Singular Value Decomposition

Not very matrix has an eigendecomposition, but every matrix has a singular value decomposition (SVD)

SVD: $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be decomposed into

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices and $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is such that the $(i, i)-$entry is a (nonnegative) singular value of $\mathbf{A}$.

Low-rank Approximation

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$ with rank k and $r \in \{1, \ldots, k-1\}$, find $\hat{\mathbf{A}} \in \mathbb{R}^{m \times n}$ with $\mathrm{rank}(\hat{\mathbf{A}}) \leq r$ such that $\|\mathbf{A} - \hat{\mathbf{A}}\|_2$ or $\|\mathbf{A} - \hat{\mathbf{A}}\|_F$ is minimum

Solution: Truncated SVD

SVD of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

where $\sigma_1 \geq \cdots \geq \sigma_k > \sigma_{k+1} = \cdots = \sigma_{\min\{m,n\}} = 0$ are the singular values of $\mathbf{A}$

$$\hat{\mathbf{A}} = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Image Compression

original image, size: 639 x 853

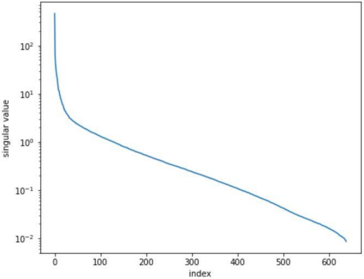


Image compression (cont'd)

compressed image with r = 10

compressed image with r = 15

compressed image with r = 20

compressed image with r = 30

Matrix Computations
Chapter 1 Introduction
Section 1.2 Review of Linear Algebra

Jie Lu
ShanghaiTech University

Notation

$\mathbb{R}$	the set of real numbers or real space
$\mathbb{C}$	the set of complex numbers or complex space
$\mathbb{R}^n$	n-dimensional real space
$\mathbb{C}^n$	n-dimensional complex space
$\mathbb{R}^{m\times n}$	the set of all $m \times n$ real-valued matrices
$\mathbb{C}^{m\times n}$	the set of all $m \times n$ complex-valued matrices
a	scalar in $\mathbb{C}$
a^*	conjugate of $a \in \mathbb{C}$
$\mathbf{x}$	vector
x_i, $[\mathbf{x}]_i$	ith entry of $\mathbf{x}$
$\mathbf{A}$	matrix
a_{ij}, $[\mathbf{A}]_{ij}$	(i, j)-entry of $\mathbf{A}$
$\mathbb{S}^n$	the set of all $n \times n$ real symmetric matrices, i.e, $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $a_{ij} = a_{ji}$ for all $i, j = 1, \ldots, n$
$\mathbb{H}^n$	the set of all $n \times n$ complex Hermitian matrices, i.e, $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $a_{ij} = a_{ji}^*$ for all i, j

$a = \alpha + j\beta$

$a^* = \alpha - j\beta$

Vector

- $\mathbf{x} \in \mathbb{R}^n$: $\mathbf{x}$ is a real-valued n-dimensional column vector, i.e.,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \qquad x_i \in \mathbb{R} \text{ for all } i$$

column vector

- $\mathbf{x} \in \mathbb{C}^n$: $\mathbf{x}$ is a complex-valued n-dimensional column vector

row vector

- Transpose: $\mathbf{x}^T = \begin{bmatrix} x_1, & x_2, & \dots, & x_n \end{bmatrix}$
- Hermitian transpose: $\mathbf{x}^H = \begin{bmatrix} x_1^*, & x_2^*, & \dots, & x_n^* \end{bmatrix}$

Matrix

- $\mathbf{A} \in \mathbb{R}^{m\times n}$: $\mathbf{A}$ is a real-valued $m \times n$ matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \qquad a_{ij} \in \mathbb{R} \text{ for all } i, j$$

- $\mathbf{A} \in \mathbb{C}^{m\times n}$: $\mathbf{A}$ is a complex-valued $m \times n$ matrix

- We may write

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1, & \mathbf{a}_2, & \dots, & \mathbf{a}_n \end{bmatrix}$$

where $\mathbf{a}_i \in \mathbb{R}^m$ is the ith column of matrix A

Matrix (Cont'd)

- A matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to be
  - square if $m = n$;
  - tall if $m > n$;
  - fat if $m < n$.

- A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be
  - upper triangular if $a_{ij} = 0$ for all $i > j$;
  - lower triangular if $a_{ij} = 0$ for all $i < j$.

Examples:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -4 & 5 \\ 0 & 0 & 6 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{8} & 3 & 0 \end{bmatrix}.$$

upper triangular

lower triangular

Matrix Transpose

- Given a $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & & & \vdots \\ a_{1n} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix}$$

  is a $n \times m$ matrix

- The following properties hold:
  - $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$
  - $(\mathbf{A}^T)^T = \mathbf{A}$
  - $(\mathbf{A}+\mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

Matrix Transpose (Cont'd)

- Hermitian transpose: Given $\mathbf{A} \in \mathbb{C}^{m \times n}$,

$$\mathbf{A}^H = \begin{bmatrix} a_{11}^* & a_{21}^* & \dots & a_{m1}^* \\ a_{12}^* & a_{22}^* & \dots & a_{m2}^* \\ \vdots & & & \vdots \\ a_{1n}^* & a_{m2}^* & \dots & a_{mn}^* \end{bmatrix} \in \mathbb{C}^{n \times m}$$

- The following properties hold:
  - $(\mathbf{AB})^H = \mathbf{B}^H \mathbf{A}^H$
  - $(\mathbf{A}^H)^H = \mathbf{A}$
  - $(\mathbf{A} + \mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H$

Matrix Trace and Matrix Power

- Given $\mathbf{A} \in \mathbb{R}^{n \times n}$, the trace of $\mathbf{A}$ is

$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$$

  - $\mathrm{tr}(\mathbf{A}^T) = \mathrm{tr}(\mathbf{A})$
  - $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$
  - $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ for $\mathbf{A}, \mathbf{B}$ of proper sizes

- Matrix power: Given $\mathbf{A} \in \mathbb{R}^{n \times n}$,

$$\mathbf{A}^k = \underbrace{\mathbf{A}\mathbf{A} \cdots \mathbf{A}}_{k \text{ times}}$$

Some Common Vectors and Matrices

- All-one vectors: We use the notation

$$\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

  to denote a vector of all 1's

- Zero vectors or matrices: We use the notation $\mathbf{0}$ to denote either a vector of all zeros or a matrix of all zeros

- Unit vectors: We use the notation

$$\mathbf{e}_i = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix}^T$$

  to denote a unit vector whose i-th entry is 1 and other entries are all zero

Some Common Vectors and Matrices (Cont'd)

- Identity matrix:

$$\mathbf{I} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

The empty entries are assumed to be zero by default

- Diagonal matrices: We use the notation

$$\mathrm{Diag}(a_1, \ldots, a_n) = \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix}$$

to denote a diagonal matrix whose diagonal entries are $a_1, \ldots, a_n$

For $\mathbf{a} = \begin{bmatrix} a_1, \ldots, a_n \end{bmatrix}^T$, we use the shorthand notation $\mathrm{Diag}(\mathbf{a})$

$\alpha x + (1-\alpha) y \quad \forall \alpha \in \mathbb{R}$

Subspace

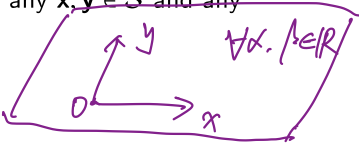
$\alpha x + (1-\alpha) y \quad \forall \alpha \in [0,1]$

$\mathbb{R}^m$ subspace

A subset $\mathcal{S}$ of $\mathbb{R}^m$ is said to be a subspace if for any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ and any $\alpha, \beta \in \mathbb{R}$,

$0 \in S$

$$\alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{S}$$



- If $\mathcal{S}$ is a subspace and $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$, then any linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$, i.e., $\sum_{i=1}^n \alpha_i \mathbf{a}_i$ for some $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, lies in $\mathcal{S}$

$\underbrace{[\alpha_1 a_1 + \alpha_2 a_2] + \alpha_3 a_3}_{\in S} + \cdots + \alpha_n a_n \in S$

- Let $\mathcal{S}_1, \mathcal{S}_2$ be subspaces of $\mathbb{R}^m$
  - $\mathcal{S}_1 + \mathcal{S}_2 := \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathcal{S}_1, \mathbf{y} \in \mathcal{S}_2\}$ is a subspace
  - $\mathcal{S}_1 \cap \mathcal{S}_2$ is a subspace

$\forall s, r \in S_1 + S_2, \ \exists x, x' \in S_1, \ y, y' \in S_2$ s.t.

$s = x + y, \quad r = x' + y'$

$\forall \alpha, \beta \in \mathbb{R}, \quad \alpha s + \beta r = \alpha(x+y) + \beta(x'+y') = \underbrace{\alpha x + \beta x'}_{\in S_1} + \underbrace{\alpha y + \beta y'}_{\in S_2}$

Span

The span of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ is defined as

$$\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \left\{ \mathbf{y} \in \mathbb{R}^m \;\middle|\; \mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i,\ \alpha_1, \ldots, \alpha_n \in \mathbb{R} \right\}$$

- $\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is the set of all linear combinations of $\mathbf{a}_1, \ldots, \mathbf{a}_n$
- $\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a subspace

Theorem

Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$. There exists a positive integer n and $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$ such that $\mathcal{S} = \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

- We can always represent a subspace by a span

$$Ax = [a_1 \cdots a_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 a_1 + \cdots + x_n a_n$$

Range and Nullspace

The range (space) of $\mathbf{A} \in \mathbb{R}^{m\times n}$ is defined as

subspace in $\mathbb{R}^m$

$$\mathcal{R}(\mathbf{A}) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = \mathbf{A}\mathbf{x},\ \mathbf{x} \in \mathbb{R}^n\}$$

- $\mathcal{R}(\mathbf{A})$ is the span of the columns of $\mathbf{A}$ column space

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

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

- $\mathcal{N}(\mathbf{A})$ is a subspace
- $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{B})$ for some integer $r > 0$ and $\mathbf{B} \in \mathbb{R}^{n\times r}$

$\forall x, x' \in \mathcal{N}(A)$,

$Ax = Ax' = 0$

$\forall \alpha, \beta \in \mathbb{R}$,

$A(\alpha x + \beta x') = 0$

$\Rightarrow \alpha x + \beta x' \in \mathcal{N}(A)$

Linear Independence

$\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ are said to be linearly independent if

$$\sum_{i=1}^{n} \alpha_i \mathbf{a}_i \neq \mathbf{0} \quad \text{for all } \alpha = \left[\alpha_1, \ldots, \alpha_n\right]^T \in \mathbb{R}^n \text{ with } \alpha \neq \mathbf{0}$$

and linearly dependent otherwise

- Equivalent definition of linear dependence: $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ are linearly dependent if there exists $\alpha \in \mathbb{R}^n$, $\alpha \neq \mathbf{0}$ such that

$$\sum_{i=1}^{n} \alpha_i \mathbf{a}_i = \mathbf{0}$$

Linear Independence (Cont'd)

- If $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathbb{R}^m$ are linearly independent, then any $\mathbf{a}_j$ *cannot* be a linear combination of the other $\mathbf{a}_i$'s
- If $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathbb{R}^m$ are linearly dependent, then *there exists* an $\mathbf{a}_j$ such that $\mathbf{a}_j$ is a linear combination of the other $\mathbf{a}_i$'s
- If $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathbb{R}^m$ are linearly independent, then $n \leq m$

Linear Independence (Cont'd)

- If $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ are linearly independent and $\mathbf{y} \in \mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, then the coefficient $\alpha = [\alpha_1, \ldots, \alpha_n]^T$ for the representation

$$\mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$$

is unique, i.e., there does *not* exist $\beta = [\beta_1, \ldots, \beta_n]^T \in \mathbb{R}^n$, $\beta \neq \alpha$ such that $\mathbf{y} = \sum_{i=1}^{n} \beta_i \mathbf{a}_i$

Assume to the contrary that $\exists \beta \neq \alpha$ s.t. $y = \sum_{i=1}^{n} \beta_i a_i$

$$0 = y - y = \sum_{i=1}^{n} \alpha_i a_i - \sum_{i=1}^{n} \beta_i a_i = \sum_{i=1}^{n} (\alpha_i - \beta_i) a_i$$

Since $a_1, \ldots a_n$ are linearly independent.

$\alpha_i - \beta_i = 0 \quad \forall i = 1, \ldots, n$ Contradiction

Linear Independence (Cont'd)

Let $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$, and denote $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ as an index subset with $k \leq n$ and $i_j \neq i_\ell$ for all $j \neq \ell$.
A vector subset $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is called a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ if both of the following conditions hold:

1. $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is linearly independent
2. $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is not contained by any other linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$

- A set of non-redundant vectors from $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$

Linear Independence (Cont'd)

Example:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

The linearly independent subsets of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ are

$$\begin{gathered} \{\mathbf{a}_1\}, \{\mathbf{a}_2\}, \{\mathbf{a}_3\}, \{\mathbf{a}_4\}, \\ \{\mathbf{a}_1, \mathbf{a}_2\}, \{\mathbf{a}_1, \mathbf{a}_3\}, \{\mathbf{a}_1, \mathbf{a}_4\}, \{\mathbf{a}_2, \mathbf{a}_3\}, \{\mathbf{a}_2, \mathbf{a}_4\}, \{\mathbf{a}_3, \mathbf{a}_4\}, \\ \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}, \quad \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}, \quad \{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\} \end{gathered}$$

The maximal linearly independent subsets are

Linear Independence (Cont'd)

Example:

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{a}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{a}_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

The linearly independent subsets of $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ are

$$\begin{gathered} \{\mathbf{a}_1\}, \{\mathbf{a}_2\}, \{\mathbf{a}_3\}, \{\mathbf{a}_4\}, \\ \{\mathbf{a}_1, \mathbf{a}_2\}, \{\mathbf{a}_1, \mathbf{a}_3\}, \{\mathbf{a}_1, \mathbf{a}_4\}, \{\mathbf{a}_2, \mathbf{a}_3\}, \{\mathbf{a}_2, \mathbf{a}_4\}, \{\mathbf{a}_3, \mathbf{a}_4\}, \\ \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}, \quad \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}, \quad \{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\} \end{gathered}$$

The maximal linearly independent subsets are

$$\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}, \quad \{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_4\}, \quad \{\mathbf{a}_1, \mathbf{a}_3, \mathbf{a}_4\}$$

Linear Independence (Cont'd)

$x = \sum_{i=1}^{n} \alpha_i a_i = \alpha_{i_1} a_{i_1} + \cdots + \alpha_{i_k} a_{i_k}$

$+ \sum_{j \neq i_1, \ldots, i_k} \alpha_j a_j$ a linear combination of $a_{i_1}, \ldots, a_{i_k}$

$\in \text{span}\{a_{i_1}, \ldots, a_{i_k}\}$

Facts:

- $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ if and only if $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}, \mathbf{a}_j\}$ is linearly dependent for any $j \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$

- If $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$, then

clearly,

$$\text{span}\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\} = \text{span}\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$$

$\text{span}\{a_{i_1}, \ldots, a_{i_k}\} \subseteq \text{span}\{a_1, \ldots, a_n\}$

We want to show $\text{span}\{a_1, \ldots, a_n\} \subseteq \text{span}\{a_{i_1}, \ldots, a_{i_k}\}$

For any $x = \sum_{i=1}^{n} \alpha_i a_i \in \text{span}\{a_1, \ldots, a_n\}$,

Basis

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a subspace with $\mathcal{S} \neq \{\mathbf{0}\}$.
A vector set $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\} \subset \mathbb{R}^m$ is called a basis for $\mathcal{S}$ if both of the following hold:

1. $\mathbf{b}_1, \ldots, \mathbf{b}_k$ are linearly independent
2. $\mathcal{S} = \mathrm{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$

- If $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent vector subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, then $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a basis for $\mathrm{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$

- Given $\mathcal{S}$, there can be multiple bases

- All bases for $\mathcal{S}$ have the same number of elements, i.e., if $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_\ell\}$ are bases for $\mathcal{S}$, then $k = \ell$

Dimension of a Subspace

The dimension of a subspace $\mathcal{S}$ with $\mathcal{S} \neq \{\mathbf{0}\}$, denoted by $\dim \mathcal{S}$, is the number of elements of any basis for $\mathcal{S}$

- $\dim\{\mathbf{0}\} = 0$
- represent effective degrees of freedom of the subspace

Examples:

- $\dim \mathbb{R}^m = m$
- If k is the number of maximal linearly independent vectors of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, then $\dim \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = k$

Dimension of a Subspace (Cont'd)

Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces

- If $\mathcal{S}_1 \subseteq \mathcal{S}_2$, then $\dim \mathcal{S}_1 \leq \dim \mathcal{S}_2$

- If $\mathcal{S}_1 \subseteq \mathcal{S}_2$ and $\dim \mathcal{S}_1 = \dim \mathcal{S}_2$, then $\mathcal{S}_1 = \mathcal{S}_2$

Suppose $\{b_1, \ldots, b_k\}$ is a basis of S_1. $\dim S_1 = k = \dim S_2$

$b_1, \ldots b_k \in S_1 \subseteq S_2$

$\Rightarrow \{b_1, \ldots, b_k\}$ is also a basis of S_2.

- $\dim \mathcal{S}_1 = m$ if and only if $\mathcal{S}_1 = \mathbb{R}^m$

$S_2 = \mathbb{R}^m$

- $\dim(\mathcal{S}_1 + \mathcal{S}_2) \leq \dim \mathcal{S}_1 + \dim \mathcal{S}_2$
    - $\dim(\mathcal{S}_1 + \mathcal{S}_2) = \dim \mathcal{S}_1 + \dim \mathcal{S}_2 - \dim(\mathcal{S}_1 \cap \mathcal{S}_2)$

$S_1 = S_2$

$= \operatorname{span}\{b_1, \ldots, b_k\}$

Rank

The rank of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, denoted by $\operatorname{rank}(\mathbf{A})$, is the number of elements of a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$

- Equivalently, $\operatorname{rank}(\mathbf{A})$ is the maximum number of linearly independent columns of $\mathbf{A}$
- $\dim \mathcal{R}(\mathbf{A}) = \operatorname{rank}(\mathbf{A})$

Facts:

- $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{A}^T)$, i.e., the rank of $\mathbf{A}$ is also the maximum number of linearly independent rows of $\mathbf{A}$
- $\operatorname{rank}(\mathbf{A} + \mathbf{B}) \leq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B})$
- $\operatorname{rank}(\mathbf{AB}) \leq \min\{\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})\}$
  - The equality holds when the columns of $\mathbf{A}$ are linearly independent and the rows of $\mathbf{B}$ are linearly independent

Rank (Cont'd)

- Matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to
  - have full column rank if all the columns of $\mathbf{A}$ are linearly independent
    - $\mathbf{A} \in \mathbb{R}^{m \times n}$ full column rank $\iff m \geq n, \mathrm{rank}(\mathbf{A}) = n$
  - have full row rank if all the rows of $\mathbf{A}$ are linearly independent
    - $\mathbf{A} \in \mathbb{R}^{m \times n}$ full row rank $\iff m \leq n, \mathrm{rank}(\mathbf{A}) = m$
  - have full rank if $\mathrm{rank}(\mathbf{A}) = \min\{m, n\}$, i.e., it has either full column rank or full row rank
  - be rank deficient if $\mathrm{rank}(\mathbf{A}) < \min\{m, n\}$

Invertible Matrices

A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be nonsingular or invertible if the columns of $\mathbf{A}$ are linearly independent, and singular or non-invertible otherwise

- Alternatively, $\mathbf{A}$ is singular if $\mathbf{A}\mathbf{x} = \mathbf{0}$ for some $\mathbf{x} \neq \mathbf{0}$

$$Ax = x_1 a_1 + \cdots + x_n a_n = 0$$

The inverse of an invertible $\mathbf{A}$, denoted by $\mathbf{A}^{-1}$, is a $n \times n$ square matrix satisfying

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Invertible Matrices (Cont'd)

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a nonsingular matrix

- $\mathbf{A}^{-1}$ always exists and is unique
- $\mathbf{A}^{-1}$ is nonsingular
- $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$
- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ are nonsingular
- $(\mathbf{A}^{T})^{-1} = (\mathbf{A}^{-1})^{T}$
    - As a shorthand notation, we denote $\mathbf{A}^{-T} = (\mathbf{A}^{T})^{-1}$