0. The Foundation of Mathematics

1. Linear algebra models

1.1 Vector basics

- Vector basics: A vector x can thus be defined as a collection of elements x1 , x2 , . . . , xn , arranged in a column or in a row.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

magnitude

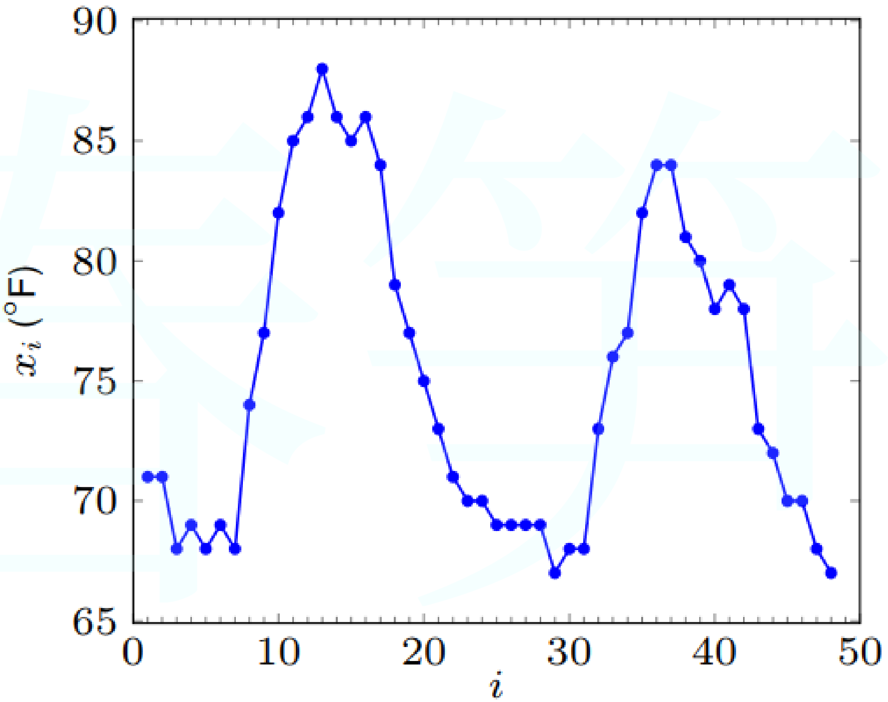
direction

use the notation $x = (x_1, \ldots, x_n)$ to denote a vector, $x \in \mathbb{R}^n$.

Signal or time series

elements of n-vector are values of some quantity at n different times

- hourly temperature over period of n hours

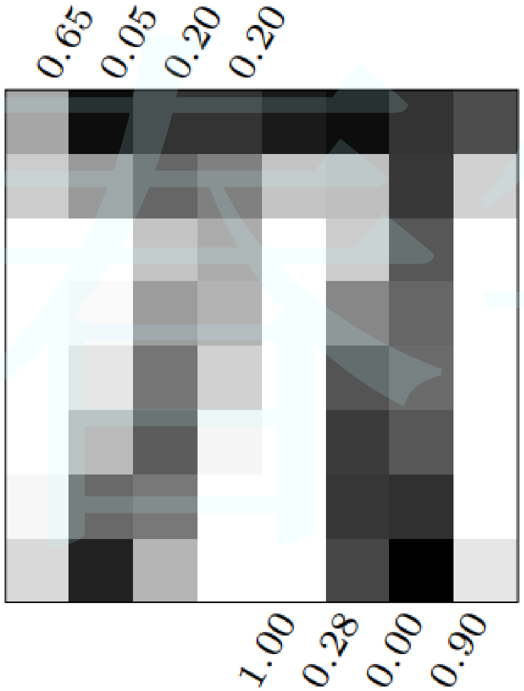


- daily return of a stock for period of n trading days
- cash flow: payments to an entity over n periods (*e.g.*, quarters)

Images, video

Monochrome (black and white) image

grayscale values of $M \times N$ pixels stored as MN-vector (*e.g.*, row-wise)



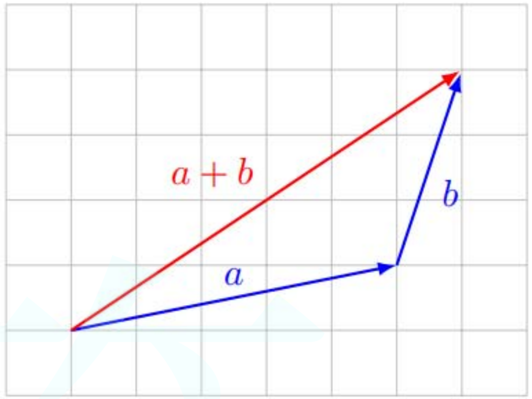
$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{62} \\ x_{63} \\ x_{64} \end{bmatrix} = \begin{bmatrix} 0.65 \\ 0.05 \\ 0.20 \\ \vdots \\ 0.28 \\ 0.00 \\ 0.90 \end{bmatrix}$$

Color image: $3MN$-vectors with R, G, B values of the MN pixels

Video: vector of size KMN represents K monochrome images of $M \times N$ pixels

The operations of sum, difference, and scalar multiplication

$$a + b = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix}, \qquad a - b = \begin{bmatrix} a_1 - b_1 \\ a_2 - b_2 \\ \vdots \\ a_n - b_n \end{bmatrix}$$



Scalar-vector multiplication: for scalar β and n-vector a,

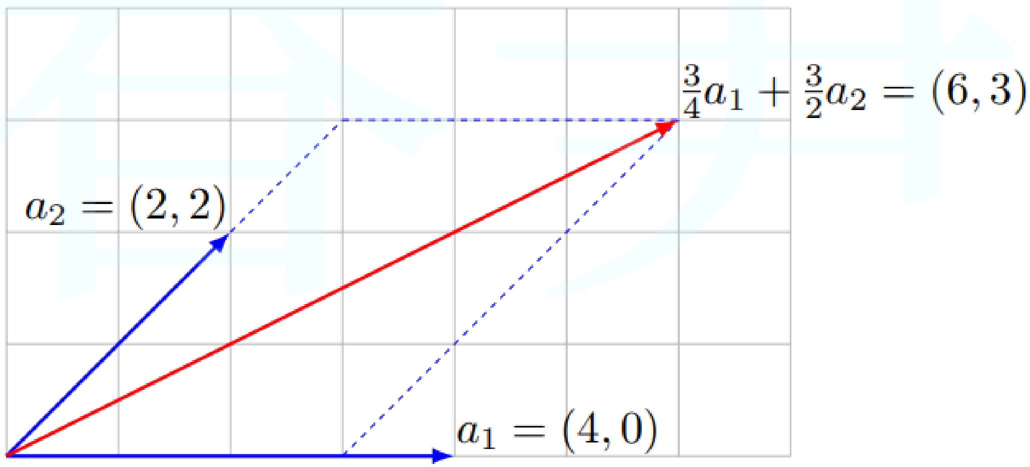
$$\beta \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \beta a_1 \\ \beta a_2 \\ \vdots \\ \beta a_n \end{bmatrix}$$

Linear combination

a *linear combination* of vectors $a_1, \ldots, a_m$ is a sum of scalar-vector products

$$\beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_m a_m$$

the scalars $\beta_1, \ldots, \beta_m$ are the *coefficients* of the linear combination



- Independence, bases, and dimensions。

A set $\{\phi_i\}_{i=1}^n$ is called a basis for $\mathbb{R}^n$ if the vectors in the set span $\mathbb{R}^n$ and are linearly independent.[2] This implies that each vector in the space has a unique representation as a linear combination of these basis vectors. Specifically, for any $x \in \mathbb{R}^n$, there exist (unique) coefficients $\{c_i\}_{i=1}^n$ such that

$$x = \sum_{i=1}^{n} c_i \phi_i.$$

$$x = \Phi c.$$

An important special case of a basis is an orthonormal basis, defined as a set of vectors $\{\phi_i\}_{i=1}^n$ satisfying

$$\langle \phi_i, \phi_j \rangle = \begin{cases} 1, & i = j; \\ 0, & i \neq j. \end{cases}$$

An orthonormal basis has the advantage that the coefficients c can be easily calculated as

$$c_i = \langle x, \phi_i \rangle,$$

or

$$c = \Phi^T x$$

Vector spaces： a vector space, X, is obtained by equipping vectors with the operations of addition and multiplication by a scalar.

$$x, y \in \mathcal{V} \Rightarrow \alpha x + \beta y \in \mathcal{V}.$$

Special vectors

Zero vector and ones vector

$$0 = (0, 0, \ldots, 0), \qquad \mathbf{1} = (1, 1, \ldots, 1)$$

size follows from context (if not, we add a subscript and write 0_n, $\mathbf{1}_n$)

Unit vectors

- there are n unit vectors of size n, written e_1, e_2, $\ldots$, e_n
- ith unit vector is zero except its ith element which is 1; for $n = 3$,

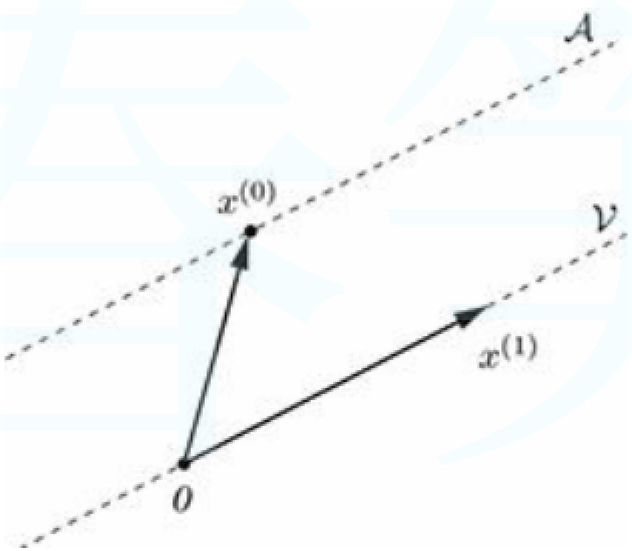
$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- size of e_i follows from context (or should be specified explicitly)

Affine sets.

$$\mathcal{A} = \{x \in \mathcal{X} : x = v + x^{(0)}, v \in \mathcal{V}\},$$

example:



几何解释?

1.2 Norms and inner products

Inner product

Inner product

Definition 2.2 *An* inner product *on a (real) vector space $\mathcal{X}$ is a real-valued function which maps any pair of elements $x, y \in \mathcal{X}$ into a scalar denoted by $\langle x, y\rangle$. The inner product satisfies the following axioms: for any $x, y, z \in \mathcal{X}$ and scalar α*

$$\langle x, x\rangle \geq 0;$$
$$\langle x, x\rangle = 0 \textit{ if and only if } x = 0;$$
$$\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle;$$
$$\langle \alpha x, y\rangle = \alpha\langle x, y\rangle;$$
$$\langle x, y\rangle = \langle y, x\rangle.$$

$$\langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i,$$

1.2 Norms and inner products

Definition 2.1 *A function from $\mathcal{X}$ to* R *is a* norm, *if*

$\|x\| \geq 0\ \forall x \in \mathcal{X}$, *and* $\|x\| = 0$ *if and only if* $x = 0$;

$\|x + y\| \leq \|x\| + \|y\|$, *for any* $x, y \in \mathcal{X}$ *(triangle inequality)*;

$\|\alpha x\| = |\alpha| \|x\|$, *for any scalar* α *and any* $x \in \mathcal{X}$.

Well-known examples of vector norm are as follows:

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|, \quad (l_\infty\text{-norm})$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|, \quad (l_1\text{-norm})$$

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2}, \quad (l_2\text{-norm}).$$

The above examples are particular cases of l_p-norm which is defined

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}, \ (l_p\text{-norm}).$$

$$
\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + 2\boldsymbol{x}^*\boldsymbol{y} + \|\boldsymbol{y}\|^2 \\
&\le \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 \\
&= (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2 .
\end{aligned}
$$

$$
\big|\, \|\boldsymbol{x}\| - \|\boldsymbol{y}\| \,\big| \le \|\boldsymbol{x}-\boldsymbol{y}\|
$$

$$
\|\boldsymbol{A}\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2} = \sqrt{\operatorname{tr}(\boldsymbol{A}\boldsymbol{A}^*)} = \sqrt{\operatorname{tr}(\boldsymbol{A}^*\boldsymbol{A})},
$$

$$
\left\| \int_a^b f(x)\, dx \right\| \le \int_a^b \|f(x)\|\, dx
$$

Norm balls. The set of all vectors with ℓ_p norm less than or equal to one,

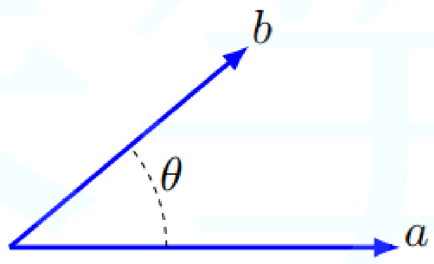
$$
\mathcal{B}_p = \{x \in \mathbb{R}^n : \|x\|_p \le 1\},
$$

Angle between vectors

the angle between nonzero real vectors a, b is defined as

$$\arccos\left(\frac{a^Tb}{\|a\|\,\|b\|}\right)$$

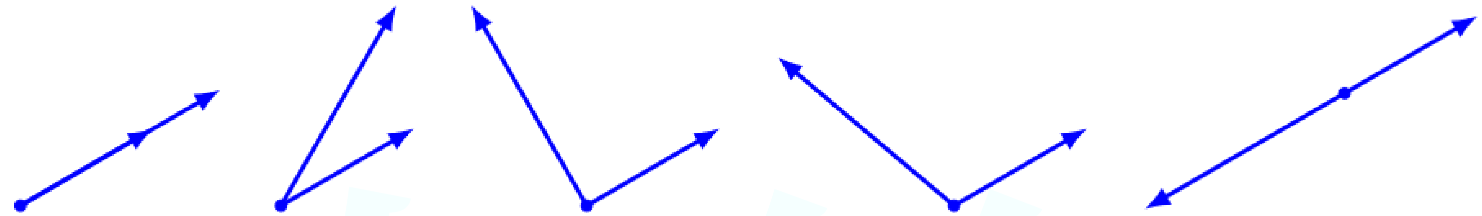
- this is the unique value of $\theta \in [0, \pi]$ that satisfies $a^Tb = \|a\|\|b\|\cos\theta$



- Cauchy-Schwarz inequality guarantees that

$$-1 \leq \frac{a^Tb}{\|a\|\,\|b\|} \leq 1$$

Terminology



$\theta = 0$	$a^T b = \|a\|\|b\|$	vectors are aligned or parallel
$0 \leq \theta < \pi/2$	$a^T b > 0$	vectors make an acute angle
$\theta = \pi/2$	$a^T b = 0$	vectors are orthogonal ($a \perp b$)
$\pi/2 < \theta \leq \pi$	$a^T b < 0$	vectors make an obtuse angle
$\theta = \pi$	$a^T b = -\|a\|\|b\|$	vectors are anti-aligned or opposed

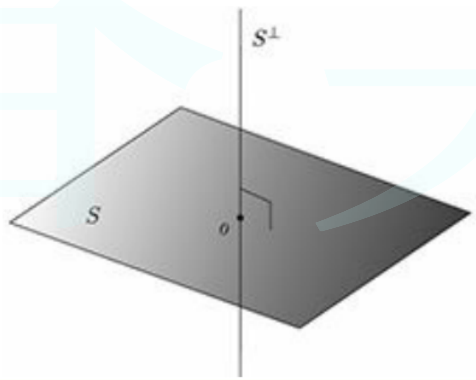
Orthonormal

Orthonormal vectors

$$\langle x^{(i)}, x^{(j)} \rangle = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

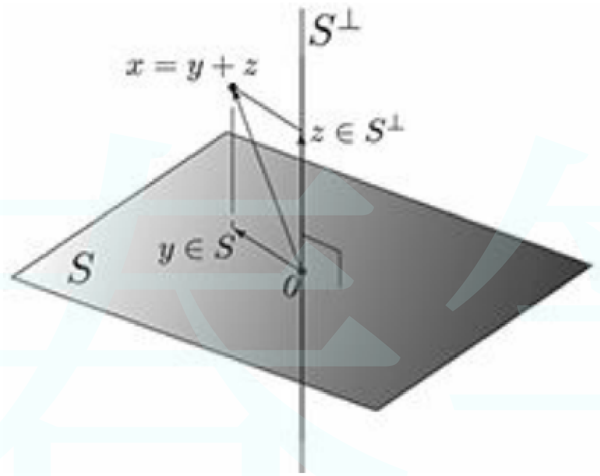
Orthogonal complement

2.2.3.3 Orthogonal complement. A vector $x \in \mathcal{X}$ is orthogonal to a subset $\mathcal{S}$ of an inner product space $\mathcal{X}$ if $x \perp s$ for all $s \in \mathcal{S}$. The set of vectors in $\mathcal{X}$ that are orthogonal to $\mathcal{S}$ is called the *orthogonal complement* of $\mathcal{S}$, and it is denoted by $\mathcal{S}^{\perp}$; see Figure 2.13.



Orthogonal decomposition

$\mathcal{X} = \mathcal{S} \oplus \mathcal{S}^{\perp}$ *for any subspace* $\mathcal{S} \subseteq \mathcal{X}$.

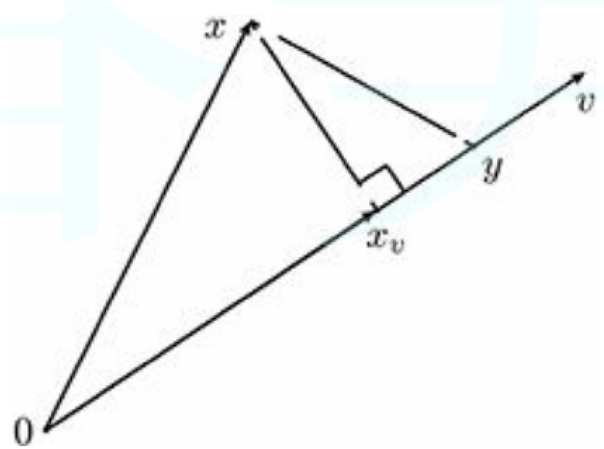


1.3 Projections

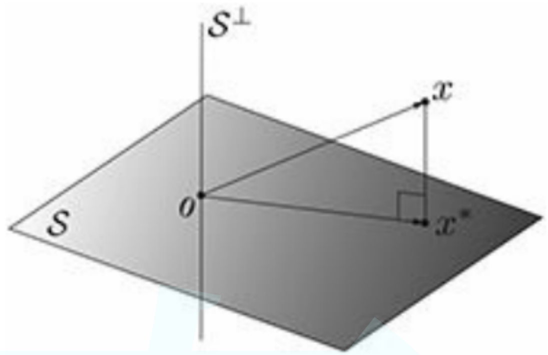
The idea of projection is central in optimization, and it corresponds to the problem of finding a point on a given set that is closest (in norm) to a given point. Formally, given a vector x in an inner product space $\mathcal{X}$ (say, e.g., $\mathcal{X} = \mathbb{R}^n$) and a closed set[9] $\mathcal{S} \subseteq \mathcal{X}$, the projection of x onto $\mathcal{S}$, denoted by $\Pi_{\mathcal{S}}(x)$, is defined as the point in $\mathcal{S}$ at minimal distance from x:

$$\Pi_{\mathcal{S}}(x) = \arg\min_{y \in \mathcal{S}} \|y - x\|,$$

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



$$(x - x_v) \perp v \iff \langle x - x_v, v \rangle = 0.$$



$$x^* \in \mathcal{S}, (x - x^*) \perp \mathcal{S}.$$

Proof Let $\mathcal{S}^\perp$ be the orthogonal subspace of $\mathcal{S}$, then, by virtue of Theorem 2.1, any vector $x \in \mathcal{X}$ can be written in a unique way as

$$x = u + z,\ u \in \mathcal{S},\ z \in \mathcal{S}^\perp.$$

Hence, for any vector y,

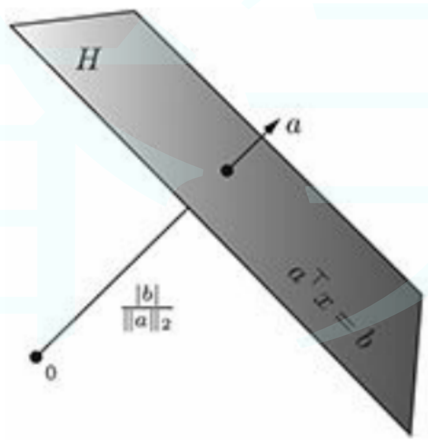
$$\|y - x\|^2 = \|(y - u) - z\|^2 = \|y - u\|^2 + \|z\|^2 - 2\langle y - u, z\rangle.$$

Hyperplanes and half-spaces

2.4.4.1 Hyperplanes. As defined in Section 2.3.2.2, a hyperplane is a set described by a single scalar product equality. Precisely, a hyperplane in $\mathbb{R}^n$ is a set of the form

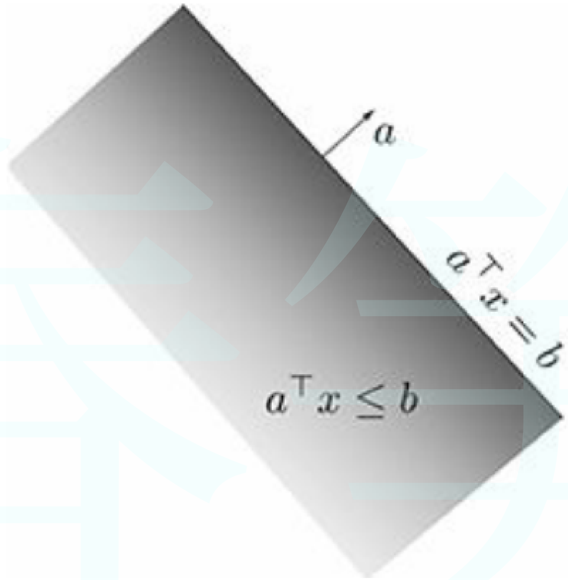
$$H = \{x \in \mathbb{R}^n : a^\top x = b\}, \tag{2.10}$$

where $a \in \mathbb{R}^n$, $a \neq 0$, and $b \in \mathbb{R}$ are given. Equivalently, we can think of hyperplanes as the level sets of linear functions, see Figure 2.23.



Half-spaces. A hyperplane H separates the whole space into two regions:

$$H_{-} \doteq \{x : a^{\top} x \leq b\}, H_{++} \doteq \{x : a^{\top} x > b\}.$$



2. Matrix Theory

Matrix

a rectangular array of numbers, for example

$$A = \begin{bmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{bmatrix}$$

- numbers in array are the *elements* (*entries*, *coefficients*, *components*)
- A_{ij} is the i, j element of A; i is its *row index*, j the *column index*
- *size* (*dimensions*) of the matrix is specified as (#rows) $\times$ (#columns)

  for example, A is a 3×4 matrix
- set of $m \times n$ matrices with real elements is written $\mathbf{R}^{m \times n}$
- set of $m \times n$ matrices with complex elements is written $\mathbf{C}^{m \times n}$

Symmetric and Hermitian matrices

Symmetric matrix: square with $A_{ij} = A_{ji}$

$$\begin{bmatrix} 4 & 3 & -2 \\ 3 & -1 & 5 \\ -2 & 5 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 4+3\mathrm{j} & 3-2\mathrm{j} & 0 \\ 3-2\mathrm{j} & -\mathrm{j} & -2\mathrm{j} \\ 0 & -2\mathrm{j} & 3 \end{bmatrix}$$

Hermitian matrix: square with $A_{ij} = \bar{A}_{ji}$ (complex conjugate of A_{ij})

$$\begin{bmatrix} 4 & 3-2\mathrm{j} & -1+\mathrm{j} \\ 3+2\mathrm{j} & -1 & 2\mathrm{j} \\ -1-\mathrm{j} & -2\mathrm{j} & 3 \end{bmatrix}$$

note: diagonal elements are real (since $A_{ii} = \bar{A}_{ii}$)

Structured matrices

matrices with special patterns or structure arise in many applications

- diagonal matrix: square with $A_{ij} = 0$ for $i \neq j$

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -5 \end{bmatrix}, \qquad \begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -5 \end{bmatrix}$$

- lower triangular matrix: square with $A_{ij} = 0$ for $i < j$

$$\begin{bmatrix} 4 & 0 & 0 \\ 3 & -1 & 0 \\ -1 & 5 & -2 \end{bmatrix}, \qquad \begin{bmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ -1 & 0 & -2 \end{bmatrix}$$

- upper triangular matrix: square with $A_{ij} = 0$ for $i > j$

Scalar-matrix multiplication and addition

Scalar-matrix multiplication:

scalar-matrix product of $m \times n$ matrix A with scalar β

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

A and β can be real or complex

Addition: sum of two $m \times n$ matrices A and B (real or complex)

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

Transpose

the *transpose* of an $m \times n$ matrix A is the $n \times m$ matrix

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

- $(A^T)^T = A$
- a symmetric matrix satisfies $A = A^T$
- A may be complex, but transpose of complex matrices is rarely needed
- transpose of matrix-scalar product and matrix sum

$$(\beta A)^T = \beta A^T, \qquad (A + B)^T = A^T + B^T$$

Matrix-vector product

product of $m \times n$ matrix A with n-vector (or $n \times 1$ matrix) x

$$Ax = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n \\ \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n \end{bmatrix}$$

- dimensions must be compatible: number of columns of A equals the size of x
- Ax is a linear combination of the columns of A:

$$Ax = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1a_1 + x_2a_2 + \cdots + x_na_n$$

each a_i is an m-vector (ith column of A)

Left and right inverse

$AB \neq BA$ in general, so we have to distinguish two types of inverses

Left inverse: X is a *left inverse* of A if

$$XA = I$$

A is *left-invertible* if it has at least one left inverse

Right inverse: X is a *right inverse* of A if

$$AX = I$$

A is *right-invertible* if it has at least one right inverse

Pseudo-inverse

- suppose $A \in \mathbf{R}^{m \times n}$ has linearly independent columns
- this implies that A is tall or square $(m \geq n)$; see page 4-13

the *pseudo-inverse* of A is defined as

$$A^\dagger = (A^T A)^{-1} A^T$$

- this matrix exists, because the Gram matrix $A^T A$ is nonsingular
- $A^\dagger$ is a left inverse of A:

$$A^\dagger A = (A^T A)^{-1} (A^T A) = I$$

(for complex A with linearly independent columns, $A^\dagger = (A^H A)^{-1} A^H$)

Matrix with orthonormal columns

$A \in \mathbf{R}^{m \times n}$ has orthonormal columns if its Gram matrix is the identity matrix:

$$
\begin{aligned}
A^T A &= \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}^T \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} \\
&= \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}
\end{aligned}
$$

Matrix-vector product

if $A \in \mathbf{R}^{m\times n}$ has orthonormal columns, then the linear function $f(x) = Ax$

- preserves inner products:

$$(Ax)^T(Ay) = x^T A^T Ay = x^T y$$

- preserves norms:

$$\|Ax\| = \left((Ax)^T(Ax)\right)^{1/2} = (x^T x)^{1/2} = \|x\|$$

- preserves distances: $\|Ax - Ay\| = \|x - y\|$

- preserves angles:

$$\angle(Ax, Ay) = \arccos\left(\frac{(Ax)^T(Ay)}{\|Ax\|\|Ay\|}\right) = \arccos\left(\frac{x^T y}{\|x\|\|y\|}\right) = \angle(x, y)$$

2. Matrix Theory

2.1Basic

$$
\begin{aligned}
(\mathbf{AB})^{-1} &= \mathbf{B}^{-1}\mathbf{A}^{-1} && (1)\\
(\mathbf{ABC}...)^{-1} &= ...\mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} && (2)\\
(\mathbf{A}^{T})^{-1} &= (\mathbf{A}^{-1})^{T} && (3)\\
(\mathbf{A}+\mathbf{B})^{T} &= \mathbf{A}^{T}+\mathbf{B}^{T} && (4)\\
(\mathbf{AB})^{T} &= \mathbf{B}^{T}\mathbf{A}^{T} && (5)\\
(\mathbf{ABC}...)^{T} &= ...\mathbf{C}^{T}\mathbf{B}^{T}\mathbf{A}^{T} && (6)\\
(\mathbf{A}^{H})^{-1} &= (\mathbf{A}^{-1})^{H} && (7)\\
(\mathbf{A}+\mathbf{B})^{H} &= \mathbf{A}^{H}+\mathbf{B}^{H} && (8)\\
(\mathbf{AB})^{H} &= \mathbf{B}^{H}\mathbf{A}^{H} && (9)\\
(\mathbf{ABC}...)^{H} &= ...\mathbf{C}^{H}\mathbf{B}^{H}\mathbf{A}^{H} && (10)
\end{aligned}
$$

2.2 Trace

$$
\begin{aligned}
\operatorname{Tr}(\mathbf{A}) &= \textstyle\sum_i A_{ii} && (11)\\
\operatorname{Tr}(\mathbf{A}) &= \textstyle\sum_i \lambda_i, \qquad \lambda_i = \operatorname{eig}(\mathbf{A}) && (12)\\
\operatorname{Tr}(\mathbf{A}) &= \operatorname{Tr}(\mathbf{A}^T) && (13)\\
\operatorname{Tr}(\mathbf{AB}) &= \operatorname{Tr}(\mathbf{BA}) && (14)\\
\operatorname{Tr}(\mathbf{A}+\mathbf{B}) &= \operatorname{Tr}(\mathbf{A}) + \operatorname{Tr}(\mathbf{B}) && (15)
\end{aligned}
$$

2.3 Derivatives

vector forms

$$\left[\frac{\partial \mathbf{x}}{\partial y}\right]_i = \frac{\partial x_i}{\partial y} \qquad \left[\frac{\partial x}{\partial \mathbf{y}}\right]_i = \frac{\partial x}{\partial y_i} \qquad \left[\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right]_{ij} = \frac{\partial x_i}{\partial y_j}$$

The following rules are general and very useful

$$\begin{aligned}
\partial \mathbf{A} &= 0 \qquad (\mathbf{A} \text{ is a constant}) && (29)\\
\partial(\alpha \mathbf{X}) &= \alpha \partial \mathbf{X} && (30)\\
\partial(\mathbf{X}+\mathbf{Y}) &= \partial \mathbf{X} + \partial \mathbf{Y} && (31)\\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial \mathbf{X}) && (32)\\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y}) && (33)\\
\partial(\mathbf{X} \circ \mathbf{Y}) &= (\partial \mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial \mathbf{Y}) && (34)\\
\partial(\mathbf{X} \otimes \mathbf{Y}) &= (\partial \mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes (\partial \mathbf{Y}) && (35)\\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial \mathbf{X})\mathbf{X}^{-1} && (36)\\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial \mathbf{X}) && (37)\\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial \mathbf{X}) && (38)\\
\partial \mathbf{X}^T &= (\partial \mathbf{X})^T && (39)\\
\partial \mathbf{X}^H &= (\partial \mathbf{X})^H && (40)
\end{aligned}$$

Derivatives of an Inverse

$$\frac{\partial \mathbf{Y}^{-1}}{\partial x} = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\mathbf{Y}^{-1} \tag{53}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^{-1}\mathbf{b}}{\partial \mathbf{X}} = -\mathbf{X}^{-T}\mathbf{a}\mathbf{b}^T\mathbf{X}^{-T} \tag{55}$$

Derivatives of Matrices, Vectors and Scalar Forms

$$\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a} \tag{61}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T \tag{62}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T \tag{63}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{a}}{\partial \mathbf{X}} = \frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{a}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{a}^T \tag{64}$$

$$\frac{\partial \mathbf{X}}{\partial X_{ij}} = \mathbf{J}^{ij} \tag{65}$$

Assume $\mathbf{W}$ is symmetric, then

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) &= -2\mathbf{A}^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) && (76)\\
\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s}) &= 2\mathbf{W}(\mathbf{x}-\mathbf{s}) && (77)\\
\frac{\partial}{\partial \mathbf{s}}(\mathbf{x}-\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{s}) &= -2\mathbf{W}(\mathbf{x}-\mathbf{s}) && (78)\\
\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) &= 2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) && (79)\\
\frac{\partial}{\partial \mathbf{A}}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s}) &= -2\mathbf{W}(\mathbf{x}-\mathbf{A}\mathbf{s})\mathbf{s}^T && (80)
\end{aligned}$$

2.4 Singular Value Decomposition

Any $n \times m$ matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{UDV}^T,$$

where

$$\begin{array}{lcll} \mathbf{U} & = & \text{eigenvectors of } \mathbf{AA}^T & n \times n \\ \mathbf{D} & = & \sqrt{\text{diag}(\text{eig}(\mathbf{AA}^T))} & n \times m \\ \mathbf{V} & = & \text{eigenvectors of } \mathbf{A}^T\mathbf{A} & m \times m \end{array}$$

2.5 Symmetric Square decomposed into squares

Assume $\mathbf{A}$ to be $n \times n$ and symmetric. Then

$$\left[\ \mathbf{A}\ \right] = \left[\ \mathbf{V}\ \right]\left[\ \mathbf{D}\ \right]\left[\ \mathbf{V}^T\ \right], \tag{271}$$

where $\mathbf{D}$ is diagonal with the eigenvalues of $\mathbf{A}$, and $\mathbf{V}$ is orthogonal and the eigenvectors of $\mathbf{A}$.

2.6 Square decomposed into squares

Assume $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then

$$\left[\ \mathbf{A}\ \right] = \left[\ \mathbf{V}\ \right] \left[\ \mathbf{D}\ \right] \left[\ \mathbf{U}^T\ \right], \tag{272}$$

where $\mathbf{D}$ is diagonal with the square root of the eigenvalues of $\mathbf{A}\mathbf{A}^T$, $\mathbf{V}$ is the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{U}^T$ is the eigenvectors of $\mathbf{A}^T\mathbf{A}$.

2.7 Cholesky-decomposition

Assume $\mathbf{A}$ is a symmetric positive definite square matrix, then

$$\mathbf{A} = \mathbf{U}^T\mathbf{U} = \mathbf{L}\mathbf{L}^T, \tag{278}$$

where $\mathbf{U}$ is a unique upper triangular matrix and $\mathbf{L}$ is a unique lower triangular matrix.

2.8 Positive Definite and Semi-definite Matrices

A matrix $\mathbf{A}$ is positive definite if and only if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > \mathbf{0}, \quad \forall \mathbf{x} \neq \mathbf{0}$$

A matrix $\mathbf{A}$ is positive semi-definite if and only if

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq \mathbf{0}, \quad \forall \mathbf{x}$$

Note that if $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is also positive semi-definite.

Eigenvalues

The following holds with respect to the eigenvalues:

$$\begin{aligned} \mathbf{A} \text{ pos. def.} \quad &\Leftrightarrow \quad \mathrm{eig}\left(\tfrac{\mathbf{A}+\mathbf{A}^H}{2}\right) > 0 \\ \mathbf{A} \text{ pos. semi-def.} \quad &\Leftrightarrow \quad \mathrm{eig}\left(\tfrac{\mathbf{A}+\mathbf{A}^H}{2}\right) \geq 0 \end{aligned}$$

Trace

The following holds with respect to the trace:

$$\begin{aligned} \mathbf{A} \text{ pos. def.} \quad &\Rightarrow \quad \mathrm{Tr}(\mathbf{A}) > 0 \\ \mathbf{A} \text{ pos. semi-def.} \quad &\Rightarrow \quad \mathrm{Tr}(\mathbf{A}) \geq 0 \end{aligned}$$

Inverse

If $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is invertible and $\mathbf{A}^{-1}$ is also positive definite.

Diagonal

If $\mathbf{A}$ is positive definite, then $A_{ii} > 0, \forall i$

Decomposition I

The matrix $\mathbf{A}$ is positive semi-definite of rank r $\Leftrightarrow$ there exists a matrix $\mathbf{B}$ of rank r such that $\mathbf{A} = \mathbf{B}\mathbf{B}^T$

The matrix $\mathbf{A}$ is positive definite $\Leftrightarrow$ there exists an invertible matrix $\mathbf{B}$ such that $\mathbf{A} = \mathbf{B}\mathbf{B}^T$