $\langle x + y, z \rangle = \langle x, y \rangle + \langle y, z \rangle$
  - $\langle x, y \rangle = \langle y, x \rangle$
  - $\langle x, x \rangle \ge 0$
  - $\langle x, x \rangle = 0 \Leftrightarrow x = 0$

Inner Product

- Parallelogram equality.

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$$



- Cauchy-Schwartz inequality. For any $x, y \in R^n$,

$$|x^T y| \le \|x\| \|y\|$$

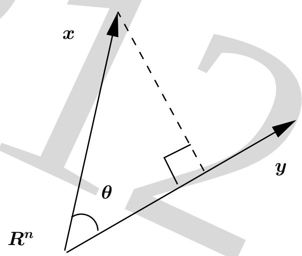
Note that this only holds for the 2-norm.

- Angle between vectors in R^n.

$$\theta = \angle(x, y) = \cos^{-1} \frac{x^T y}{\|x\| \|y\|}$$

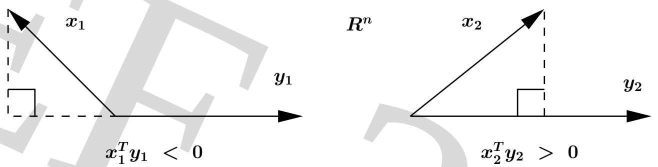
Thus,

$$x^T y = \|x\| \|y\| cos\theta$$



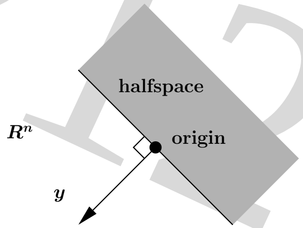
- Special cases and the meaning of $x^T y$.
  - x and y are aligned.
    $\theta = 0 \Rightarrow x^T y = \|x\| \|y\|$.
    if $x \neq 0$, $y = \alpha x$ for some $\alpha \ge 0$
  - x and y are in opposite directions.
    $\theta = \pi \Rightarrow x^T y = -\|x\| \|y\|$.
    if $x \neq 0$, $y = -\alpha x$ for some $\alpha \ge 0$
  - x and y are orthogonal ($x \perp y$).
    $\theta = \pm\pi/2 \Rightarrow x^T y = 0$
  - $x^T y > 0$ means $\angle(x, y)$ is acute.
  - $x^T y < 0$ means $\angle(x, y)$ is obtuse.

- Example. $x_1^T y_1 < 0$ and $x_2^T y_2 > 0$.



- Halfspace with outward normal vector y and boundary passing through the origin.

$$\{x \mid x^T y \le 0\}$$



- Definition. $\lambda \in C$ is an eigenvalue of $A \in R^{n \times n}$ if

$$\mathcal{X}(\lambda) = \det(\lambda I - A) = 0$$

- If λ is an eigenvalue $\Leftrightarrow$

  there exists a nonzero $v \in C^n$ such that $(\lambda I - A)v = 0$, i.e.,

$$Av = \lambda v$$

  Such a v is termed an eigenvector of A associated with the eigenvalue λ.

- **If λ is an eigenvalue $\Leftrightarrow$**

  there exists a nonzero $w \in C^n$ such that $w^T(\lambda I - A) = 0$, i.e.,

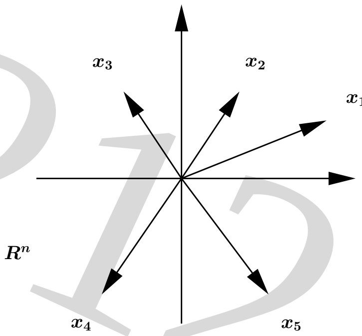
  $$w^T A = \lambda w^T$$

  Such a w is called a left eigenvector of A.

- **Conjugate symmetry. If $v \in C^n$ is an eigenvector associated with $\lambda \in C$, then $\bar{v}$ is the eigenvector associated with $\bar{\lambda}$.**

  $$Av = \lambda v \Rightarrow \bar{Av} = \bar{\lambda v} \Rightarrow A\bar{v} = \bar{\lambda}\bar{v}$$

- **Scaling. If v is an eigenvector, the effect of A on v is similar to scaling v by λ.**
  - **$\lambda \in R$, $\lambda > 0$.** **v and Av point in the same direction.**
  - **$\lambda \in R$, $\lambda < 0$.** **v and Av point in opposite directions.**
  - **$\lambda \in R$, $|\lambda| < 1$.** **v is larger than Av.**
  - **$\lambda \in R$, $|\lambda| > 1$.** **v is smaller than Av.**



- **Diagonalization.**

  Suppose $v_1, v_2, \ldots, v_n$ are linearly independent eigenvectors of $A \in R^{n\times n}$.

  $$Av_i = \lambda_i v_i, \qquad i = 1, \ldots, n$$

- **We can express this as**

  $$A[v_1\ v_2\ \ldots\ v_n] = [v_1\ v_2\ \ldots\ v_n]\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

- **Define $T = [v_1\ v_2\ \ldots\ v_n]$ and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, so that**

  $$AT = T\Lambda$$

- **Thus,**

  $$T^{-1}AT = \Lambda$$

  T^{-1} exists since $v_1, v_2, \ldots, v_n$ are linearly independent.

- **T is a similarity transformation that diagonalizes A.**

Eigenvectors and Eigenvalues

- Converse. If there exist a vector $T = [v_1 \; v_2 \; \ldots \; v_n]$ such that

$$T^{-1}AT = \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

then $AT = T\Lambda$, i.e.,

$$Av_i = \lambda_i v_i, \qquad i = 1, \ldots, n$$

such that $v_1, v_2, \ldots, v_n$ are linearly independent eigenvectors of A.

- Briefly, A is diagonalizable if
  - there exists T such that $T^{-1}AT = \Lambda$ is diagonal.
  - A has a set of linearly independent eigenvectors.

Eigenvectors and Eigenvalues

- We can also use the left eigenvectors for diagonalization. Rewrite $T^{-1}AT = \Lambda$ as $T^{-1}A = \Lambda T^{-1}$.

Assigning, $w_1^T, w_2^T, \ldots, w_n^T$ as the rows of T^{-1} we get

$$\begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_n^T \end{bmatrix} A = \Lambda \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \\ w_n^T \end{bmatrix}$$

- Thus, $w_i^T A = \Lambda_i w_i^T$.

The rows of T^{-1} are left eigenvectors. They are also normalized so that

$$w_i^T v_j = \delta_{ij}$$

Eigenvectors and Eigenvalues

- Definition. If A is not diagonalizable, it is called defective.

- Take for example

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

The characteristic polynomial is $\mathcal{X}(s) = s^2$.

$\Rightarrow \lambda = 0$ is the only eigenvalue.

Eigenvectors satisfy $Av = 0v = 0$.

Eigenvectors and Eigenvalues

- Thus, $Av = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = 0$

and the eigenvectors have the form

$$v = \begin{bmatrix} v_1 \\ 0 \end{bmatrix} \qquad \text{where } v_1 \neq 0$$

$\Rightarrow$ A does not have two linearly independent eigenvectors.
$\Rightarrow$ A is not diagonalizable.

- If A has distinct eigenvalues, i.e., $\lambda_i \neq \lambda_j$ for $i \neq j$, then A is diagonalizable.

- The diagonal form is important in investigating how a control system behaves.

  What if A cannot be diagonalized?

  Alternatively, we can use the Jordan canonical form.

- Any matrix $A \in R^{n\times n}$ can be put in the Jordan canonical form by the a similarity transformation.

$$T^{-1}AT = \begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_q \end{bmatrix}$$

  where J_i is a Jordan block of size n_i with eigenvalue λ_i.

- A Jordan block J_i of size n_i with eigenvalue λ_i.

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

- Properties of the Jordan form.
  - J is upper diagonal.
  - J is the special case of N Jordan blocks of size $n_i = 1$.
  - Jordan form is unique (up to the permutations of the blocks).
  - we can have multiple blocks with the same eigenvalue.

- The Jordan form is useful for investigating system concepts.

  However, it is never used for numerical computations.

- How do we determine the Jordan canonical form.

  Given

$$\mathcal{X}(\lambda) = \det(\lambda I - A) = (s - \lambda_1)^{n_1} \ldots (s - \lambda_q)^{n_q}$$

  with distinct eigenvalues.

  $\Rightarrow n_i = 1 \Rightarrow A$ is diagonalizable.

- The dimension of $\mathcal{N}(\lambda I - A)$ is the number of Jordan blocks associated with eigenvalue λ.

- In general,

$$\dim \mathcal{N}(\lambda I - A)^k = \sum_{\lambda_i = \lambda} \min\{k, n_i\}$$

  Using $k = 1, 2, \ldots$ and $\dim \mathcal{N}(\lambda I - A)^k$, we can deduce the sizes of the Jordan blocks associated with λ.

Jordan Canonical Form

- Suppose $T^{-1}AT = J = \text{diag}(J_1, \ldots, J_q)$.

  Express T as
  $$T = [T_1 \; T_2 \; \ldots \; T_q]$$
  where $T_i \in C_{n \times n_i}$ are the columns of T associated with ith Jordan block J_i.

- Expanding
  $$A[T_1 \; T_2 \; \ldots \; T_q] = [T_1 \; T_2 \; \ldots \; T_q] \cdot \text{diag}(J_1, \ldots, J_q)$$
  we can write $AT_i = T_i J_i$.

Jordan Canonical Form

- Let $T_i = [v_{i1} \; v_{i2} \; \ldots \; v_{in_i}]$, then we have
  $$Av_{i1} = \lambda_i v_{i1}$$
  i.e., the first column of each T_i is an eigenvector associated with eigenvalue λ_i.

- For $j = 1, \ldots, n_i - 1$,
  $$Av_{ij} = v_{i,j-1} + \lambda_i v_{ij}$$
  Vectors $v_{i1}, \ldots, v_{in_i}$ are called generalized eigenvectors.

Cayley-Hamilton Theorem

- Useful for deriving solutions to linear equations.

- If $p(s) = a_0 + a_1 s + \ldots + a_k s^k$ is a polynomial and $A \in R^{n \times n}$, we define
  $$p(A) = a_0 I + a_1 A + \ldots + a_k A^k$$

- Cayley-Hamilton theorem.

  For any $A \in R^{n \times n}$, we have $\mathcal{X}(A) = 0$, where $\mathcal{X}(s) = \det(sI - A)$.

Cayley-Hamilton Theorem

- Example.

  For $A = \begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix}$, we have
  $$\mathcal{X}(s) = \det(sI - A) = s^2 - 3s - 10$$
  Thus,
  $$\begin{aligned} \mathcal{X}(A) &= A^2 - 3A - 10I \\ &= \begin{bmatrix} 13 & 9 \\ 12 & 16 \end{bmatrix} - 3\begin{bmatrix} 1 & 3 \\ 4 & 2 \end{bmatrix} - 10I \\ &= 0 \end{aligned}$$

Cayley-Hamilton Theorem

- **Corollary.** For every $p \in Z_+$, we have

$$A^p \in \text{span}\{I, A, A^2, \ldots, A^{n-1}\}$$

Additionally, if A is nonsingular, the above holds for $p \in Z$.

- Every power of A can be expressed as a linear combination of $I, A, A^2, \ldots, A^{n-1}$.

Proof. Divide s^p by $\mathcal{X}(s)$ to get

$$\frac{s^p}{\mathcal{X}(s)} = q(s) + \frac{r(s)}{\mathcal{X}(s)} \Rightarrow s^p = q(s)\mathcal{X}(s) + r(s)$$

Cayley-Hamilton Theorem

- Where $q(s)$ is the quotient polynomial and $r(s) = \alpha_0 + \alpha_1 s + \ldots + \alpha_{n-1}s^{n-1}$ is the remainder polynomial.

- Then,

$$\begin{aligned} A^p &= q(A)\mathcal{X}(A) + r(A) \\ &= r(A) \\ &= \alpha_0 I + \alpha_1 A + \ldots + \alpha_{n-1}A^{n-1} \end{aligned}$$

- You could also show this by taking $\mathcal{X}(A) = 0$ and expressing higher powers of A in terms of lower powers of A.

Cayley-Hamilton Theorem

- For $p = -1$, rewrite Cayley-Hamilton theorem

$$\mathcal{X}(A) = A^n + a_{n-1}A^{n-1} + \ldots + a_0 I = 0$$

as

$$I = A\left[-\frac{a_1}{a_0} - \frac{a_2}{a_0}A - \ldots - \frac{1}{a_0}A^{n-1}\right]$$

- Note : since A is nonsigular, $a_0 \neq 0$. Thus,

$$A^{-1} = -\frac{a_1}{a_0} - \frac{a_2}{a_0}A - \ldots - \frac{1}{a_0}A^{n-1}$$

i.e., the inverse of A may be expressed as a linear combination of $I, A, A^2, \ldots, A^{n-1}$.

Proof of Cayley-Hamilton Theorem

- First assume A is diagonalizable, i.e., $T^{-1}AT = \Lambda$.

$$\mathcal{X}(s) = (s - \lambda_1)(s - \lambda_2)\ldots(s - \lambda_n)$$

- Since $\mathcal{X}(A) = \mathcal{X}(T\Lambda T^{-1}) = T\mathcal{X}(\Lambda)T^{-1}$, it is sufficient to show $\mathcal{X}(\Lambda) = 0$.

$$\begin{aligned} \mathcal{X}(\Lambda) &= (\Lambda - \lambda_1 I)(\Lambda - \lambda_2 I)\ldots(\Lambda - \lambda_n I) \\ &= \text{diag}(0, \lambda_2 - \lambda_1, \ldots, \lambda_n - \lambda_1) \\ &\quad \ldots\text{diag}(\lambda_1 - \lambda_n, \ldots, \lambda_{n-1} - \lambda_n, 0) \\ &= 0 \end{aligned}$$

Proof of Cayley-Hamilton Theorem

- For the general case, use the Jordan form.
  With $T^{-1}AT = J$, we get

$$\mathcal{X}(s) = (s - \lambda_1)^{n_1}(s - \lambda_2)^{n_2}\dots(s - \lambda_n)^{n_q}$$

- It is sufficient to show $\mathcal{X}(J_i) = 0$.

$$\mathcal{X}(J_i) = (J_i - \lambda_1 I)^{n_1}\dots \underbrace{\begin{bmatrix} 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ & & & \dots \end{bmatrix}}_{(J_i - \lambda_i I)^{n_i}} \cdots (J_i - \lambda_q I)^{n_q}$$

$$= 0$$

Summary

- Linear algebra is the main tool for dealing with linear systems.
- Definitions.
  - vector space and subspace
  - Independence, basis and dimension
  - Range, nullspace, inverse and rank
  - Similarity transform, norms and inner product
- Important concepts.
  - Eigenvectors and eigenvalues
  - Jordan canonical form and Cayley-Hamilton theorem