

Linear Functions

• System of linear equations

- Consider a function or mapping, $\mathbf{y} = \mathbf{Ax}$, i.e.,

$$y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n$$

$$\vdots$$

$$y_m = a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n$$

$$\text{where } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \in \mathbf{R}^m, \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \in \mathbf{R}^{m \times n}, \text{ and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbf{R}^n$$

(1) \mathbf{y} is measurement or observation; \mathbf{x} is unknown to be determined

(2) \mathbf{x} is input; \mathbf{y} is output

• Linear functions

- A function or mapping $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ is linear if

$$(1) f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}), \forall \mathbf{x}, \mathbf{y} \in \mathbf{R}^n$$

$$(2) f(\alpha \mathbf{x}) = \alpha f(\mathbf{x}), \forall \mathbf{x} \in \mathbf{R}^n \forall \alpha \in \mathbf{R}$$

- Any linear function $f: \mathbf{R}^n \rightarrow \mathbf{R}^m$ can be written as $f(\mathbf{x}) = \mathbf{Ax}$ for some $\mathbf{A} \in \mathbf{R}^{m \times n}$

• Interpretation of $y_i = \sum_{j=1}^n a_{ij}x_j$

(1) a_{ij} is a gain from j th input (x_j) to i th output (y_i)

(2) i th row of $\mathbf{A} \Leftrightarrow i$ th output

(3) j th column of $\mathbf{A} \Leftrightarrow j$ th input

(4) $a_{ij} = 0 \Leftrightarrow i$ th output (y_i) does not depend on j th input (x_j)

(5) a_{ip} dominates all a_{ij} for $j \neq p \Leftrightarrow y_i$ depends mainly on x_p

(6) a_{qi} dominates all a_{ij} for $i \neq q \Leftrightarrow x_q$ affects mainly y_i

(7) \mathbf{A} is diagonal, i.e., $a_{ij} = 0$ for $i \neq j \Leftrightarrow y_i$ depends only on x_i

(8) Sparsity pattern of \mathbf{A} determines input-output interactions

• Interpretation of $\mathbf{y} = \mathbf{A}\mathbf{x}$

- Sum (linear combination) of columns. Write \mathbf{A} as $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \cdots \mathbf{a}_n]$ where

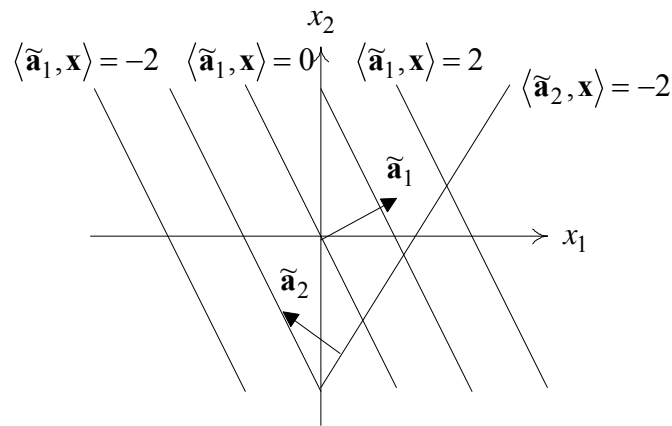
$$\mathbf{a}_j = [a_{1j} \ a_{2j} \ \cdots \ a_{mj}]^T \in \mathbf{R}^m. \text{ Then,}$$

$$\mathbf{y} = x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \sum_{j=1}^n x_j \mathbf{a}_j$$

- Inner product with rows. Write \mathbf{A} as

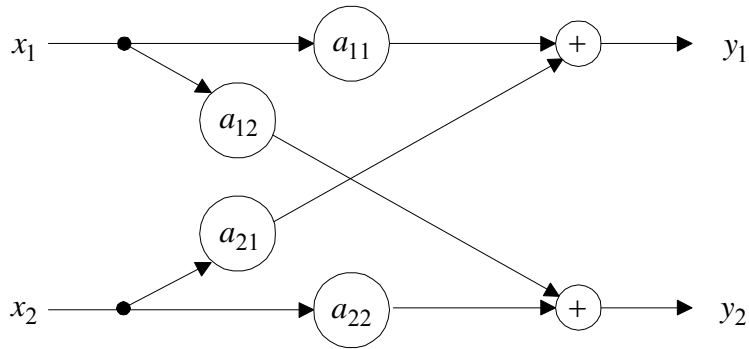
$$\mathbf{A} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \tilde{\mathbf{a}}_2^T \\ \vdots \\ \tilde{\mathbf{a}}_m^T \end{bmatrix} \text{ where } \tilde{\mathbf{a}}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \in \mathbf{R}^n. \text{ Then } \mathbf{y} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{x} \\ \tilde{\mathbf{a}}_2^T \mathbf{x} \\ \vdots \\ \tilde{\mathbf{a}}_m^T \mathbf{x} \end{bmatrix} \text{ or } y_i = \langle \tilde{\mathbf{a}}_i, \mathbf{x} \rangle.$$

- \mathbf{x} is on intersection of all hyperplanes $\tilde{\mathbf{a}}_i^T \mathbf{x} = y_i$ for $i = 1, 2, \dots, m$.



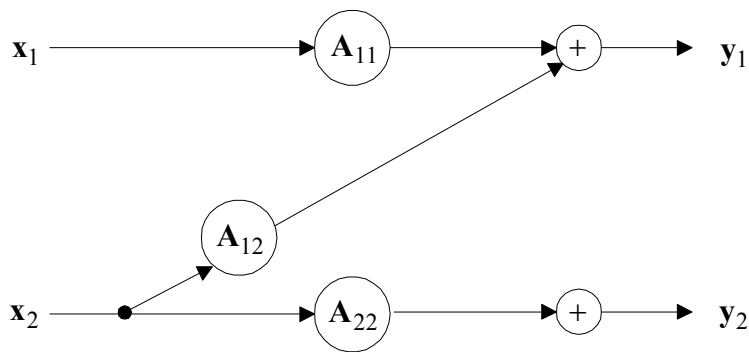
- Block diagram or signal flow graph

(1) For a system,
$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$



(2) For a system, $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, we can partition as $y_1 = \mathbf{A}_{11}x_1 + \mathbf{A}_{12}x_2$

and $y_2 = \mathbf{A}_{22}x_2$

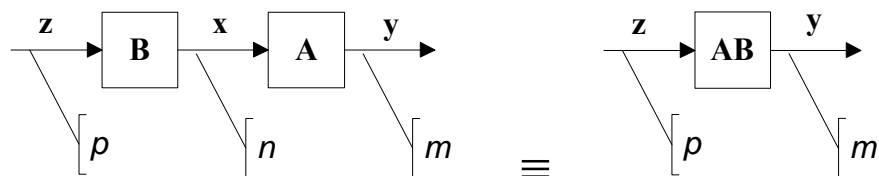


• Composite functions or systems $y = \mathbf{AB}z$

- Matrix multiplication, $\mathbf{C} = \mathbf{AB} \in \mathbf{R}^{m \times p}$ where $\mathbf{A} \in \mathbf{R}^{m \times n}$ and $\mathbf{B} \in \mathbf{R}^{n \times p}$. Then,

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj} = \tilde{\mathbf{a}}_i^T \mathbf{b}_j = \langle \tilde{\mathbf{a}}_i, \mathbf{b}_j \rangle, \quad \mathbf{C} = \mathbf{AB} = [\mathbf{A}\mathbf{b}_1 \cdots \mathbf{A}\mathbf{b}_p], \quad \text{and} \quad \mathbf{C} = \mathbf{AB} = \begin{bmatrix} \tilde{\mathbf{a}}_1^T \mathbf{B} \\ \vdots \\ \tilde{\mathbf{a}}_m^T \mathbf{B} \end{bmatrix}$$

- Composite interpretation.



• Many engineering problems can be modelled as $y = \mathbf{Ax}$

- Estimation or inversion problems

(1) Model

- (a) y_i is i th measurement or sensor reading
- (b) x_j is j th parameter to be estimated
- (c) a_{ij} is sensitivity of i th sensor to j th parameter

(2) Problems

- (a) Given \mathbf{y} , find \mathbf{x}
- (b) Find all \mathbf{x} 's that can result in \mathbf{y}
- (c) If there is no \mathbf{x} such that $\mathbf{y} = \mathbf{Ax}$, find the best $\hat{\mathbf{x}}$ such that $\mathbf{y} \approx \mathbf{A}\hat{\mathbf{x}} \Rightarrow$ least square error solution and minimum norm solution

- Control or design problems

(1) Model

- (a) \mathbf{x} is design parameter or inputs or controls
- (b) \mathbf{y} is results or outputs or states
- (c) \mathbf{A} describes how input affects output

(2) Problems

- (a) Given a desirable \mathbf{y}^* (specifications), find \mathbf{x} so that $\mathbf{y} = \mathbf{y}^*$
- (b) Find all \mathbf{x} 's that can result in $\mathbf{y} = \mathbf{y}^*$
- (c) Among all \mathbf{x} 's in (b), find a small or efficient one

- Signal processing problems

(1) Model

- (a) y_i is output signal at time i
- (b) x_j is input signal at time j

(2) Problems

- (a) Given a desirable \mathbf{y}^* , find \mathbf{A} so that $\mathbf{Ax} = \mathbf{y}^*$
- (b) Given \mathbf{A} , find \mathbf{B} such that $\mathbf{By} = \mathbf{x}$ or $\mathbf{By} \approx \mathbf{x}$

Linear Algebra

• Vector space and subspace (see Vector Space)

- $V_1 = \mathbf{R}^n$, subspace of \mathbf{R}^n
- $V_2 = \{\mathbf{0}\}$ where $\mathbf{0} \in \mathbf{R}^n$, subspace of \mathbf{R}^n
- $V_3 = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ where $\mathbf{v}_i \in \mathbf{R}^n$, subspace of \mathbf{R}^n
- $V_4 = \{\mathbf{x} : \mathbf{R}_+ \rightarrow \mathbf{R}^n \mid \mathbf{x} \text{ is differentiable}\}$ where $(\mathbf{x} + \mathbf{z})(t) = \mathbf{x}(t) + \mathbf{z}(t)$,
 $(\alpha \mathbf{x})(t) = \alpha \mathbf{x}(t)$ (a point in V_4 is a trajectory in \mathbf{R}^n)
- $V_5 = \{\mathbf{x} \in V_4 : \dot{\mathbf{x}} = \mathbf{A}\mathbf{x}\}$, subspace of V_4

• Linear independence, basis, dimension (see Vector Space)

- Linearly independent vs. linearly dependent
- $V = \text{span}\{\mathbf{v}_i\}_{i=1}^n$ and $\{\mathbf{v}_i\}_{i=1}^n$ are linearly independent
 - (a) $\{\mathbf{v}_i\}_{i=1}^n$ is a basis and dimension of V is n .
 - (b) $\forall \mathbf{x} \in V, \mathbf{x} = \sum_{i=1}^n c_i \mathbf{v}_i, \{c_i\}_{i=1}^n$ is unique
 - (c) Basis is not unique, i.e., V may have infinite number of bases
 - (d) All bases have the same number of vectors (equal to dimension)
 - (e) Any linearly independent set of vectors in V can be extended to a basis
 - (f) Any spanning set of V can be reduced to a basis

• Real matrix and vector

- $\mathbf{R}^{m \times n}$ denotes the vector space of all $m \times n$ real matrices:

$$\mathbf{A} \in \mathbf{R}^{m \times n} \Leftrightarrow \mathbf{A} = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}.$$

- \mathbf{R}^m (i.e., $\mathbf{R}^{m \times 1}$) denotes the vector space of all $m \times 1$ real column vectors.

- For $\mathbf{x} \in \mathbf{R}^m$ and $\mathbf{y} \in \mathbf{R}^n$, $\mathbf{xy}^T \in \mathbf{R}^{m \times n}$ is the outer product.
- For $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{y} \in \mathbf{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} \in \mathbf{R}^{1 \times 1}$ is the inner product.

• Complex matrices

- For $\mathbf{A} \in \mathbf{C}^{m \times n}$, then $\mathbf{A}^H = \left[\overline{a_{ji}} \right]$.
- For $\mathbf{x} \in \mathbf{C}^n$ and $\mathbf{y} \in \mathbf{C}^n$, $\mathbf{x}^H \mathbf{y} = \sum_{i=1}^n \overline{x_i} y_i = \overline{\mathbf{y}^H \mathbf{x}}$ is the inner product.
- $\mathbf{A} \in \mathbf{C}^{n \times n}$ is unitary if $\mathbf{A}^H \mathbf{A} = \mathbf{I}_n$, Hermitian if $\mathbf{A}^H = \mathbf{A}$, and positive definite if $\mathbf{x}^H \mathbf{A} \mathbf{x} > 0, \forall \mathbf{x} \in \mathbf{C}^n$ and $\mathbf{x} \neq \mathbf{0}$.

• Vector norm (Euclidean norm, see normed vector space)

- For $\mathbf{x} \in \mathbf{R}^n$, $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{\mathbf{x}^T \mathbf{x}}$
- $\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$

• Inner product (see Inner Product Space)

- For $\mathbf{x}, \mathbf{y} \in \mathbf{C}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^H \mathbf{y} = x_1^* y_1 + \dots + x_n^* y_n = \sum_{i=1}^n x_i^* y_i$
- For $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = x_1 y_1 + \dots + x_n y_n = \sum_{i=1}^n x_i y_i$
- Row vector \mathbf{x}^T represents a linear functional: $\mathbf{R}^n \rightarrow \mathbf{R}$
- $\{\mathbf{x} : \mathbf{x}^T \mathbf{y} \leq 0\}$ defines a halfspace with boundary passing $\mathbf{0}$ and outward normal vector \mathbf{y}

• Null space of a matrix

- For a matrix $\mathbf{A} \in \mathbf{R}^{m \times n}$, the null space of \mathbf{A} is

$$N(\mathbf{A}) = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{A} \mathbf{x} = \mathbf{0}\}.$$

(1) $N(\mathbf{A})$ is set of vectors mapped to zero by $\mathbf{y} = \mathbf{A} \mathbf{x}$

(2) $N(\mathbf{A})$ is set of vectors orthogonal to all rows of \mathbf{A} ; $y_i = \langle \tilde{\mathbf{a}}_i, \mathbf{x} \rangle = 0$ for all i

- If $N(\mathbf{A}) = \{\mathbf{0}\}$,

(1) $\mathbf{y} = \mathbf{A}\mathbf{x}$ uniquely determines \mathbf{x} (the linear transformation does not lose information, i.e., can be inverted)

(2) Since $\mathbf{y} = \sum_{i=1}^n x_i \mathbf{a}_i$ and $\{x_i\}_{i=1}^n$ is unique, $\{\mathbf{a}_i\}_{i=1}^n$ is a basis

(3) \mathbf{A} has a left inverse, i.e., $\exists \mathbf{B} \in \mathbf{R}^{n \times m}$ s.t. $\mathbf{BA} = \mathbf{I}_n$

(4) $\det(\mathbf{A}^T \mathbf{A}) \neq 0$

- Meanings of $\mathbf{z} \in N(\mathbf{A})$

(1) Ambiguity in \mathbf{x} . Given $\mathbf{y} = \mathbf{A}\mathbf{x}$ and $\mathbf{z} \in N(\mathbf{A})$,

(a) $\mathbf{y} = \mathbf{A}(\mathbf{x} + \mathbf{z}) \Rightarrow \mathbf{z}$ is undetectable

(b) If $\mathbf{y} = \mathbf{A}\hat{\mathbf{x}}$, then $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{z} \Rightarrow \mathbf{x}$ and $\hat{\mathbf{x}}$ are indistinguishable

(2) Freedom of input choice. Given $\mathbf{y} = \mathbf{A}\mathbf{x}$ and $\mathbf{z} \in N(\mathbf{A})$

(a) $\mathbf{0} = \mathbf{A}\mathbf{z} \Rightarrow \mathbf{z}$ is input with no result

(b) $\hat{\mathbf{x}} = \mathbf{x} + \mathbf{z}$ provides different input choice for the same result

• Range of a matrix

- For a matrix $\mathbf{A} \in \mathbf{R}^{m \times n}$, the range of \mathbf{A} is

$$R(\mathbf{A}) = \{\mathbf{y} \in \mathbf{R}^m : \mathbf{y} = \mathbf{A}\mathbf{x} \text{ for some } \mathbf{x} \in \mathbf{R}^n\} = \text{span}\{\mathbf{a}_i\}_{i=1}^n$$

- \mathbf{A} is called onto if $R(\mathbf{A}) = \mathbf{R}^m$

(1) $\text{span}\{\mathbf{a}_i\}_{i=1}^n = \mathbf{R}^m$ and $n \geq m$

(2) $\mathbf{y} = \mathbf{A}\mathbf{x}$ can be solved for \mathbf{x}

(3) \mathbf{A} has a right inverse, i.e., $\exists \mathbf{B} \in \mathbf{R}^{n \times m}$ s.t. $\mathbf{AB} = \mathbf{I}_m$

(4) $\{\tilde{\mathbf{a}}_j\}_{j=1}^m$ are linearly independent

(5) $N(\mathbf{A}^T) = \{\mathbf{0}\}$

(6) $\det(\mathbf{AA}^T) \neq 0$

- Meanings of $\mathbf{v} \in R(\mathbf{A})$ and $\mathbf{w} \notin R(\mathbf{A})$ given $\mathbf{y} = \mathbf{A}\mathbf{x}$

(1) \mathbf{y} is a measurement of \mathbf{x} , $R(\mathbf{A})$ is the possible results

(a) $\mathbf{y} = \mathbf{v}$ is a possible or consistent sensor signal

(b) $\mathbf{y} = \mathbf{w}$ is an impossible or inconsistent sensor signal (sensor failure or wrong

model)

(2) \mathbf{y} is an output for input \mathbf{x} , $R(\mathbf{A})$ is the achievable outputs

(a) \mathbf{v} is a possible output

(b) \mathbf{w} cannot be an output

• **Rank of a matrix**

- For any matrix $\mathbf{A} \in \mathbf{R}^{m \times n}$, $\text{rank}(\mathbf{A}) = \dim[R(\mathbf{A})]$

- $\text{rank}(\mathbf{A})$ is the maximal number of independent columns or rows \Rightarrow

$$\text{rank}(\mathbf{A}) \leq \min(m, n)$$

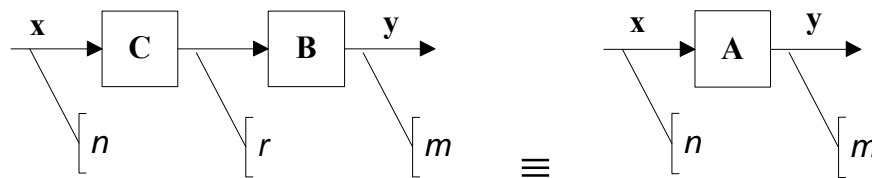
- $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T)$

- $\text{rank}(\mathbf{A}) + \dim[N(\mathbf{A})] = n$ or $\text{dim}[R(\mathbf{A})] + \dim[N(\mathbf{A})] = n \Rightarrow$ each dimension of input \mathbf{x} is either crushed to zero or ends up in nonzero output

- Rank of product; $\text{rank}(\mathbf{BC}) \leq \min\{\text{rank}(\mathbf{B}), \text{rank}(\mathbf{C})\}$

(1) If $\mathbf{A} = \mathbf{BC}$ with $\mathbf{A} \in \mathbf{R}^{m \times n}$, $\mathbf{B} \in \mathbf{R}^{m \times r}$, and $\mathbf{C} \in \mathbf{R}^{r \times n}$, then $\text{rank}(\mathbf{A}) \leq r$

(2) If $\text{rank}(\mathbf{A}) = r$, then $\mathbf{A} = \mathbf{BC}$ with $\mathbf{A} \in \mathbf{R}^{m \times n}$, $\mathbf{B} \in \mathbf{R}^{m \times r}$, and $\mathbf{C} \in \mathbf{R}^{r \times n}$



(a) $\text{rank}(\mathbf{A}) = r$ is the minimal size of vector needed to faithfully reconstruct \mathbf{y} from \mathbf{x}

(b) Two step computation, i.e., $\mathbf{z} = \mathbf{C}\mathbf{x}$ and $\mathbf{y} = \mathbf{B}\mathbf{z}$ needs $(m + n)r$ operations compared to mn operations in $\mathbf{y} = \mathbf{A}\mathbf{x}$

- \mathbf{A} is full rank if $\text{rank}(\mathbf{A}) = \min\{m, n\}$

(1) If $m = n$, full rank means nonsingular

(2) If $m > n$ (narrow), full rank means columns are independent

(3) If $m < n$ (wide), full rank means rows are independent

• **Inverse of a matrix**

- A matrix $\mathbf{A} \in \mathbf{R}^{n \times n}$ is invertible or nonsingular if $\det(\mathbf{A}) \neq 0$. Followings are equivalent to $\det(\mathbf{A}) \neq 0$.

- (1) $\{\mathbf{a}_i\}_{i=1}^n$ are a basis for \mathbf{R}^n
- (2) $\{\tilde{\mathbf{a}}_j\}_{j=1}^n$ are a basis for \mathbf{R}^n
- (3) $\mathbf{y} = \mathbf{A}\mathbf{x}$ has a unique solution \mathbf{x} for every \mathbf{y}
- (4) $\exists \mathbf{A}^{-1}$ s.t. $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$
- (5) $N(\mathbf{A}) = \{\mathbf{0}\}$
- (6) $R(\mathbf{A}) = \mathbf{R}^n$
- (7) $\det(\mathbf{A}^T \mathbf{A}) = \det(\mathbf{A}\mathbf{A}^T) \neq 0$
- (8) $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^T) = n$

• Coordinates

- Standard basis vectors in \mathbf{R}^n are $\{\mathbf{e}_i\}_{i=1}^n$ where $\mathbf{e}_i = [00 \cdots 010 \cdots 0]^T$
- $\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$ and $\{x_i\}_{i=1}^n$ are coordinates of \mathbf{x} in the standard basis
- Change of coordinates

(1) If $\{\mathbf{t}_i\}_{i=1}^n$ are another basis for \mathbf{R}^n , $\mathbf{x} = \tilde{x}_1\mathbf{t}_1 + \cdots + \tilde{x}_n\mathbf{t}_n$

(2) Define $\mathbf{T} = [\mathbf{t}_1 \mathbf{t}_2 \cdots \mathbf{t}_n]$, then $\mathbf{x} = \mathbf{T}\tilde{\mathbf{x}}$ with $\tilde{\mathbf{x}} = [\tilde{x}_1 \cdots \tilde{x}_n]^T$

(3) $\tilde{\mathbf{x}} = \mathbf{T}^{-1}\mathbf{x}$

- Consider a linear transform $\mathbf{y} = \mathbf{A}\mathbf{x}$ for $\mathbf{A} \in \mathbf{R}^{n \times n}$,

$$\mathbf{x} = \mathbf{T}\tilde{\mathbf{x}}, \quad \mathbf{y} = \mathbf{T}\tilde{\mathbf{y}}, \quad \text{and} \quad \tilde{\mathbf{y}} = (\mathbf{T}^{-1}\mathbf{A}\mathbf{T})\tilde{\mathbf{x}}$$

(1) $\mathbf{A} \rightarrow \mathbf{T}^{-1}\mathbf{A}\mathbf{T}$: similarity transformation

(2) Similarity transformation by \mathbf{T} expresses \mathbf{y} in coordinates $\mathbf{T} = [\mathbf{t}_1 \mathbf{t}_2 \cdots \mathbf{t}_n]$

• Orthogonal vectors

- The set of vectors $\{\mathbf{u}_i\}_{i=1}^k$ in \mathbf{R}^n are orthogonal if $\mathbf{u}_i^T \mathbf{u}_j = 0$ whenever $i \neq j$ and orthonormal if $\mathbf{u}_i^T \mathbf{u}_j = \delta_{ij}$.

- Let $\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_k]$, then $\mathbf{U}^T \mathbf{U} = \mathbf{I}_k$ and $\text{span}\{\mathbf{u}_i\}_{i=1}^k = R(\mathbf{U})$

- Geometric properties

(1) If $\mathbf{w} = \mathbf{Uz}$, then $\|\mathbf{w}\| = \|\mathbf{Uz}\| = \|\mathbf{z}\|$. Mapping \mathbf{U} is *isometric*.

(2) $\langle \mathbf{Uz}, \mathbf{Uw} \rangle = \langle \mathbf{z}, \mathbf{w} \rangle$ and hence $\angle(\mathbf{Uz}, \mathbf{Uw}) = \angle(\mathbf{z}, \mathbf{w})$

(3) If $\mathbf{w} = \mathbf{Uz}$ and $\tilde{\mathbf{w}} = \mathbf{U}\tilde{\mathbf{z}}$, $\langle \mathbf{w}, \tilde{\mathbf{w}} \rangle = \langle \mathbf{z}, \tilde{\mathbf{z}} \rangle$

(4) If $k = n$, the mapping \mathbf{U} is either *rotation* or *reflection*.

• Orthonormal basis and expansion

- The set of vectors $\{\mathbf{u}_i\}_{i=1}^n$ in \mathbf{R}^n is an *orthonormal basis* if $\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_n]$ is square

and orthogonal, i.e., $\mathbf{U}^T \mathbf{U} = \mathbf{I}_n$.

- Then, $\mathbf{U}^{-1} = \mathbf{U}^T$, $\mathbf{U}\mathbf{U}^T = \mathbf{I}_n$, and $\sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i^T = \mathbf{I}_n$

- $\mathbf{x} = \mathbf{U}\mathbf{U}^T \mathbf{x} = \sum_{i=1}^n (\mathbf{u}_i^T \mathbf{x}) \mathbf{u}_i = \sum_{i=1}^n \langle \mathbf{u}_i, \mathbf{x} \rangle \mathbf{u}_i$

(1) $\langle \mathbf{u}_i, \mathbf{x} \rangle = \mathbf{u}_i^T \mathbf{x}$ is the component (or projection) of \mathbf{x} in the direction of \mathbf{u}_i

(2) $\mathbf{a} = \mathbf{U}^T \mathbf{x} = [\mathbf{u}_1^T \mathbf{x} \ \mathbf{u}_2^T \mathbf{x} \ \cdots \ \mathbf{u}_n^T \mathbf{x}]^T$ resolves \mathbf{x} into \mathbf{u}_i components

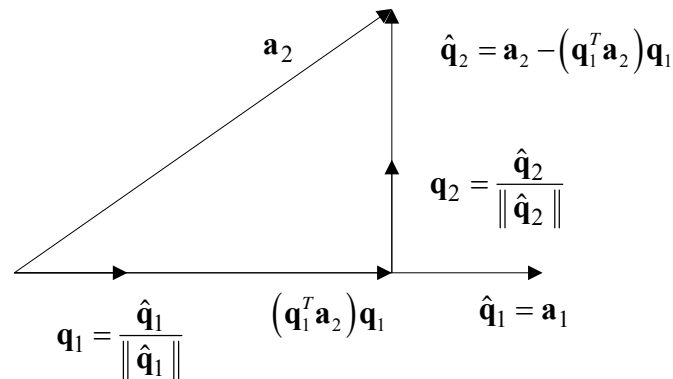
(3) $\mathbf{x} = \mathbf{U}\mathbf{a}$ reconstructs \mathbf{x} from its \mathbf{u}_i components

(4) $\mathbf{x} = \mathbf{U}\mathbf{U}^T \mathbf{x} = \mathbf{U}\mathbf{a} = \sum_{i=1}^n a_i \mathbf{u}_i$ is the expansion of \mathbf{x} in $\{\mathbf{u}_i\}_{i=1}^n$ basis

• Gram-Schmidt procedure

- Any independent vectors $\{\mathbf{a}_i\}_{i=1}^k$ in \mathbf{R}^n can be transformed to orthonormal vectors

$\{\mathbf{q}_i\}_{i=1}^k$ such that $\text{span}\{\mathbf{a}_i\}_{i=1}^r = \text{span}\{\mathbf{q}_i\}_{i=1}^r$ for $r \leq k$.



- QR decomposition. Let $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \cdots \mathbf{a}_k]$ and $\mathbf{Q} = [\mathbf{q}_1 \mathbf{q}_2 \cdots \mathbf{q}_k]$, then $\mathbf{A} = \mathbf{Q}\mathbf{R}$ with $\mathbf{Q}^T \mathbf{Q} = \mathbf{I}_k$ and \mathbf{R} is upper triangular, invertible.

• Orthogonal complement

- A collection of subspaces S_1, S_2, \dots, S_p of \mathbf{R}^n is mutually orthogonal if $\mathbf{x}^T \mathbf{y} = 0$, $\forall \mathbf{x} \in S_i$ and $\forall \mathbf{y} \in S_j$ for $i \neq j$.

- The orthogonal complement of a subspace $S \subset \mathbf{R}^n$ is

$$S^\perp = \{\mathbf{y} \in \mathbf{R}^n : \mathbf{y}^T \mathbf{x} = 0, \forall \mathbf{x} \in S\}.$$

- Note that $R(\mathbf{A})^\perp = N(\mathbf{A}^T)$.

- If $\{\mathbf{x}_i\}_{i=1}^p$ is an orthonormal basis for a subspace $S \subset \mathbf{R}^n$, then

$$\text{span}\{\mathbf{x}_i\}_{i=1}^p \oplus \text{span}\{\mathbf{x}_i\}_{i=p+1}^n = \mathbf{R}^n \quad \text{and} \quad S^\perp = \text{span}\{\mathbf{x}_i\}_{i=p+1}^n.$$

• Eigenvalues and eigenvectors

- For $\mathbf{A} \in \mathbf{R}^{n \times n}$, $\lambda \in \mathbf{C}$ is an eigenvalue of \mathbf{A} if $\chi(\lambda) = \det(\lambda \mathbf{I} - \mathbf{A}) = 0$.
- For such $\lambda \in \mathbf{C}$, \exists eigenvector, $\mathbf{v} \in \mathbf{C}^n$, $\mathbf{v} \neq \mathbf{0}$ such that $\mathbf{A}\mathbf{v} = \lambda \mathbf{v}$
- For such $\lambda \in \mathbf{C}$, \exists left eigenvector, $\mathbf{w} \in \mathbf{C}^n$, $\mathbf{w} \neq \mathbf{0}$ such that $\mathbf{w}^T \mathbf{A} = \lambda \mathbf{w}^T$
- Conjugate symmetry: $\mathbf{A}\mathbf{v} = \lambda \mathbf{v} \Leftrightarrow \mathbf{A}\bar{\mathbf{v}} = \bar{\lambda} \bar{\mathbf{v}}$
- Interpretation: scaling by λ

- (1) $\lambda \in \mathbf{R}, \lambda > 0 \Rightarrow \mathbf{v}$ and $\mathbf{A}\mathbf{v}$ are in the same direction
- (2) $\lambda \in \mathbf{R}, \lambda < 0 \Rightarrow \mathbf{v}$ and $\mathbf{A}\mathbf{v}$ are in the opposite direction
- (3) $\lambda \in \mathbf{R}, |\lambda| < 1 \Rightarrow \mathbf{A}\mathbf{v}$ is smaller than \mathbf{v}
- (4) $\lambda \in \mathbf{R}, |\lambda| > 1 \Rightarrow \mathbf{A}\mathbf{v}$ is larger than \mathbf{v}

• Diagonalization

- Suppose $\{\mathbf{v}_i \in \mathbf{R}^{n \times 1}\}_{i=1}^n$ are linearly independent and $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for $\mathbf{A} \in \mathbf{R}^{n \times n}$, then

$$\mathbf{A}\mathbf{V} = \mathbf{V}\mathbf{\Lambda} \quad \text{with } \mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_n] \quad \text{and } \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$$

- Similarity transform by \mathbf{V} diagonalize \mathbf{A} since \mathbf{V} is invertible $\Rightarrow \mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{\Lambda}$
- Conversely, $\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{\Lambda} \Rightarrow \{\mathbf{v}_i\}_{i=1}^n$ are linearly independent and $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$
- \mathbf{A} is *diagonalizable* if such \mathbf{V} exists or \mathbf{A} has linearly independent set of eigenvectors.
- \mathbf{A} is *diagonalizable* if \mathbf{A} has distinct eigenvalues.
- \mathbf{A} is *defective* if it is not diagonalizable.
- Defective matrix can be put in *Jordan canonical form*.

• Symmetric matrix

- $\mathbf{A} \in \mathbf{R}^{n \times n}$ and $\mathbf{A} = \mathbf{A}^T$

(1) Eigenvalues of \mathbf{A} are real.

(2) There is a set of orthonormal eigenvectors of \mathbf{A}

(a) $\exists \{\mathbf{q}_i\}_{i=1}^n$ such that $\mathbf{A}\mathbf{q}_i = \lambda_i \mathbf{q}_i$ and $\mathbf{q}_i^T \mathbf{q}_j = \delta_{ij}$

(b) If $\{\lambda_i\}_{i=1}^n$ are distinct, corresponding eigenvectors are orthogonal. If not, choose eigenvectors so that they are orthogonal.

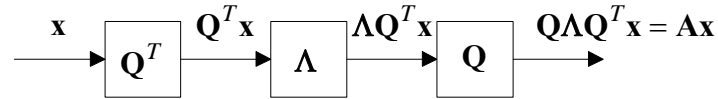
(c) $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$

(d) $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T = \sum_{i=1}^n \lambda_i \mathbf{q}_i \mathbf{q}_i^T$: decomposition of linear combination of one-dimensional orthogonal projections

(3) Interpretations of linear mapping $\mathbf{y} = \mathbf{A}\mathbf{x}$

(a) Resolve \mathbf{x} into \mathbf{q}_i coordinates

- (b) Scale coordinates by λ_i
- (c) Reconstruct with basis \mathbf{q}_i



(4) Geometric interpretation

- (a) Rotate by \mathbf{Q}^T
- (b) Diagonal real scale by Λ
- (c) Rotate back by \mathbf{Q}

• Normal matrix

- $\mathbf{A} \in \mathbf{R}^{n \times n}$ is normal if $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}$
 - (1) Symmetric matrix is normal; $\mathbf{A} = \mathbf{A}^T$
 - (2) Skew symmetric matrix is normal; $\mathbf{A} = -\mathbf{A}^T$
- $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} \Leftrightarrow \exists \mathbf{Q}$ such that $\mathbf{Q}^T\mathbf{A}\mathbf{Q} = \Lambda$, $\mathbf{A} = \mathbf{Q}\Lambda\mathbf{Q}^T = \sum_{i=1}^n \lambda_i \mathbf{q}_i \mathbf{q}_i^T$, $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}_n$,

$$\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$$

• Gram matrix

- For $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_r] \in \mathbf{R}^{n \times r}$, Gram matrix or Grammian is $\mathbf{X}^T\mathbf{X} \in \mathbf{R}^{r \times r}$
- $\{\mathbf{x}_i\}_{i=1}^r$ are linearly independent $\Leftrightarrow \det(\mathbf{X}^T\mathbf{X}) \neq 0$
- Factorization
 - (1) QR factorization: $\mathbf{X} = \mathbf{Q}\mathbf{R}$, $\mathbf{Q} = \mathbf{X}\mathbf{R}^{-1}$, $\mathbf{Q}^T\mathbf{X} = \mathbf{R}$
 - (2) Cholesky factorization: $\mathbf{G} = \mathbf{X}^T\mathbf{X} = \mathbf{R}^T\mathbf{Q}^T\mathbf{Q}\mathbf{R} = \mathbf{R}^T\mathbf{R} (= \mathbf{LU} = \mathbf{LDU})$

• Quadratic forms

- A function $f: \mathbf{R}^n \rightarrow \mathbf{R}$ of the form $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i,j=1}^n a_{ij}x_ix_j$ is a quadratic form.
- Examples: $\|\mathbf{B}\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{B}^T\mathbf{B}\mathbf{x}$, $\sum_{i=2}^n (x_{i+1} - x_i)^2$, $\|\mathbf{F}\mathbf{x}\|^2 - \|\mathbf{G}\mathbf{x}\|^2$
- Uniqueness: if $\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T\mathbf{B}\mathbf{x}$ for all $\mathbf{x} \in \mathbf{R}^n$ and $\mathbf{A} = \mathbf{A}^T, \mathbf{B} = \mathbf{B}^T$, then $\mathbf{A} = \mathbf{B}$.

- $\{\mathbf{x} : f(\mathbf{x}) = a\}$ is a quadratic surface.
- $\{\mathbf{x} : f(\mathbf{x}) \leq a\}$ is a quadratic region.
- If $\mathbf{A} = \mathbf{A}^T$, $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$ with $\lambda_1 \geq \dots \geq \lambda_n$, then
 - (1) $\lambda_n \mathbf{x}^T \mathbf{x} \leq \mathbf{x}^T \mathbf{A} \mathbf{x} \leq \lambda_1 \mathbf{x}^T \mathbf{x}$, $\lambda_n = \lambda_{\min}$ and $\lambda_1 = \lambda_{\max}$
 - (2) $\mathbf{q}_1^T \mathbf{A} \mathbf{q}_1 = \lambda_1 \|\mathbf{q}_1\|^2$ and $\mathbf{q}_n^T \mathbf{A} \mathbf{q}_n = \lambda_n \|\mathbf{q}_n\|^2$
- For $\mathbf{A} = \mathbf{A}^T \in \mathbf{R}^n$,
 - (1) If $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$ for all \mathbf{x} ,
 - (a) \mathbf{A} is positive semidefinite and $\mathbf{A} \geq 0$
 - (b) $\mathbf{A} \geq 0$ iff $\lambda_{\min}(\mathbf{A}) \geq 0$
 - (2) If $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ for all \mathbf{x} ,
 - (a) \mathbf{A} is positive definite and $\mathbf{A} > 0$
 - (b) $\mathbf{A} > 0$ iff $\lambda_{\min}(\mathbf{A}) > 0$
 - (3) \mathbf{A} is negative semidefinite if $-\mathbf{A} \geq 0$
 - (4) \mathbf{A} is negative definite if $-\mathbf{A} > 0$
 - (5) If $\mathbf{B} = \mathbf{B}^T$,
 - (a) $\mathbf{A} \geq \mathbf{B}$ if $\mathbf{A} - \mathbf{B} \geq 0$
 - (b) $\mathbf{A} < \mathbf{B}$ if $\mathbf{A} - \mathbf{B} < 0$
 - (c) $\mathbf{A} > \mathbf{B}$ means $\mathbf{x}^T \mathbf{A} \mathbf{x} > \mathbf{x}^T \mathbf{B} \mathbf{x}$ for all $\mathbf{x} \neq \mathbf{0}$
- Ellipsoids: with $\mathbf{A} = \mathbf{A}^T$, the set $\{\mathbf{x} : \mathbf{x}^T \mathbf{A} \mathbf{x} \leq 1\}$ is an ellipsoid in \mathbf{R}^n centered at $\mathbf{0}$.
 - (1) Semi-axes: $\mathbf{s}_i = \lambda_i^{-1/2} \mathbf{q}_i$
 - (a) Eigenvectors define directions of semiaxes
 - (b) Eigenvalues determine lengths of semiaxes
 - (c) $(\mathbf{q}_1, \lambda_1 = \lambda_{\max})$ direction: smallest length, thin
 - (d) $(\mathbf{q}_n, \lambda_n = \lambda_{\min})$ direction: largest length, fat
 - (2) $\sqrt{\lambda_{\max} / \lambda_{\min}}$ = maximum eccentricity

• **Matrix norms**

- For $\mathbf{A} \in \mathbf{R}^{m \times n}$,

(1) Matrix norm or spectral norm of \mathbf{A} : $\|\mathbf{A}\| = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|}$ = maximum gain or amplification factor

(2) Since $\max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|^2}{\|\mathbf{x}\|^2} = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\mathbf{x}^T \mathbf{A}^T \mathbf{A} \mathbf{x}}{\|\mathbf{x}\|^2} = \lambda_{\max}(\mathbf{A}^T \mathbf{A})$, $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})}$

(3) Similarly, $\min_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|}{\|\mathbf{x}\|} = \sqrt{\lambda_{\min}(\mathbf{A}^T \mathbf{A})}$

(4) Observations

(a) $\mathbf{A}^T \mathbf{A} \in \mathbf{R}^{n \times n}$ is symmetric and $\mathbf{A}^T \mathbf{A} \geq \mathbf{0} \Rightarrow \lambda_{\max}, \lambda_{\min} \geq 0$

(b) $(\mathbf{q}_1, \lambda_{\max} = \lambda_1(\mathbf{A}^T \mathbf{A}))$: maximum gain input direction

(c) $(\mathbf{q}_n, \lambda_{\min} = \lambda_n(\mathbf{A}^T \mathbf{A}))$: minimum gain input direction

- Properties of matrix norm

(1) For $\mathbf{x} \in \mathbf{R}^{n \times 1}$, $\sqrt{\lambda_{\max}(\mathbf{x}^T \mathbf{x})} = \sqrt{\mathbf{x}^T \mathbf{x}} = \|\mathbf{x}\|$

(2) For any \mathbf{x} , $\|\mathbf{Ax}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|$

(3) Scaling: $\|\alpha \mathbf{A}\| \leq |\alpha| \|\mathbf{A}\|$

(4) Triangle inequality: $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$

(5) $\|\mathbf{A}\| = 0 \Leftrightarrow \mathbf{A} = \mathbf{0}$

(6) $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$

- For $\mathbf{A} \in \mathbf{C}^{m \times n}$,

(1) F-norm (*Frobenius norm*) is $\|\mathbf{A}\|_F = \left[\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right]^{1/2}$

$$(2) \text{ } p\text{-norm is } \|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}.$$

• **Matrix inversion formula**

- Inverse of a partitioned matrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \Leftrightarrow \mathbf{R}^{-1} = \begin{bmatrix} \mathbf{E}^{-1} & \mathbf{F}\mathbf{H}^{-1} \\ \mathbf{H}^{-1}\mathbf{G} & \mathbf{H}^{-1} \end{bmatrix}$$

$$\mathbf{E} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$$

$$\mathbf{A}\mathbf{F} = -\mathbf{B}$$

$$\mathbf{G}\mathbf{A} = -\mathbf{C}$$

$$\mathbf{H} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$$

- Matrix inversion lemma

$$\mathbf{E} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \Leftrightarrow \mathbf{E}^{-1} = \mathbf{A}^{-1} + \mathbf{F}\mathbf{H}^{-1}\mathbf{G}$$

- Partitioned matrix inverse

$$\mathbf{R} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \Leftrightarrow \mathbf{R}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \mathbf{F} \\ \mathbf{I} \end{bmatrix} \mathbf{H}^{-1} [\mathbf{G} \quad \mathbf{I}]$$

- Woodbury's identity (rank 1 update)

$$\mathbf{R} = \mathbf{R}_0 + \gamma^2 \mathbf{u}\mathbf{u}^T \Leftrightarrow \mathbf{R}^{-1} = \mathbf{R}_0^{-1} + \frac{\gamma^2}{1 + \gamma^2 \mathbf{u}^T \mathbf{R}_0^{-1} \mathbf{u}} \mathbf{R}_0^{-1} \mathbf{u}\mathbf{u}^T \mathbf{R}_0^{-1}$$

Singular Value Decomposition (SVD)

• SVD of \mathbf{A}

- $\mathbf{A} \in \mathbf{R}^{m \times n}$, $\text{rank}(\mathbf{A}) = r$

- $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$

(1) $\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_r] \in \mathbf{R}^{m \times r}$, $\mathbf{U}^T \mathbf{U} = \mathbf{I}_r$, \mathbf{u}_i are right or input singular vectors of \mathbf{A}

(2) $\mathbf{V} = [\mathbf{v}_1 \cdots \mathbf{v}_r] \in \mathbf{R}^{n \times r}$, $\mathbf{V}^T \mathbf{V} = \mathbf{I}_r$, \mathbf{v}_i are left or output singular vectors of \mathbf{A}

(3) $\mathbf{\Sigma} = \text{diag}(\sigma_1, \dots, \sigma_r)$ with $\sigma_1 \geq \dots \geq \sigma_r > 0$, σ_i are nonzero singular values of

\mathbf{A}

$$\begin{array}{c}
 \begin{array}{ccc}
 & n & \\
 & \boxed{\mathbf{A}} & \\
 m & &
 \end{array}
 = m \begin{array}{c}
 \begin{array}{ccc}
 & r & \\
 & \boxed{\mathbf{U}} & \\
 & &
 \end{array}
 \begin{array}{c}
 \begin{array}{ccc}
 & r & \\
 & \boxed{\mathbf{\Sigma}} & \\
 & &
 \end{array}
 \begin{array}{c}
 \begin{array}{ccc}
 & n & \\
 & \boxed{\mathbf{V}^T} & \\
 & & r
 \end{array}
 \end{array}
 \end{array}$$

$$\begin{array}{c}
 \begin{array}{ccc}
 & n & \\
 & \boxed{\mathbf{A}} & \\
 m & &
 \end{array}
 = m \begin{array}{c}
 \begin{array}{ccc}
 & r & \\
 & \boxed{\mathbf{U}} & \\
 & &
 \end{array}
 \begin{array}{c}
 \begin{array}{ccc}
 & r & \\
 & \boxed{\mathbf{\Sigma}} & \\
 & &
 \end{array}
 \begin{array}{c}
 \begin{array}{ccc}
 & n & \\
 & \boxed{\mathbf{V}^T} & \\
 & & r
 \end{array}
 \end{array}
 \end{array}$$

- $\mathbf{A}^T \mathbf{A} = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T) = \mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T$

(1) \mathbf{v}_i are eigenvectors of $\mathbf{A}^T \mathbf{A}$

(2) $\sigma_i = \sqrt{\lambda_i(\mathbf{A}^T \mathbf{A})}$ and $\lambda_i(\mathbf{A}^T \mathbf{A}) = 0$ for $i > r$

(3) $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})} = \sigma_1$

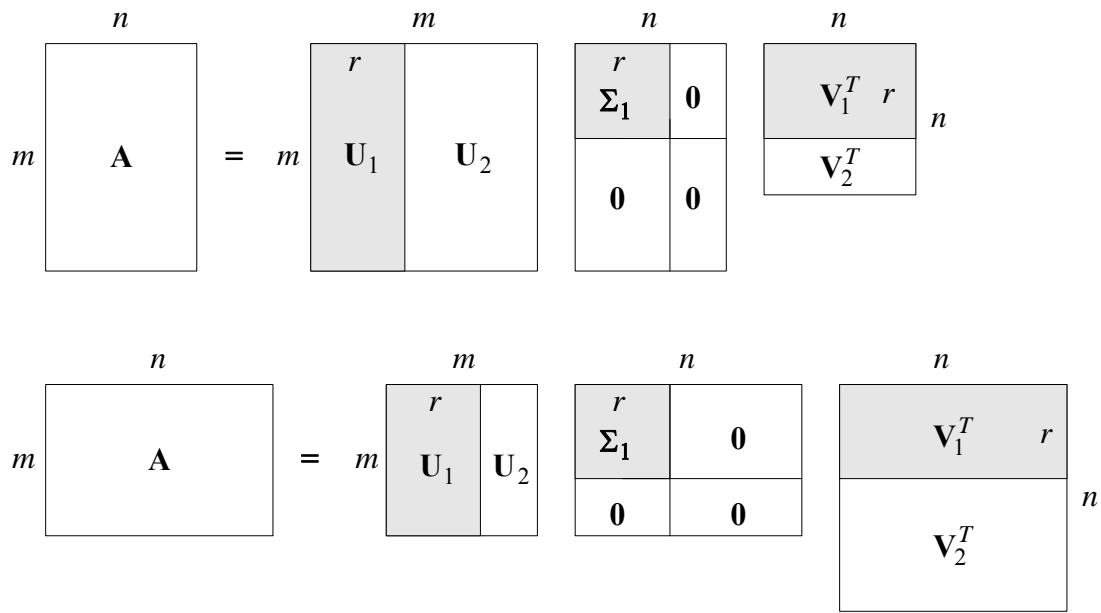
- $\mathbf{A}\mathbf{A}^T = (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T$

- (1) \mathbf{u}_i are eigenvectors of $\mathbf{A}\mathbf{A}^T$
- (2) $\sigma_i = \sqrt{\lambda_i(\mathbf{A}\mathbf{A}^T)}$ and $\lambda_i(\mathbf{A}\mathbf{A}^T) = 0$ for $i > r$
- $\{\mathbf{u}_i\}_{i=1}^r$ are orthonormal basis for $R(\mathbf{A})$
 - $\{\mathbf{v}_i\}_{i=1}^r$ are orthonormal basis for $N(\mathbf{A})^\perp$

• **Full SVD**

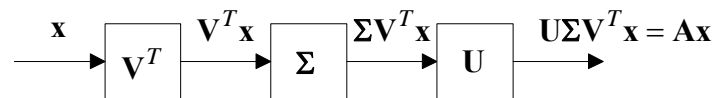
- $\mathbf{A} \in \mathbf{R}^{m \times n}$, $\text{rank}(\mathbf{A}) = r$
 - $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
- (1) $\mathbf{U}_1 = [\mathbf{u}_1 \cdots \mathbf{u}_r] \in \mathbf{R}^{m \times r}$, find $\mathbf{U}_2 = [\mathbf{u}_{r+1} \cdots \mathbf{u}_m] \in \mathbf{R}^{m \times (m-r)}$ such that $\mathbf{U} \in \mathbf{R}^{m \times m}$ and
- $$\mathbf{U}^T \mathbf{U} = \mathbf{I}_m$$
- (2) $\mathbf{V}_1 = [\mathbf{v}_1 \cdots \mathbf{v}_r] \in \mathbf{R}^{n \times r}$, find $\mathbf{V}_2 = [\mathbf{v}_{r+1} \cdots \mathbf{v}_n] \in \mathbf{R}^{n \times (n-r)}$ such that $\mathbf{V} \in \mathbf{R}^{n \times n}$ and
- $$\mathbf{V}^T \mathbf{V} = \mathbf{I}_n$$
- (3) $\mathbf{\Sigma}_1 = \text{diag}(\sigma_1, \dots, \sigma_r)$ with $\sigma_1 \geq \dots \geq \sigma_r > 0$, σ_i are nonzero singular values

of \mathbf{A} , construct $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0}_{r \times (n-r)} \\ \mathbf{0}_{(m-r) \times r} & \mathbf{0}_{(m-r) \times (n-r)} \end{bmatrix} \in \mathbf{R}^{m \times n}$



• Interpretations of SVD

- Decomposition of linear mapping $\mathbf{y} = \mathbf{Ax} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}$
 - (1) $\mathbf{V}^T\mathbf{x}$: coefficients of \mathbf{x} along input directions, $\{\mathbf{v}_i\}_{i=1}^r$
 - (2) $\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}$: scaling by $\{\sigma_i\}_{i=1}^r$
 - (3) $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}$: reconstruction along output directions, $\{\mathbf{u}_i\}_{i=1}^r$



- $\mathbf{A}\mathbf{v}_i = \sigma_i\mathbf{u}_i$
 - (1) \mathbf{v}_1 is the most sensitive (highest gain) input direction
 - (2) \mathbf{u}_1 is the most sensitive (highest gain) output direction
- Geometric interpretation of $\mathbf{y} = \mathbf{Ax} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}$
 - (1) \mathbf{V}^T : rotation
 - (2) $\mathbf{\Sigma}$: scaling; stretched or compressed or squashed to zero ($\sigma_i = 0$)
 - (3) Zero padding (if $m > n$) or truncation (if $m < n$) to get m -vector
 - (4) \mathbf{U} : rotation

• Pseudo-inverse

- For $\mathbf{A} \in \mathbf{R}^{m \times n}$, $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ and $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$ in compact form

(1) If $m > n$ (narrow) and $\text{rank}(\mathbf{A}) = n$, $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \Rightarrow$ least squares solution

(2) If $m < n$ (wide) and $\text{rank}(\mathbf{A}) = m$, $\mathbf{A}^+ = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \Rightarrow$ minimum norm solution

- Properties

(1) $\mathbf{A}^+\mathbf{A} = \mathbf{I}_n$

(2) $\mathbf{A}\mathbf{A}^+ = \mathbf{P}_A$ (see Projection)

(3) $\mathbf{A}^+\mathbf{A}\mathbf{A} = \mathbf{A}$

(4) $\mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+$

• Projection

- Assume $R(\mathbf{A}) = \mathbf{R}^n$ for some matrix $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_p \mathbf{a}_{p+1} \cdots \mathbf{a}_n] \in \mathbf{R}^{m \times n}$, then

(1) $\text{rank}(\mathbf{A}) = n$

(2) $\{\mathbf{a}_i\}_{i=1}^n$ are linearly independent.

(3) $\text{span}\{\mathbf{a}_i\}_{i=1}^n = \mathbf{R}^n$

- Partition \mathbf{A} into \mathbf{S} and \mathbf{N}

(1) $\mathbf{A} = [\mathbf{S} \quad \mathbf{N}]$

(2) $\mathbf{S} = [\mathbf{a}_1 \cdots \mathbf{a}_p] \in \mathbf{R}^{m \times p}$ and let the subspace $S = \text{span}\{\mathbf{a}_i\}_{i=1}^p = R(\mathbf{S})$

(3) $\mathbf{N} = [\mathbf{a}_{p+1} \cdots \mathbf{a}_n] \in \mathbf{R}^{m \times (n-p)}$ and let the subspace $N = \text{span}\{\mathbf{a}_i\}_{i=p+1}^n = R(\mathbf{N})$

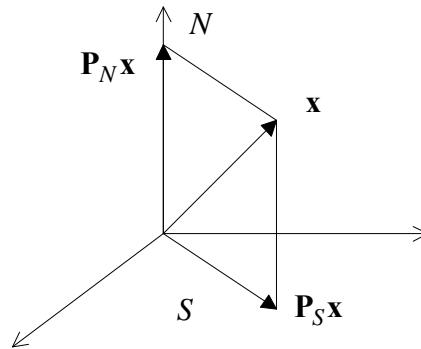
- By SVD, $\mathbf{S} = \mathbf{U}_S \mathbf{\Sigma}_S \mathbf{V}_S^T$ and $\mathbf{N} = \mathbf{U}_N \mathbf{\Sigma}_N \mathbf{V}_N^T$

(1) $S = \text{span}\{\mathbf{a}_i\}_{i=1}^p = R(\mathbf{S}) = R(\mathbf{U}_S)$

(2) $N = \text{span}\{\mathbf{a}_i\}_{i=p+1}^n = R(\mathbf{N}) = R(\mathbf{U}_N)$

(3) $S \oplus N = \mathbf{R}^n$ and $S^\perp = N$

- Projection: for any $\mathbf{x} \in \mathbf{R}^n$, $\mathbf{x} = \mathbf{P}_S \mathbf{x} + \mathbf{P}_N \mathbf{x}$



$$(1) \mathbf{P}_S = \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T = \mathbf{U}_S \mathbf{U}_S^T \quad \text{and} \quad \mathbf{P}_N = \mathbf{N}(\mathbf{N}^T \mathbf{N})^{-1} \mathbf{N}^T = \mathbf{U}_N \mathbf{U}_N^T = \mathbf{I}_n - \mathbf{U}_S \mathbf{U}_S^T$$

$$(2) \text{ Idempotent: } \mathbf{P}_S^T = \mathbf{P}_S = \mathbf{P}_S^2 \quad \text{and} \quad \mathbf{P}_N^T = \mathbf{P}_N = \mathbf{P}_N^2$$

$$(3) \mathbf{P}_S \mathbf{P}_N = \mathbf{P}_N \mathbf{P}_S = \mathbf{0} \quad \text{and} \quad \mathbf{P}_S + \mathbf{P}_N = \mathbf{I}_n$$

$$(4) \text{ Let } \mathbf{U} = [\mathbf{U}_S \quad \mathbf{U}_N]$$

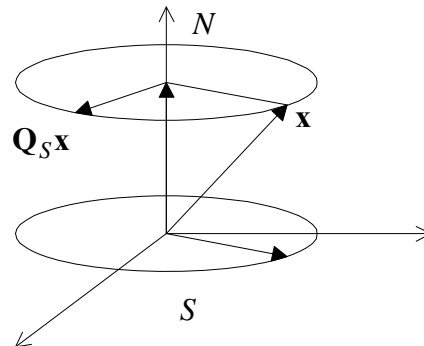
$$(a) \mathbf{U}^T \mathbf{P}_S \mathbf{U} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$(b) \mathbf{U}^T \mathbf{P}_N \mathbf{U} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{(n-p)} \end{bmatrix}$$

$$(5) \text{ Eigenvalues of } \mathbf{P}_S \text{ and } \mathbf{P}_N \text{ are 1 and 0.}$$

• Rotation

- Same setup as above in projection
- Rotate in the subspace S or rotate around the subspace N
- $\tilde{\mathbf{x}} = \mathbf{Q}_S \mathbf{x}$ with $\mathbf{Q}_S = \mathbf{U}_S \mathbf{Q} \mathbf{U}_S^T + \mathbf{P}_N$ and $\mathbf{Q}^T \mathbf{Q} = \mathbf{Q} \mathbf{Q}^T = \mathbf{I}_n$
- $\mathbf{P}_N \mathbf{x}$ component is not changed.
- $\mathbf{P}_S \mathbf{Q}_S = \mathbf{Q}_S \mathbf{P}_S$



• SVD in estimation and inversion

- Model: $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$

(1) \mathbf{x} is what we want to estimate.

(2) \mathbf{y} is sensor measurements.

(3) \mathbf{v} is unknown noise or measurement error with $\|\mathbf{v}\| \leq \alpha$

- Optimality of least squares estimator

(1) Consider a linear estimator $\hat{\mathbf{x}} = \mathbf{B}\mathbf{y}$ with $\mathbf{B}\mathbf{A} = \mathbf{I}$

(2) Error is $\mathbf{e} = \hat{\mathbf{x}} - \mathbf{x} = \mathbf{B}\mathbf{v}$ and the set of possible estimation errors is an ellipsoid

$$\mathbf{e} \in E = \{\mathbf{B}\mathbf{v} : \|\mathbf{v}\| \leq \alpha\}$$

(3) Uncertainty ellipsoid E must be small since $\|\mathbf{e}\| = \|\hat{\mathbf{x}} - \mathbf{x}\| \leq \alpha \|\mathbf{B}\|$

(4) Using SVD, we can prove that $\|\mathbf{A}^+\| = \|\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T\| = \|(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\| \leq \|\mathbf{B}\|$

• Condition number

- Error analysis of $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ for $\mathbf{A} \in \mathbf{R}^{n \times n}$ and invertible

(1) Assume error or noise in \mathbf{y} as $\mathbf{y} + \delta\mathbf{y}$

(2) $\mathbf{x} = \mathbf{x} + \delta\mathbf{x}$ with $\delta\mathbf{x} = \mathbf{A}^{-1}\delta\mathbf{y}$ and $\|\delta\mathbf{x}\| = \|\mathbf{A}^{-1}\delta\mathbf{y}\| \leq \|\mathbf{A}^{-1}\| \|\delta\mathbf{y}\|$

$$(3) \frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}\| \|\mathbf{A}^{-1}\| \frac{\|\delta\mathbf{y}\|}{\|\mathbf{y}\|} = \kappa(\mathbf{A}) \frac{\|\delta\mathbf{y}\|}{\|\mathbf{y}\|}$$

(4) Condition number of \mathbf{A} is $\kappa(\mathbf{A}) = \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})} \Rightarrow$ can compute using SVD

(5) If $\kappa(\mathbf{A})$ is large, \mathbf{A} is practically singular and ill-conditioned. If small, well-

conditioned.

(6) $\sigma_{\min}(\mathbf{A})$ is distance to nearest singular matrix.

- Use regularization for the solution of $\mathbf{y} = \mathbf{A}\mathbf{x}$ with ill-conditioned \mathbf{A} .

• **Low rank approximation or model simplification**

- $\mathbf{A} \in \mathbf{R}^{m \times n}$, $\text{rank}(\mathbf{A}) = r$, and $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T$
- Assume $\sigma_1 > \sigma_2 > \dots > \sigma_p$ and $\{\sigma_i\}_{i=p+1}^r$ are all small, then we can approximate \mathbf{A} by the optimal rank p approximator,

$$\hat{\mathbf{A}} = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^T \quad \text{and} \quad \|\mathbf{A} - \hat{\mathbf{A}}\| = \left\| \sum_{i=p+1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T \right\| = \sigma_{p+1} + \dots + \sigma_r$$

Solution of $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{A} \in \mathbf{R}^{n \times n}$

• Matrix multiplication, $\mathbf{C} = \mathbf{AB}$

- $c_{ij} = \tilde{\mathbf{a}}_i^T \mathbf{b}_j$, $\mathbf{c}_j = \mathbf{A} \mathbf{b}_j$, $\tilde{\mathbf{c}}_i = \tilde{\mathbf{a}}_i^T \mathbf{B}$
- $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
- $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$ and $(\mathbf{B} + \mathbf{C})\mathbf{D} = \mathbf{BD} + \mathbf{CD}$
- $\mathbf{FE} \neq \mathbf{EF}$, in general

• Matrix transpose, \mathbf{A}^T

- $(\mathbf{A}^T)_{ij} = (\mathbf{A})_{ji}$
- $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$

• System of linear equations

- Consider a system of linear equations, $\mathbf{y} = \mathbf{Ax}$, i.e.,

$$y_1 = a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n$$

$$\vdots$$

$$y_n = a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n$$

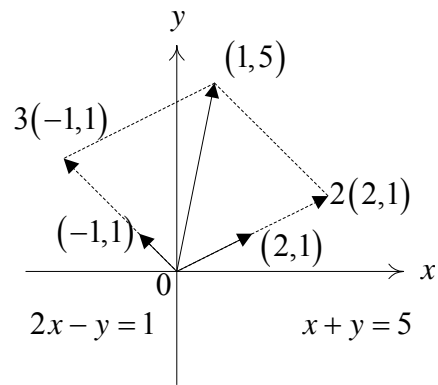
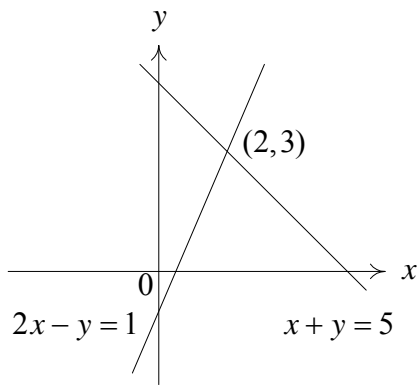
$$\text{where } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \in \mathbf{R}^n, \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \in \mathbf{R}^{n \times n}, \quad \text{and } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbf{R}^n$$

- Three possibilities: one solution, no solution, infinite solutions

• Geometry

- Example:

$$\begin{aligned} 2x - y = 1 \\ x + y = 5 \end{aligned} \quad \text{or} \quad x \begin{bmatrix} 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$



- In \mathbf{R}^3 , $ax + by + cz = d$ is a plane. If $d = 0$, it passes the origin. If $d \neq 0$, it is parallel to the plane with $d = 0$. All of these planes are perpendicular to the vector $[a, b, c]^T$.
- $\mathbf{Ax} = \mathbf{b}$, $\mathbf{A} \in \mathbf{R}^{n \times n}$

(1) Intersection of n planes:

$$\begin{bmatrix} \tilde{\mathbf{a}}_1^T \\ \tilde{\mathbf{a}}_2^T \\ \vdots \\ \tilde{\mathbf{a}}_n^T \end{bmatrix} \mathbf{x} = \mathbf{b} \quad \text{or} \quad \begin{cases} \tilde{\mathbf{a}}_1^T \mathbf{x} = b_1 \\ \tilde{\mathbf{a}}_2^T \mathbf{x} = b_2 \\ \vdots \\ \tilde{\mathbf{a}}_n^T \mathbf{x} = b_n \end{cases}$$

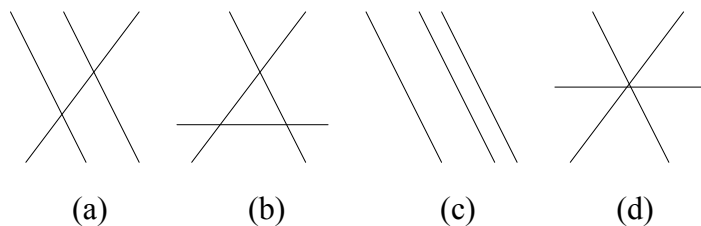
(2) Linear combination of column vectors equals to \mathbf{b} :

$$[\mathbf{a}_1 \mathbf{a}_2 \cdots \mathbf{a}_n] \mathbf{x} = \mathbf{b} \quad \text{or} \quad x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \cdots + x_n \mathbf{a}_n = \mathbf{b}$$

(3) Coordinates of intersection point = coefficients of linear combination

- Singularity: no solution or infinite solutions

(1) No intersection point or infinite intersection points



- (a) 2 parallel: no solution
- (b) No intersection: no solution
- (c) All parallel: no solution
- (d) Line of intersection: infinite solutions

(2) $\mathbf{b} \notin \text{span}\{\mathbf{a}_i\}_{i=1}^n$

• Gaussian elimination

- (Forward elimination + backsubstitution) or (factorization + forward substitution + backsubstitution)
- Example:

$$\begin{cases} 2x + y + z = 5 \\ 4x - 6y = 2 \\ -2x + 7y + 2z = 9 \end{cases} \Rightarrow$$

$$\begin{bmatrix} 2 & 1 & 1 & 5 \\ 4 & -6 & 0 & -2 \\ -2 & 7 & 2 & 9 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 8 & 3 & 14 \end{bmatrix} \Rightarrow \begin{bmatrix} 2 & 1 & 1 & 5 \\ 0 & -8 & -2 & -12 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

- (1) Pivots cannot be zero.
- (2) If pivot is zero, change order (rows) if possible (permutation).
- (3) If all possible pivots are zero, \mathbf{A} is singular (i.e., \mathbf{A} is not invertible).
- After elimination with $O(n^3)$, solve for \mathbf{x} by back substitution with $O(n^2)$.

• LU or LDU factorization

- For repeated solutions with different RHS, store multipliers during elimination process in a lower triangular matrix \mathbf{L} . Main diagonal elements of \mathbf{L} are all one.
- Store pivots and the rest of each row in an upper triangular matrix \mathbf{U} . Main diagonal elements of \mathbf{U} are pivots.

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \text{ and } \mathbf{U} = \begin{bmatrix} 2 & 1 & 1 \\ 0 & -8 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

- We can split \mathbf{U} into \mathbf{D} and \mathbf{U} as

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -8 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } \mathbf{U} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

- Using LU factorization, $\mathbf{A} = \mathbf{LU}$. For a given \mathbf{b} ,
 - (a) Forward substitution with $O(n^2)$: solve $\mathbf{Lc} = \mathbf{b}$ for \mathbf{c}
 - (b) Backsubstitution with $O(n^2)$: solve $\mathbf{Ux} = \mathbf{c}$ for \mathbf{x}
- Using LDU factorization, $\mathbf{A} = \mathbf{LDU}$. For a given \mathbf{b} ,

- (a) Forward substitution with $O(n^2)$: solve $\mathbf{Lc} = \mathbf{b}$ for \mathbf{c}
- (b) Simple divisions with $O(n)$: solve $\mathbf{De} = \mathbf{c}$ for \mathbf{e}
- (c) Backsubstitution with $O(n^2)$: solve $\mathbf{Ux} = \mathbf{e}$ for \mathbf{x}

• **Inverse matrix**

- $\mathbf{Ax} = \mathbf{b}$ has a unique solution. $\Leftrightarrow \mathbf{A}$ is nonsingular. $\Leftrightarrow \mathbf{A}^{-1}$ exists.
- If \mathbf{A}^{-1} exists, $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ and $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ and $\mathbf{AA}^{-1} = \mathbf{I}$.
- All pivots are nonzero. $\Leftrightarrow \mathbf{A}^{-1}$ exists.
- All eigenvalues are nonzero. $\Leftrightarrow \mathbf{A}^{-1}$ exists.
- Determinant of \mathbf{A} is nonzero. $\Leftrightarrow \mathbf{A}^{-1}$ exists.
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, $(\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$

• **Symmetric nonsingular matrix**

- \mathbf{A} is symmetric $\Leftrightarrow \mathbf{A} = \mathbf{A}^T$
- If \mathbf{A} is symmetric and nonsingular, $\mathbf{A} = \mathbf{LU} = \mathbf{LL}^T = \mathbf{U}^T\mathbf{U}$: Cholesky factorization
- If \mathbf{A} is symmetric and nonsingular, $\mathbf{A} = \mathbf{LDU} = \mathbf{LDL}^T = \mathbf{U}^T\mathbf{DU}$

• **Comments**

- If all pivots are positive, \mathbf{A} is called positive definite.
- If \mathbf{A} is ill-conditioned,
 - (1) Computation is sensitive to a small numerical error.
 - (2) Use partial pivoting where we pick up the largest pivot among all possible pivots.

This requires permutation.
- If \mathbf{A} is not ill-conditioned, it is well-conditioned.

Solution of $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{A} \in \mathbf{R}^{m \times n}$

• $\mathbf{Ax} = \mathbf{b}$

- Scalar example: $ax = b$
 - (a) $a \neq 0 \Rightarrow$ unique solution, nonsingular
 - (b) $a = 0$ and $b = 0 \Rightarrow$ infinite number of solutions, underdetermined
 - (c) $a = 0$ and $b \neq 0 \Rightarrow$ no solution, inconsistent
- $\mathbf{Ax} = \mathbf{b}$, $\mathbf{A} \in \mathbf{R}^{m \times n}$
 - (a) Infinite number of solutions for every \mathbf{b}
 - (b) Infinite number of solutions for some \mathbf{b} and no solution for other \mathbf{b}
 - (c) One solution for some \mathbf{b} and no solution for other \mathbf{b}
- Example

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 6 & 2 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \mathbf{U}$$

$$\mathbf{U} = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} : \text{echelon form, } \mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix}, \text{ and } \mathbf{A} = \mathbf{LU}$$

- $\mathbf{PA} = \mathbf{LU}$ for any $\mathbf{A} \in \mathbf{R}^{m \times n}$ with permutation \mathbf{P}

• Homogeneous solution, $\mathbf{Ax} = \mathbf{0}$

- $\mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{Ux} = \mathbf{0}$
- Example

$$\begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} x_3 = -\frac{1}{3}x_4 \\ x_1 = -3x_2 - x_4 \end{cases} \Rightarrow \mathbf{x} = \begin{bmatrix} -3x_2 - x_4 \\ x_2 \\ -\frac{1}{3}x_4 \\ x_4 \end{bmatrix} = x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{3} \\ 1 \end{bmatrix}$$

- Two kinds of variables
 - (a) Column with pivot \Rightarrow basic variable (x_1 and x_3)
 - (b) Column without pivot \Rightarrow free variable (x_2 and x_4)
- Procedure for homogeneous solution

- (a) Identify basic and free variables from $\mathbf{Ux} = \mathbf{0}$
- (b) For each free variable, set it to 1 and set all other free variables to 0. Solve $\mathbf{Ux} = \mathbf{0}$
- (c) Combinations of all solutions from (b) equal to $N(\mathbf{A})$
- Comments
 - (a) If $n > m$ (wide matrix), nontrivial homogeneous solutions ($\mathbf{x} \neq \mathbf{0}$) of $\mathbf{Ax} = \mathbf{0}$ exist.
 - (b) Number of free variables = $\dim\{N(\mathbf{A})\}$
 - (c) Finding homogeneous solutions is equivalent to finding $N(\mathbf{A})$.

• **Inhomogeneous or particular solution, $\mathbf{Ax} = \mathbf{b} \neq \mathbf{0}$**

- $\mathbf{Ax} = \mathbf{b} \Rightarrow \mathbf{Ux} = \mathbf{c}$
- Example

$$\begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 - 2b_1 \\ b_3 - 2b_2 + 5b_1 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 5 \\ -1 & -3 & 3 & 0 \end{bmatrix} = [\mathbf{a}_1 \mathbf{a}_2 \mathbf{a}_3 \mathbf{a}_4]$$

$$R(\mathbf{A}) = \text{span}\{\mathbf{a}_1, \mathbf{a}_3\} \text{ or } R(\mathbf{A}) = \{(b_1, b_2, b_3) : b_3 - 2b_2 + 5b_1 = 0\} \text{ : plane}$$

$$\text{If } \mathbf{b} = [1 \ 5 \ 5]^T, \begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \text{ and } \left\{ \begin{array}{l} x_3 = 1 - \frac{1}{3}x_4 \\ x_1 = -2 - 3x_2 - x_4 \end{array} \right\}.$$

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \underbrace{\begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\text{particular solution}} + \underbrace{x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 0 \\ -\frac{1}{3} \\ 1 \end{bmatrix}}_{\text{homogeneous solution}}$$

- Procedure for particular solution
 - (a) $\mathbf{Ax} = \mathbf{b} \Rightarrow \mathbf{Ux} = \mathbf{c}$

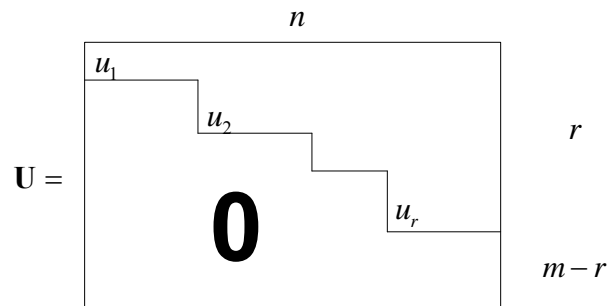
(b) Set all free variables to 0 and solve $\mathbf{U}\mathbf{x} = \mathbf{c}$ for \mathbf{x}

• **Total solution, $\mathbf{A}\mathbf{x} = \mathbf{b} \neq \mathbf{0}$**

- Total solution = particular solution + homogeneous solution ($\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$)
- Number of basic variables = r = number of pivots = $\text{rank}(\mathbf{A})$ = number of independent columns = number of independent rows
- Number of free variables = $n - r$

• **Comments**

- Echelon form with r pivots, $\{u_i\}_{i=1}^r$



- $r = n \Rightarrow$ no free variable, $N(\mathbf{A}) = \{\mathbf{0}\}$, nontrivial homogeneous solution does not exist, $\{\mathbf{a}_i\}_{i=1}^n$ are linearly independent
- $r < n \Rightarrow (n - r)$ free variables, nontrivial homogeneous solution exists
- $r = m \Rightarrow$ particular solution exists for any $\mathbf{b} \in \mathbf{R}^m$, $R(\mathbf{A}) = \mathbf{R}^m$
- $r < m$
 - (a) $c_{m-r} = c_{m-r+1} = \dots = c_m = 0 \Rightarrow$ system is consistent and particular solution exists
 - (b) $c_i \neq 0$ for some $m - r \leq i \leq m \Rightarrow$ system is inconsistent and particular solution does not exist
- $\text{rank}(\mathbf{A}) = r \leq \min(m, n)$

• **Inverse of $\mathbf{A} \in \mathbf{R}^{m \times n}$**

- Left inverse: $\mathbf{B}\mathbf{A} = \mathbf{I}_n, \mathbf{B} \in \mathbf{R}^{n \times m}$

- Right inverse: $\mathbf{AC} = \mathbf{I}_m, \mathbf{C} \in \mathbf{R}^{m \times n}$
- If both \mathbf{B} and \mathbf{C} exist, $\mathbf{B} = \mathbf{B}(\mathbf{AC}) = (\mathbf{BA})\mathbf{C} = \mathbf{C} = \mathbf{A}^{-1}$ and $m = n$
- **Existence and uniqueness of solution of $\mathbf{Ax} = \mathbf{b}$, $\mathbf{A} \in \mathbf{R}^{m \times n}$**
 - Existence ($m \geq n$, narrow matrix): one or many solutions
 - (a) $\mathbf{Ax} = \mathbf{b}$ has “at least” one solution for every \mathbf{b} iff $R(\mathbf{A}) = \mathbf{R}^m$.
 - (b) In this case, $r = m$ and right inverse $\mathbf{C} \in \mathbf{R}^{m \times n}$ exists such that $\mathbf{AC} = \mathbf{I}_m$.
 - (c) $\mathbf{B} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$ and $\text{rank}(\mathbf{A}^T \mathbf{A}) = n$
 - (d) Least squares (LS) solution
 - Uniqueness ($m \leq n$, wide matrix): one or zero solution
 - (a) $\mathbf{Ax} = \mathbf{b}$ has “at most” one solution for every \mathbf{b} iff $n = r = \dim R(\mathbf{A})$.
 - (b) In this case, $r = n$ and left inverse $\mathbf{B} \in \mathbf{R}^{n \times m}$ exists such that $\mathbf{BA} = \mathbf{I}_n$.
 - (c) $\mathbf{C} = \mathbf{A}^T (\mathbf{AA}^T)^{-1}$ and $\text{rank}(\mathbf{AA}^T) = m$
 - (d) Minimum norm solution
-
- **Square matrix ($m = n$)**
 - $\mathbf{B} = \mathbf{C} = \mathbf{A}^{-1}$ iff $m = n = r$
 - Followings are equivalent
 - (a) $R(\mathbf{A}) = \mathbf{R}^m$, $\mathbf{Ax} = \mathbf{b}$ has at least one solution for any \mathbf{b}
 - (b) $N(\mathbf{A}) = \{\mathbf{0}\}$, columns of \mathbf{A} are linearly independent
 - (c) $R(\mathbf{A}^T) = \mathbf{R}^n$
 - (d) Rows of \mathbf{A} are linearly independent
 - (e) $\mathbf{PA} = \mathbf{LDU}$ with all $d_{ii} \neq 0$ (nonzero pivot)
 - (f) \mathbf{A}^{-1} exists such that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$
 - (g) $\det(\mathbf{A}) \neq 0$
 - (h) All eigenvalues of \mathbf{A} are nonzero
 - (i) $\mathbf{A}^T \mathbf{A}$ is positive definite
 - Example: polynomial of degree $(n - 1)$ has $(n - 1)$ roots \Rightarrow Vandermonde’s matrix

$$\mathbf{Ax} = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \mathbf{b}, \mathbf{A} \text{ is nonsingular}$$

Four Subspaces of $\mathbf{A} \in \mathbf{R}^{m \times n}$

• Four subspaces of $\mathbf{A} \in \mathbf{R}^{m \times n}$

- Column space of \mathbf{A} = range of \mathbf{A} = $R(\mathbf{A}) = \text{span}\{\mathbf{a}_i\}_{i=1}^n \subset \mathbf{R}^m$, $\dim R(\mathbf{A}) = r$
- Nullspace of \mathbf{A} = $N(\mathbf{A}) = \{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{0}, \mathbf{x} \in \mathbf{R}^n\} \subset \mathbf{R}^n$, $\dim N(\mathbf{A}) = n - r$
- Row space of \mathbf{A} = range of \mathbf{A}^T = $R(\mathbf{A}^T) = \text{span}\{\tilde{\mathbf{a}}_i\}_{i=1}^m \subset \mathbf{R}^n$, $\dim R(\mathbf{A}^T) = r$
- Left nullspace of \mathbf{A} = nullspace of \mathbf{A}^T = $N(\mathbf{A}^T) = \{\mathbf{y} : \mathbf{A}^T\mathbf{y} = \mathbf{0}, \mathbf{y} \in \mathbf{R}^m\} \subset \mathbf{R}^m$,
 $\dim N(\mathbf{A}^T) = m - r$

• $\mathbf{A} \in \mathbf{R}^{m \times n}$ and echelon form \mathbf{U} ($\mathbf{PA} = \mathbf{LU}$)

- $R(\mathbf{A}^T) = R(\mathbf{U}^T) \subset \mathbf{R}^n$
 - (a) $\dim R(\mathbf{A}^T) = \dim R(\mathbf{U}^T) = \text{number of pivots} = r = \text{row rank} = \text{column rank}$
 - (b) Basis of $R(\mathbf{A}^T) = R(\mathbf{U}^T) = r$ nonzero rows of \mathbf{U}
- $N(\mathbf{A}) = N(\mathbf{U}) \subset \mathbf{R}^n$ ($\mathbf{A}\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{U}\mathbf{x} = \mathbf{0}$)
 - (a) $\dim N(\mathbf{A}) = \dim N(\mathbf{U}) = n - r = \text{number of free variables} = \text{nullity of } \mathbf{A}$
 - (b) $\ker(\mathbf{A}) = N(\mathbf{A}) = N(\mathbf{U}) \subset \mathbf{R}^n$: kernel of \mathbf{A}
 - (c) For each free variable, we can generate one basis vector
- $R(\mathbf{A}) \neq R(\mathbf{U}), R(\mathbf{A}) \subset \mathbf{R}^m$
 - (a) $R(\mathbf{U}) = \text{span}\{\mathbf{u}_i\}_{i=1}^n \subset \mathbf{R}^m$
 - (b) $\dim R(\mathbf{A}) = \dim R(\mathbf{U}) = \text{number of pivots} = r = \text{column rank} = \text{row rank}$

• Dimensions

- $\dim R(\mathbf{A}) = \dim R(\mathbf{A}^T) = \text{number of pivots} = \text{rank}(\mathbf{A})$
- Number of basic variables (r) + number of free variables ($n - r$) = number of columns (n) $\Leftrightarrow \dim R(\mathbf{A}) + \dim N(\mathbf{A}) = n$
- $r + (m - r) = m \Leftrightarrow \dim R(\mathbf{A}^T) + \dim N(\mathbf{A}^T) = m$

Linear Transformation, $y = Ax$

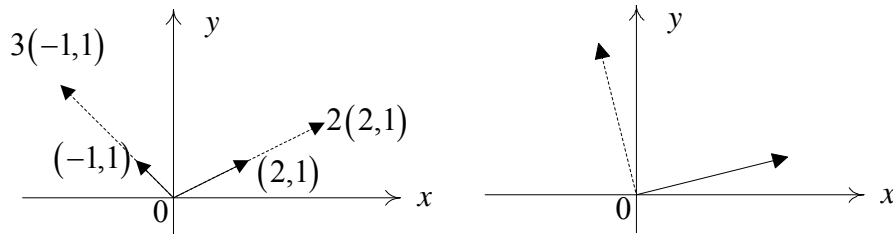
• **Linear transformation, $y = Ax$, $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$**

- Linearity: $A(cx + dy) = c(Ax) + d(Ay)$

- Examples

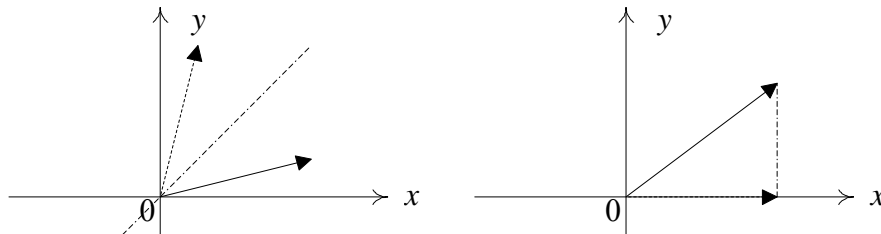
(a) Stretching: $A = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix} = cI$

(b) Rotation: $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$



(c) Reflection: $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

(d) Projection: $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$



- Linear transformation matrix A depends on basis

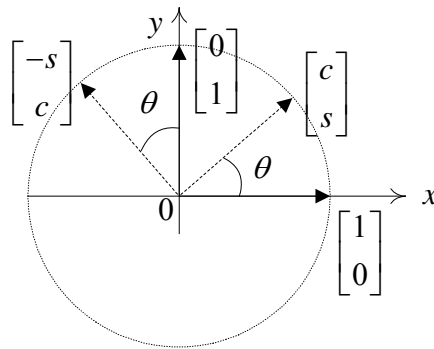
• **Basis and linear transformation**

- $v = \sum_{i=1}^n c_i x_i$ with a basis $\{x_i\}_{i=1}^n$, then $Av = \sum_{i=1}^n c_i Ax_i$

- $A: V \rightarrow W$ with $V = \text{span}\{x_i\}_{i=1}^n$ and $W = \text{span}\{y_j\}_{j=1}^m \Rightarrow Ax_i = \sum_{j=1}^m a_{ji} y_j$

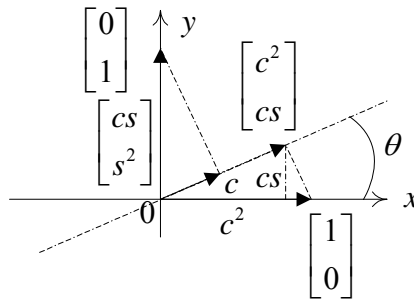
- $A: V \rightarrow W$ and $B: U \rightarrow V \Rightarrow AB: U \rightarrow W$ (i.e., $z = Ay = ABx$)

• **Rotation**



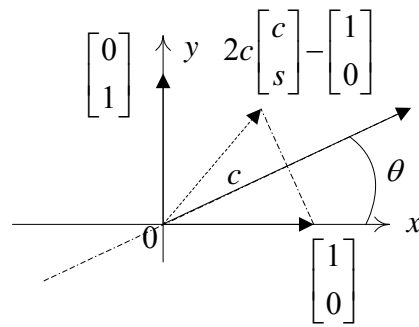
- $\mathbf{Q}_\theta = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$
- $\mathbf{Q}_\theta \mathbf{Q}_{-\theta} = \begin{bmatrix} c & -s \\ s & c \end{bmatrix} \begin{bmatrix} c & s \\ -s & c \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
- $\mathbf{Q}_\theta^2 = \mathbf{Q}_\theta \mathbf{Q}_\theta = \mathbf{Q}_{2\theta}$ and $\mathbf{Q}_\theta \mathbf{Q}_\varphi = \mathbf{Q}_{\theta+\varphi}$

• **Projection**



- $\mathbf{P}_\theta = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix}$
- $\mathbf{P}_\theta^2 = \mathbf{P}_\theta \mathbf{P}_\theta = \mathbf{P}_\theta$: idempotent
- $\mathbf{P}_\theta^T = \mathbf{P}_\theta$: symmetric

• **Reflection**



- $\mathbf{H}_\theta = \begin{bmatrix} 2c^2 - 1 & 2cs \\ 2cs & 2s^2 - 1 \end{bmatrix}$
- $\mathbf{H}_\theta = 2\mathbf{P}_\theta - \mathbf{I}$ or $\mathbf{H}_\theta + \mathbf{I} = 2\mathbf{P}_\theta$
- $\mathbf{H}_\theta^2 = \mathbf{H}_\theta \mathbf{H}_\theta = (2\mathbf{P}_\theta - \mathbf{I})^2 = 4\mathbf{P}_\theta^2 - 4\mathbf{P}_\theta + \mathbf{I} = \mathbf{I} \Rightarrow \mathbf{H}_\theta^{-1} = \mathbf{H}_\theta$