1. Vectors and matrices

- vectors
- vector operations
- inner product and norm
- matrices
- matrix operations
- functions
- linear equations

Vector

a (column) *vector* is an ordered list of numbers arranged in a vertical array, written as:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \quad \text{or} \quad a = (a_1, a_2, \ldots, a_n)$$

- a_i is the ith *entry* (*element, coefficient, component*) of vector a
- i is the *index* of the ith entry a_i
- number of entries n is the *size* (*length, dimension*) of the vector
- a vector of size n is called an n-*vector*

the **transpose** of an n-vector a is a *row* vector arranged in a horizontal array:

$$a^T = [a_1 \;\; a_2 \;\; \cdots \;\; a_n]$$

- $(\cdot)^T$ is transpose operation
- $(a^T)^T = a$ (transpose of row vector is a column vector)

Notes and conventions

- all vectors are column vectors unless otherwise stated
  - for row vector we use the transpose notation (*e.g.*, a^T)
- $\mathbb{R}^n$ is set of n-vectors with real entries
- $a \in \mathbb{R}^n$ means a is n-vector with real entries
- two n-vectors a and b are equal, denoted as $a = b$, if $a_i = b_i$ for all i
- a_i can refer to an ith vector in a collection of vectors
  - in this case, we use $(a_i)_j$ to denote the jth entry of vector a_i
  - example: if $a_2 = (-1, 2, -5)$, then $(a_2)_3 = -5$

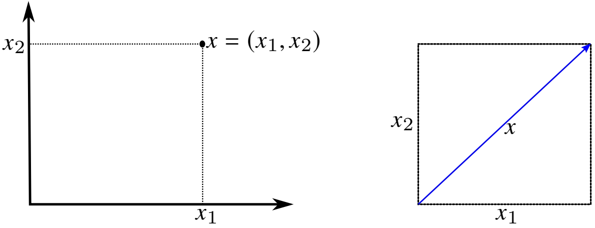
Conventions

- parentheses are also used instead of rectangular brackets to represent a vector
- other notations exist to distinguish vectors from numbers (*e.g.*, $\boldsymbol{a}, \vec{a}, \mathbf{a}$)
- conventions vary; be prepared to distinguish scalars from vectors

Examples of vectors

Location and displacement

- location (position): coordinates of a point in 2-D (plane) or 3-D space
- displacement: vector represents the change in position from one point to another (shown as an arrow in plane or 3-D space)



Examples of vectors

Time series or signal: entries are values of some quantity at n different times

- hourly temperature over a period of n hours
- audio signal: entries give the acoustic pressure values at equally spaced times

Feature vector: entries are quantities that relate to a single object

- example: age, height, weight, blood pressure, gender, etc., of patients
- entries are called the *features* or *attributes*

Portfolio: entries can represent stock portfolio (*e.g.*, investment in n assets)

- ith entry is the number of shares of asset i held (or invested in asset i)
- entries can be the no. of shares, dollar values, fractions of total dollar amount
- shares you owe another party (short positions) are represented by negative values

Special vectors

Zero vector and ones vector

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

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, denoted by $e_1, e_2, \ldots, e_n$

$$(e_i)_j = \begin{cases} 1 & j = i \\ 0 & j \neq i \end{cases}$$

- the ith unit vector is zero except its ith entry which is 1
- example: for $n = 3$,

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

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

Block vectors, subvectors

Stacking

- vectors can be *stacked* (*concatenated*) to create larger vectors
- stacking vectors b, c, d of size m, n, p gives an $(m+n+p)$-vector

$$a = \begin{bmatrix} b \\ c \\ d \end{bmatrix} = (b, c, d) = (b_1, \ldots, b_m, c_1, \ldots, c_n, d_1, \ldots, d_p)$$

- we call b, c, and d as *subvectors* or *slices* of a
- example: if $a = 1$, $b = (2, -1)$, $c = (4, 2, 7)$, then $(a, b, c) = (1, 2, -1, 4, 2, 7)$

Subvectors slicing

- colon (:) notation is used to define subvectors (slices) of a vector
- for vector a, we define $a_{r:s} = (a_r, \ldots, a_s)$
- example: if $a = (1, -1, 2, 0, 3)$, then $a_{2:4} = (-1, 2, 0)$

Outline

- vectors
- **vector operations**
- inner product and norm
- matrices
- matrix operations
- functions
- linear equations

Addition and subtraction

for n-vectors a and b,

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

Example

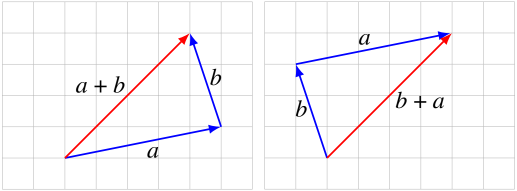
$$\begin{bmatrix} 0 \\ 7 \\ 3 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 9 \\ 3 \end{bmatrix}$$

Properties: for vectors a, b of equal size

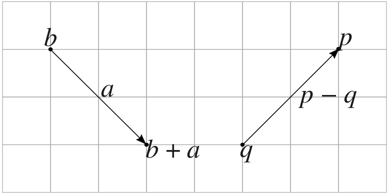
- commutative: $a+b=b+a$
- associative: $a+(b+c)=(a+b)+c$

Geometric interpretation: displacements addition

- if a and b are displacements, $a + b$ is the net displacement



- position displacements



Scalar-vector multiplication

for scalar β and n-vector a,

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

example:

$$(-2) \begin{bmatrix} 1 \\ 9 \\ 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -18 \\ -12 \end{bmatrix}$$

Properties: for vectors a, b of equal size, scalars β, γ

- commutative: $\beta a = a\beta$
- associative: $(\beta\gamma)a = \beta(\gamma a)$, we write as $\beta\gamma a$
- distributive with scalar addition: $(\beta + \gamma)a = \beta a + \gamma a$
- distributive with vector addition: $\beta(a + b) = \beta a + \beta b$

Linear combination

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

$$\beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_k a_k$$

- scalars $\beta_1, \ldots, \beta_k$ are the *coefficients* of the linear combination
- example: any n-vector b can be written as

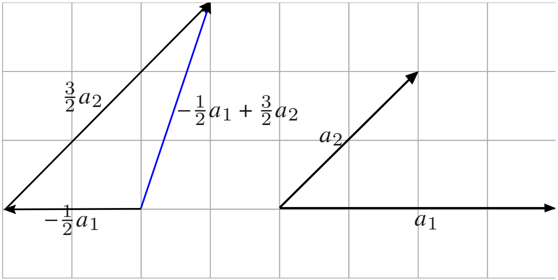
$$b = b_1 e_1 + \cdots + b_n e_n$$

Special linear combinations

- *affine combination*: when $\beta_1 + \cdots + \beta_k = 1$
- *convex combination* or *weighted average*: when $\beta_1 + \cdots + \beta_k = 1$ and $\beta_i \geq 0$

Example: combination of displacements

$$b = -0.5a_1 + 1.5a_2$$

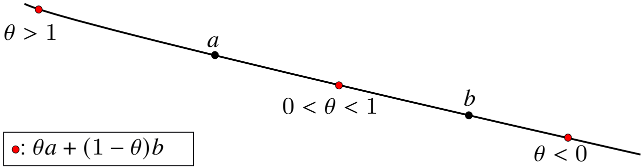


Line segment

any point on the line passing through distinct a and b can be written as

$$c = \theta a + (1 - \theta) b$$

- θ is a scalar
- for $0 \leq \theta \leq 1$, point c lie on the segment between a and b



Outline

- vectors
- vector operations
- **inner product and norm**
- matrices
- matrix operations
- functions
- linear equations

Inner product

the (Euclidean) *inner product* (or *dot product*) of two n-vectors a, b is

$$a^T b = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

- a scalar
- other notation exists: $\langle a, b \rangle$, $\langle a \mid b \rangle$, $a \cdot b$
- example:

$$\begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}^T \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = (-1)(1) + (2)(0) + (2)(-3) = -7$$

Properties of inner product

for vectors a, b, c of equal size, scalar γ

- nonnegativity: $a^T a \geq 0$, and $a^T a = 0$ if and only if $a = 0$
- commutative: $a^T b = b^T a$
- associative with scalar multiplication: $(\gamma a)^T b = \gamma(a^T b)$
- distributive with vector addition: $(a + b)^T c = a^T c + b^T c$

Useful combination: for vectors a, b, c, d

$$(a + b)^T (c + d) = a^T c + a^T d + b^T c + b^T d$$

Block vectors: if vectors a, b are block vectors, and corresponding blocks $a_i, b_i \in \mathbb{R}^{n_i}$ have the same sizes (they conform),

$$a^T b = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}^T \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} = a_1^T b_1 + \cdots + a_k^T b_k$$

Simple examples

Inner product with unit vector

$$e_i^T a = a_i$$

Differencing

$$(e_i - e_j)^T a = a_i - a_j$$

Sum and average

$$\mathbf{1}^T a = a_1 + a_2 + \cdots + a_n$$

$$\mathrm{avg}(a) = \frac{a_1 + a_2 + \cdots + a_n}{n} = \left(\tfrac{1}{n}\mathbf{1}\right)^T a$$

Inner product examples

Polynomial evaluation

- n-vector c represents the coefficients of a polynomial p of degree $n-1$ or less:

$$p(x) = c_1 + c_2 x + \cdots + c_{n-1} x^{n-2} + c_n x^{n-1}$$

- t is number, and let $z = (1, t, t^2, \ldots, t^{n-1})$ be the n-vector of powers of t
- $c^T z = p(t)$ is the value of the polynomial p at the point t

Price quantity (cost)

- vectors of prices p and quantities q of n goods
- $p^T q = p_1 q_1 + p_2 q_2 + \cdots + p_n q_n$ is the total cost

Portfolio value

- s is an n-vector of holdings in shares of a portfolio of n assets
- p is an n-vector for the prices of the assets
- $p^T s$ is the total (or net) value of the portfolio

Euclidean norm

Euclidean norm of vector $a \in \mathbb{R}^n$:

$$\|a\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} = \sqrt{a^T a}$$

- reduces to absolute value $|a| = \max\{a, -a\}$ when $n = 1$
- measures the magnitude of a
- examples

$$\left\| \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} \right\| = \sqrt{9} = 3, \quad \left\| \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\| = 1$$

Properties

Positive definiteness

$$\|a\| \geq 0 \quad \text{for all } a, \quad \|a\| = 0 \quad \text{only if } a = 0$$

Homogeneity

$$\|\beta a\| = |\beta| \|a\| \quad \text{for all vectors } a \text{ and scalars } \beta$$

Triangle inequality

$$\|a + b\| \leq \|a\| + \|b\| \text{ for all vectors } a \text{ and } b \text{ of equal length}$$

- any real function that satisfies these properties is called a (general) *norm* (we will see other norms)
- Euclidean norm is often written as $\|a\|_2$ to distinguish from other norms

Norm of block vector and norm of sum

Norm of block vector: for vectors a, b, c,

$$\left\| \begin{bmatrix} a \\ b \\ c \end{bmatrix} \right\| = \sqrt{\|a\|^2 + \|b\|^2 + \|c\|^2}$$

Norm of sum: for vectors a, b,

$$\|a + b\| = \sqrt{\|a\|^2 + 2a^T b + \|b\|}$$

Cauchy-Schwarz inequality

$$|a^T b| \leq \|a\| \|b\| \quad \text{for all } a, b \in \mathbb{R}^n$$

moreover, equality $|a^T b| = \|a\| \|b\|$ holds if:

- $a = 0$ or $b = 0$; in this case $a^T b = 0 = \|a\| \|b\|$
- $b = \gamma a$ for some $\gamma > 0$; in this case

$$0 < a^T b = \gamma \|a\|^2 = \|a\| \|b\|$$

- $b = -\gamma a$ for some $\gamma > 0$; in this case

$$0 > a^T b = -\gamma \|a\|^2 = -\|a\| \|b\|$$

Proof of Cauchy-Schwarz inequality

1. trivial if $a = 0$ or $b = 0$

2. assume $\|a\| = \|b\| = 1$; we show that $-1 \leq a^T b \leq 1$

$$\begin{aligned} 0 &\leq \|a - b\|^2 \\ &= (a-b)^T(a-b) \\ &= \|a\|^2 - 2a^Tb + \|b\|^2 \\ &= 2(1 - a^Tb) \end{aligned}$$

with equality only if $a = b$

$$\begin{aligned} 0 &\leq \|a + b\|^2 \\ &= (a+b)^T(a+b) \\ &= \|a\|^2 + 2a^Tb + \|b\|^2 \\ &= 2(1 + a^Tb) \end{aligned}$$

with equality only if $a = -b$

3. for general nonzero a, b, apply case 2 to the unit-norm vectors

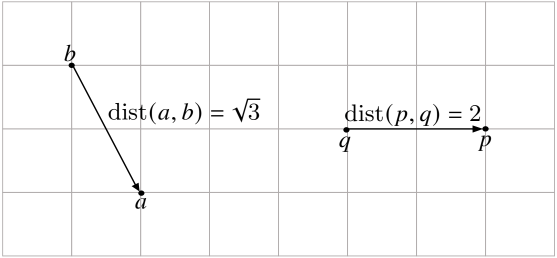
$$\frac{1}{\|a\|}a, \quad \frac{1}{\|b\|}b$$

Euclidean distance

Euclidean distance between two vectors a and b,

$$\mathrm{dist}(a, b) = \|a - b\|$$

- agrees with ordinary distance for $n = 1, 2, 3$

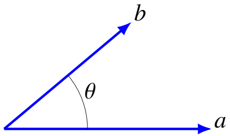


- when the distance between two vectors is small, we say they are ‘close’ or ‘nearby’, and when the distance is large, we say they are ‘far’

Angle between vectors

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

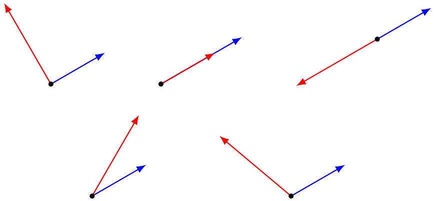
$$\theta = \angle(a, b) = \arccos\left(\frac{a^T b}{\|a\|\|b\|}\right)$$



- this is the unique value of $\theta \in [0, \pi]$ that satisfies $a^T b = \|a\|\|b\| \cos\theta$
- coincides with ordinary angle between vectors in 2-D and 3-D
- symmetric: $\angle(a, b) = \angle(b, a)$
- unaffected by positive scaling: $\angle(\beta a, \gamma b) = \angle(a, b)$ for $\beta, \gamma > 0$

Classification of angles

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



Orthonormal vectors

set of vectors $a_1, a_2, \ldots, a_k$ is *orthonormal* if:

$$a_i^T a_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

- vectors are mutually orthogonal and have unit norm
- vector of norm one is called *normalized*
- process of dividing a vector by its norm is known as *normalizing*

Examples

- standard unit vectors $e_1, \ldots, e_n$ are orthonormal
- vectors

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

  are orthonormal

Outline

- vectors
- vector operations
- inner product and norm
- **matrices**
- matrix operations
- functions
- linear equations

Matrices

a *matrix* is an ordered rectangular array of numbers, written as

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

- scalars in array are the *entries* (*elements*, *coefficients*, *components*)
- a_{ij} is the i, jth entry of A (i is row index, j is column index)
- *size* (*dimensions*) of the matrix is $m \times n$ = (#rows) $\times$ (#columns)

Example

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

- $a_{23} = -0.1$
- a 3×4 matrix

Notes and conventions

Notes

- a matrix of size $m \times n$ is called an $m \times n$-matrix
- $\mathbb{R}^{m\times n}$ is set of $m \times n$ matrices with real entries
- we use $a_{i,j}$ when i or j are more than one digit
- two matrices with same size are equal if corresponding entries are all equal
- sometimes A_k is a matrix; in this case, we use $(A_k)_{ij}$ to denote its i, j entry

Conventions

- matrices are typically denoted by capital letters
- parentheses are also used instead of rectangular brackets to represent a matrix
- sometimes A_{ij} is used to denote the i, jth entry of A
- some authors use bold capital letter for matrices (*e.g.*, $\mathbf{A}$, $\boldsymbol{A}$)
- be prepared to figure out whether a symbol represents a matrix, vector, or a scalar

Matrix examples

Images

- $m \times n$ matrix denote a monochrome (black and white) image
- x_{ij} is i, j pixel value in a monochrome image

Multiple asset returns

- $T \times n$ matrix R gives the returns of n assets over T periods
- r_{ij} is return of asset j in period i
- jth column of R is a T-vector that is the return time series for asset j

Feature matrix

- $X = [x_1 \ \cdots \ x_N]$ is $n \times N$ feature matrix
- column x_j is feature n-vector for object or example j
- x_{ij} is value of feature i for example j

Matrix shapes

Scalar: a 1×1 matrix is a scalar

Row and column vectors

- a $1 \times n$ matrix is called a row vector
- an $n \times 1$ matrix is called a column vector (or just vector)

Tall, wide, square matrices: an $m \times n$ matrix is

- tall, skinny, or thin if $m > n$
- wide or fat if $m < n$
- square if $m = n$

Transpose of a matrix

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

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

- $(A^T)_{ij} = a_{ji}$
- $(A^T)^T = A$
- example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

Columns and rows

an $m \times n$ matrix can be viewed as a matrix with row/column vectors

Columns representation

$$A = [a_1 \ a_2 \ \cdots \ a_n], \qquad a_j = \begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}$$

each a_j is an m-vector (the jth column of A)

Rows representation

$$A = \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_m^T \end{bmatrix}, \qquad b_i^T = [a_{i1} \ \cdots \ a_{in}]$$

each b_i^T is a $1 \times n$ row vector (the ith row of A)

Block matrix and submatrices

- a *block* matrix is a rectangular array of matrices
- entries in the array are the *blocks* or *submatrices* of the block matrix

Example: a 2×2 block matrix

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

- submatrices can be referred to by their block row and column (C is $1, 2$ block of A)
- dimensions of the blocks must be compatible
- if the blocks are

$$B = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 2 & 3 \\ 5 & 4 & 7 \end{bmatrix}, \quad D = \begin{bmatrix} 1 \end{bmatrix}, \quad E = \begin{bmatrix} -1 & 6 & 0 \end{bmatrix}$$

  then

$$A = \begin{bmatrix} 2 & 0 & 2 & 3 \\ 1 & 5 & 4 & 7 \\ 1 & -1 & 6 & 0 \end{bmatrix}$$

Slice of matrix

$$A_{p:q,r:s} = \begin{bmatrix} a_{pr} & a_{p,r+1} & \cdots & a_{ps} \\ a_{p+1,r} & a_{p+1,r+1} & \cdots & a_{p+1,s} \\ \vdots & \vdots & & \vdots \\ a_{qr} & a_{q,r+1} & \cdots & a_{qs} \end{bmatrix}$$

- an $(q-p+1)\times(s-r+1)$ matrix
- obtained by extracting from A entries in rows p to q and columns r to s
- from last page example, we have

$$A_{2:3,3:4} = \begin{bmatrix} 4 & 7 \\ 6 & 0 \end{bmatrix}$$

Transpose of block matrix

the transpose of a block matrix (shown for a 2×2 block matrix)

$$\left[\begin{array}{cc} A & B \\ C & D \end{array}\right]^T = \left[\begin{array}{cc} A^T & C^T \\ B^T & D^T \end{array}\right]$$

- A, B, C, and D are matrices with compatible sizes
- concept holds for any number of blocks

Special matrices

Zero matrix

- matrix with $a_{ij} = 0$ for all i, j
- notation: 0 or $0_{m \times n}$ (if dimension is not clear from context)

Identity matrix

- square matrix with $a_{ij} = 1$ if $i = j$ and $a_{ij} = 0$ if $i \neq j$
- notation: I or I_n (if dimension is not clear from context)
- columns of I_n are unit vectors $e_1, e_2, \ldots, e_n$; for example,

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [e_1 \;\; e_2 \;\; e_3]$$

Structured matrices

matrices with special patterns or structure arise in many applications

Diagonal matrix

- square with $a_{ij} = 0$ for $i \neq j$
- represented as $A = \mathrm{diag}(a_1, \ldots, a_n)$ where a_i are diagonal entries

$$\mathrm{diag}(0.2, -3, 1.2) = \begin{bmatrix} 0.2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 1.2 \end{bmatrix}$$

Lower triangular matrix: square with $a_{ij} = 0$ for $i < j$

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

Upper triangular matrix: square with $a_{ij} = 0$ for $i > j$

(a triangular matrix is **unit** upper/lower triangular if $a_{ii} = 1$ for all i)

Symmetric matrices

a square matrix is *symmetric* if

$$A = A^T$$

- $a_{ij} = a_{ji}$
- examples:

$$\begin{bmatrix} 3 & 7 & -2 \\ 7 & -1 & 5 \\ -2 & 5 & 0 \end{bmatrix}$$

Outline

- vectors
- vector operations
- inner product and norm
- matrices
- **matrix operations**
- functions
- linear equations

Matrix addition

sum of two $m \times n$ matrices A and B

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

Properties

- *commutativity:* $A + B = B + A$
- *associativity:* $(A + B) + C = A + (B + C)$
- *addition with zero matrix:* $A + 0 = 0 + A = A$
- *transpose of sum:* $(A + B)^T = A^T + B^T$

Scalar-matrix multiplication

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

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

Properties: for matrices A, B, scalars β, γ

- associativity: $(\beta\gamma)A = \beta(\gamma A)$
- *distributivity:* $(\beta + \gamma)A = \beta A + \gamma A$ and $\gamma(A + B) = \gamma A + \gamma B$
- *transposition:* $(\beta A)^T = \beta A^T$

Matrix-vector product

product of $m \times n$ matrix A with n-vector x

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1^T x \\ b_2^T x \\ \vdots \\ b_m^T x \end{bmatrix}$$

- b_i^T is ith row of A
- dimensions must be compatible (number of columns of A equals the size of x)
- Ax is a linear combination of the columns of A:

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

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

Properties of matrix-vector multiplication

for matrices A, B, vectors x, y and scalar β

- *associativity:* $(\beta A)x = A(\beta x) = \beta(Ax)$ (we write βAx)
- *distributivity:* $A(x + y) = Ax + Ay$ and $(A + B)x = Ax + Bx$

General examples

- $0x = 0$, *i.e.*, multiplying by zero matrix gives zero
- $Ix = x$, *i.e.*, multiplying by identity matrix does nothing
- inner product $a^T b$ is matrix-vector product of $1 \times n$ matrix a^T and n-vector b
- $Ae_j = a_j$, the jth column of A [$A^T e_i = b_i$ where b_i^T is ith row]
- the product $A\mathbf{1}$ is the sum of the columns of A
- for the $n \times n$ matrix

$$A = \begin{bmatrix} 1 - 1/n & -1/n & \cdots & -1/n \\ -1/n & 1 - 1/n & \cdots & -1/n \\ \vdots & & \cdots & \vdots \\ -1/n & -1/n & \cdots & 1 - 1/n \end{bmatrix},$$

  $\tilde{x} = Ax$ is de-meaned version of x (*i.e.*, $\tilde{x} = x - \mathrm{avg}(x)\mathbf{1}$)

Difference matrix

$(n-1) \times n$ *difference matrix* is

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ & & \ddots & \ddots & & & \\ & & & \ddots & \ddots & & \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

$y = Dx$ is $(n-1)$-vector of differences of consecutive entries of x:

$$Dx = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_n - x_{n-1} \end{bmatrix}$$

Vandermonde matrix

consider a polynomial of degree $n-1$ or less with coefficients $x_1, x_2, \ldots, x_n$:

$$p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

- values of $p(t)$ at m points $t_1, \ldots, t_m$ can be written as

$$\begin{bmatrix} p(t_1) \\ p(t_2) \\ \vdots \\ p(t_m) \end{bmatrix} = \begin{bmatrix} 1 & t_1 & \cdots & t_1^{n-1} \\ 1 & t_2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & t_m & \cdots & t_m^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = Ax$$

- the matrix A is called a *Vandermonde matrix*
- the product Ax maps coefficients of polynomial to function values

Matrix multiplication

product of $m \times n$ matrix A and $n \times p$ matrix B

$$C = AB$$

is the $m \times p$ matrix with i, j entry

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{in} b_{nj}$$

- to get c_{ij} : move along ith row of A, jth column of B
- dimensions must be compatible:

  #columns in A = #rows in B

- example:

$$\begin{bmatrix} -1.5 & 3 & 2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & -2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3.5 & -4.5 \\ -1 & 1 \end{bmatrix}$$

Special cases of matrix multiplication

- scalar-vector product (with scalar on right!) $x\alpha$
- inner product $a^T b$
- matrix-vector multiplication Ax
- outer product of m-vector a and n-vector b is the $m \times n$ matrix

$$ab^T = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & & \vdots \\ a_m b_1 & a_m b_2 & \cdots & a_m b_n \end{bmatrix}$$

- multiplication by identity $AI_n = A$ and $I_m A = A$
- matrix power: multiplication of matrix with itself p times: $A^p = AA \cdots A$

Properties of matrix-matrix product

- associativity: $(AB)C = A(BC)$ so we write ABC
- associativity: with scalar multiplication: $(\gamma A)B = \gamma(AB) = \gamma AB$
- distributivity with sum:

$$A(B + C) = AB + AC, \quad (A + B)C = AC + BC$$

- transpose of product:

$$(AB)^T = B^T A^T \quad \text{and} \quad (Ax)^T = x^T A^T$$

- **not** commutative: $AB \neq BA$ in general; for example,

$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

  there are exceptions, e.g., $AI = IA$ for square A

Product of block matrices

block-matrices can be multiplied as regular matrices

Example: product of two 2×2 block matrices

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} W & Y \\ X & Z \end{bmatrix} = \begin{bmatrix} AW + BX & AY + BZ \\ CW + DX & CY + DZ \end{bmatrix}$$

if the dimensions of the blocks are compatible

Column and row representations

Column representation

- A is $m \times n$, B is $n \times p$ with columns b_i

$$AB = A \begin{bmatrix} b_1 & b_2 & \cdots & b_p \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_p \end{bmatrix}$$

- so AB is 'batch' multiply of A times columns of B

Row representation

- with a_i^T the rows of A

$$AB = \begin{bmatrix} a_1^T B \\ a_2^T B \\ \vdots \\ a_m^T B \end{bmatrix} = \begin{bmatrix} (B^T a_1)^T \\ (B^T a_2)^T \\ \vdots \\ (B^T a_m)^T \end{bmatrix}$$

- row i is $(B^T a_i)^T$

Inner and outer product representations

Inner product representation: A is $m \times n$ with rows a_i^T, B is $n \times p$ with columns b_i

$$AB = \begin{bmatrix} a_1^T b_1 & a_1^T b_2 & \cdots & a_1^T b_p \\ a_2^T b_1 & a_2^T b_2 & \cdots & a_2^T b_p \\ \vdots & \vdots & & \vdots \\ a_m^T b_1 & a_m^T b_2 & \cdots & a_m^T b_p \end{bmatrix}$$

i, jth entry is $a_i^T b_j$

Outer product representation: A is $m \times n$ with rows a_i^T, B is $n \times p$ with rows b_i^T

$$AB = [a_1 \ \ \cdots \ \ a_n] \begin{bmatrix} b_1^T \\ \vdots \\ b_n^T \end{bmatrix} = a_1 b_1^T + \cdots + a_n b_n^T$$

Trace of a matrix

the *trace* of a square matrix $A \in \mathbb{R}^{n\times n}$ is the sum of its diagonal entries:

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

some properties of the trace are:

- $\mathrm{tr}(A) = \mathrm{tr}(A^T)$
- $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$ for square and equal size matrices A and B
- $\mathrm{tr}(\beta A) = \beta\,\mathrm{tr}(A)$ for any scalar β
- if A is an $m \times n$ matrix and B is an $n \times m$ matrix, then

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

- $\mathrm{tr}(ab^T) = \mathrm{tr}(b^Ta) = b^Ta$ for any n-vectors a and b

Inner product of matrices: the standard inner product between $A, B \in \mathbb{R}^{m\times n}$

$$\langle A, B\rangle = \mathrm{tr}(A^TB) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}$$

Determinant of a matrix

the determinant of a square matrix for value of i ($i = 1, 2, \ldots, n$) is

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}$$

- A_{ij} is the ij*th submatrix* of A obtained by removing row i and column j from A; for example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, \quad A_{32} = \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix}$$

- $\det A_{ij}$ is called the ijth *minor* of A
- $(-1)^{i+j} \det(A_{ij})$ is called the ijth *cofactor* of A

Examples

- for a scalar matrix $A = [a_{11}]$, we have $\det A = a_{11}$
- for a 2×2 matrix:

$$\det A = \det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{21}a_{12}$$

- for the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

we have for $i = 1$

$$A_{11} = \begin{bmatrix} 5 & 6 \\ 8 & 9 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 9 \end{bmatrix}, \quad A_{13} = \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}$$

thus, the determinant is

$$\begin{aligned} \det A &= (-1)^2 a_{11}(\det A_{11}) + (-1)^3 a_{12}(\det A_{12}) + (-1)^4 a_{13}(\det A_{13}) \\ &= a_{11}(\det A_{11}) - a_{12}(\det A_{12}) + a_{13}(\det A_{13}) \\ &= 1(-3) - 2(-6) + 3(-3) = 0 \end{aligned}$$

Determinant properties

- $\det A = \det A^T$
- $\det \beta A = \beta^n \det A$ for any scalar β
- $\det AB = \det A \times \det B$ for square matrices A and B
- if A is lower/upper triangular, then $\det A = a_{11} \cdots a_{nn}$
- if A is block upper/lower triangular, with square diagonal blocks $A_{11}, \ldots, A_{kk}$ (of possibly different sizes), then $\det A = \det A_{11} \cdots \det A_{kk}$
- determinant unchanged if we add to a column a linear comb. of other columns
- swapping two rows/columns changes the sign of $\det(A)$

Outline

- vectors
- vector operations
- inner product and norm
- matrices
- matrix operations
- **functions**
- linear equations

Functions

- $f : \mathcal{X} \to \mathcal{Y}$ denotes a *function* f that maps an element from set $\mathcal{X}$ to set $\mathcal{Y}$
- $f : \mathbb{R}^n \to \mathbb{R}^m$ means that f maps a real n-vector to a real m-vector:

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}$$

where the entry $f_i : \mathbb{R}^n \to \mathbb{R}$ is itself a scalar-valued function of x

Function domain

- the *domain* of f, denoted by $\mathrm{dom}\, f \subseteq \mathcal{X}$, is the set where f is defined and finite
- for example, the functions

$$f_1(x) = \begin{cases} 1/x & \text{if } x \neq 0 \\ \infty & \text{otherwise} \end{cases}, \quad f_2(x) = \begin{cases} 1/x & \text{if } x > 0 \\ \infty & \text{otherwise} \end{cases}$$

are different since they have different domains

Examples

Defined everywhere ($\mathrm{dom}\, f = \mathbb{R}^n$)

- $f : \mathbb{R} \to \mathbb{R}$: $f(x) = x^2 + x + 1$ maps a scalar x to a scalar $f(x)$
- $f : \mathbb{R}^3 \to \mathbb{R}$: $f(x_1, x_2, x_3) = x_1 + x_2 + x_3^2$
- $f : \mathbb{R}^n \to \mathbb{R}^m$: $f(x) = Ax$ where $x \in \mathbb{R}^n$ and A is an $m \times n$ matrix
- $f : \mathbb{R}^2 \to \mathbb{R}^3$: $f(x_1, x_2) = (x_1, x_2, x_1 + x_2^2)$

Undefined everywhere

- $f(x) = \log x$ is valid only for $x > 0$, hence $\mathrm{dom}\, f = \{x \mid x > 0\}$
- $f(x_1, x_2) = x_1/(x_1 + x_2)$ has domain $\mathrm{dom}\, f = \{(x_1, x_2) \mid x_1 + x_2 \neq 0\}$

Linear functions

Linear functions: f is linear if it satisfies the *superposition* property

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

for all numbers α, β, and all n-vectors x, y

Extension: if f is linear, then

$$f(\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_m u_m) = \alpha_1 f(u_1) + \alpha_2 f(u_2) + \cdots + \alpha_m f(u_m)$$

for all n-vectors $u_1, \ldots, u_m$ and all scalars $\alpha_1, \ldots, \alpha_m$

Linear functions as matrix-vector product

define $f(x) = Ax$ for fixed $A \in \mathbb{R}^{m\times n}$ ($f : \mathbb{R}^n \to \mathbb{R}^m$)

- any function of this type is linear: $A(\alpha x + \beta y) = \alpha(Ax) + \beta(Ay)$
- every linear function f can be written as $f(x) = Ax$:

$$\begin{aligned} f(x) &= f(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n) \\ &= x_1 f(e_1) + x_2 f(e_2) + \cdots + x_n f(e_n) \\ &= [f(e_1)\ f(e_2)\ \cdots\ f(e_n)] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = Ax \end{aligned}$$

  where $A = [f(e_1)\ f(e_2)\ \cdots\ f(e_n)]$ and $f(e_i)$ is an m-vector

- for $f : \mathbb{R}^n \to \mathbb{R}$, we get inner product function $f(x) = a^T x$

Examples

Linear

- average function of an n-vector, $f(x) = (\mathbf{1}/n)^T x = (x_1 + \cdots + x_n)/n$
- f reverses the order of the components of x is linear

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

- f scales x_1 by a given number d_1, x_2 by d_2, x_3 by d_3 is linear

$$A = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix}$$

Nonlinear

- f sorts the components of x in decreasing order: not linear
- f replaces each x_i by its absolute value $|x_i|$: not linear

Affine function

a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is *affine* if it satisfies

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

for all n-vectors x, y and all scalars α, β with $\alpha + \beta = 1$

Extension: if f is affine, then

$$f(\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_m u_m) = \alpha_1 f(u_1) + \alpha_2 f(u_2) + \cdots + \alpha_m f(u_m)$$

for all n-vectors $u_1, \ldots, u_m$ and all scalars $\alpha_1, \ldots, \alpha_m$ with

$$\alpha_1 + \alpha_2 + \cdots + \alpha_m = 1$$

Affine functions and matrix-vector product

$f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is affine, if and only if it can be expressed as

$$f(x) = Ax + b$$

for some $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$

- to see it is affine, let $\alpha + \beta = 1$ then

$$A(\alpha x + \beta y) + b = \alpha(Ax + b) + \beta(Ay + b)$$

- using the definition, we can show

$$A = [f(e_1) - f(0) \;\; f(e_2) - f(0) \;\; \cdots \;\; f(e_n) - f(0)], \quad b = f(0)$$

- for $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ the above becomes $f(x) = a^T x + b$

Quadratic functions

a function $f : \mathbb{R}^n \to \mathbb{R}$ is *quadratic* if it can be expressed as

$$f(x) = x^T Q x + x^T r + s$$

- Q is an $n \times n$ matrix
- r is an n-vector
- s is a scalar

Quadratic form

- a quadratic form is a special case: $x^T Q x$ where Q is symmetric
- we can always assume Q is symmetric because:

$$x^T Q x = (1/2) x^T (Q + Q^T) x$$

hence, $x^T Q x = x^T P x$ with $P = \frac{1}{2}(Q + Q^T)$ being symmetric

Some sets notation

- *nonnegative orthant:*

$$\mathbb{R}^n_+ = \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n \geq 0\}$$

- *positive orthant:*

$$\mathbb{R}^n_{++} = \{(x_1, x_2, \ldots, x_n) \mid x_1, x_2, \ldots, x_n > 0\}$$

- *symmetric matrices:*

$$\mathbb{S}^n = \{X \in \mathbb{R}^{n \times n} \mid X = X^T\}$$

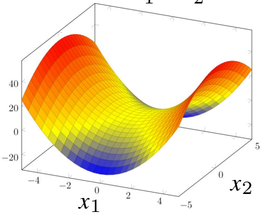
Level sets

the *level set* (*sublevel set* or *contour lines*) of a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ at level γ is

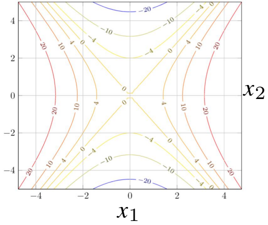
$$\mathcal{S}_\gamma = \{x \mid f(x) = \gamma\}$$

- the set of points with function value equal to γ
- for $n = 2$, this level set is called a *curve*; for $n = 3$, it is a *surface*
- for larger values of n, it is referred to as a *hyper-surface*
- example:

$f(x) = 2x_1^2 - x_2^2$



Level sets (controur lines) of $f(x)$



Outline

- vectors
- vector operations
- inner product and norm
- matrices
- matrix operations
- functions
- **linear equations**

Systems of linear equations

set (system) of m linear equations in n variables $x_1, \ldots, x_n$:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

- can express compactly as $Ax = b$
- a_{ij} are the *coefficients*; A is the *coefficient matrix*
- b is called the *right-hand side*
- may have no solution, a unique solution, infinitely many solutions

Classification

- under-determined if $m < n$ (A wide; more unknowns than equations)
- square if $m = n$ (A square)
- over-determined if $m > n$ (A tall; more equations than unknowns)

Examples

- no solution

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 2 \\ 2x_1 + 3x_3 &= 1 \end{aligned}$$

- unique solution

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 2 \\ x_2 + 3x_3 &= 1 \end{aligned}$$

- infinitely many solutions

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 2 \end{aligned}$$

Example: polynomial interpolation

- polynomial of degree at most $n-1$ with coefficients $x_1, x_2, \ldots, x_n$:

$$p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

- fit polynomial to m given points $(t_1, y_1), \ldots (t_m, y_m)$
- this is a system of linear equations:

$$Ax = \begin{bmatrix} 1 & t_1 & \cdots & t_1^{n-1} \\ 1 & t_2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & t_m & \cdots & t_m^{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

where A is the *Vandermonde matrix*

Particular and general solution

$$\begin{bmatrix} 1 & 0 & 8 & -4 \\ 0 & 1 & 2 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \end{bmatrix}$$

- first two columns consist of a 1 and a 0, so a particular solution is $\hat{x} = (42, 8, 0, 0)$
- to find a general solution, we find $Ax_0 = 0$; for any x_3, x_4

$$x_1 = -8x_3 + 4x_4, \quad x_2 = -2x_3 - 12x_4$$

  so $x_0 = (-8x_3 + 4x_4, -2x_3 - 12x_4, x_3, x_4)$ satisfies $Ax_0 = 0$
- combining solutions, the set of all solution, called *general solution*, is

$$x = \begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -8x_3 + 4x_4 \\ -2x_3 - 12x_4 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 42 \\ 8 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -8 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 4 \\ -12 \\ 0 \\ 1 \end{bmatrix}, \quad x_3, x_4 \in \mathbb{R}$$

Elementary row transformation

the solution of $Ax = b$ is invariant under the elementary operations:

- exchange of two equations (rows of *augmented* matrix $[A\ b]$)
- multiplication of an equation (row of $[A\ b]$) with a nonzero constant
- addition of two equations (rows of $[A\ b]$)

Row echelon form: system is in *row-echelon form* if it has staircase structure:

- all rows that contain only zeros are below the nonzero rows (bottom of matrix)
- in nonzero rows, *leading coefficient* or *pivot* is to right of pivot of row above it

it is in **reduced row-echelon form** or *row canonical form* (as in page 1.69) if further

- every pivot is 1
- pivot is the only nonzero entry in its column

Basic and free variables

- variables corresponding to the pivots are called *basic variables*
- other variables are called *free variables*

Gaussian elimination

Gaussian elimination is an algorithm that solves $Ax = b$ by transforming $[A\ b]$ into (reduced) row-echelon form

to find all solutions to $Ax = b$:

1. find a particular solution to $Ax = b$ by Guassian elimination
   - obtained from pivot columns (basic variables) with free variables set to zero
2. find all solutions to the homogeneous equation $Ax = 0$
   - by expressing basic variables in term of free variables
3. combine the solutions to the general solution

Example

$$\begin{aligned} -3x_1 + 2x_3 &= -1 \\ x_1 - 2x_2 + 2x_3 &= -5/3 \\ -x_1 - 4x_2 + 6x_3 &= -13/3 \end{aligned}$$

- r_i: ith equation or row of $[A\ b]$
- transform system into row echelon-form

$$\left[\begin{array}{ccc|c} -3 & 0 & 2 & -1 \\ 1 & -2 & 2 & -5/3 \\ -1 & -4 & 6 & -13/3 \end{array}\right] \underset{\substack{(1/3)\mathrm{r}_1+\mathrm{r}_2 \\ -(1/3)\mathrm{r}_1+\mathrm{r}_3}}{\longrightarrow} \left[\begin{array}{ccc|c} -3 & 0 & 2 & -1 \\ 0 & -2 & 8/3 & -2 \\ 0 & -4 & 16/3 & -4 \end{array}\right]$$

$$\xrightarrow{-2\mathrm{r}_2+\mathrm{r}_3} \left[\begin{array}{ccc|c} -3 & 0 & 2 & -1 \\ 0 & -2 & 8/3 & -2 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

we can work backward to solve this system or continue to make it into reduced row echelon form

- multiplying row 1 by $-1/3$ and row 2 by $1/{-2}$, we obtain the canonical form

$$\left[\begin{array}{ccc|c} 1 & 0 & -2/3 & 1/3 \\ 0 & 1 & -4/3 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

basic variables are x_1, x_2 and free variable is x_3

- a particular solution is $x = (1/3, 1, 0)$ and the homogeneous solution is

$$x_0 = \begin{bmatrix} (2/3)x_3 \\ (4/3)x_3 \\ x_3 \end{bmatrix}$$

- the set of all solutions is

$$\left\{ \begin{bmatrix} 1/3 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 2/3 \\ 4/3 \\ 1 \end{bmatrix} z \quad \middle| \quad z \in \mathbb{R} \right\}$$

each value of z gives a different solution

Example

suppose after Gaussian elimination, we obtain

$$[A\ b] = \left[\begin{array}{ccccr|c} 1 & 3 & 0 & 0 & 3 & 1 \\ 0 & 0 & 1 & 0 & 9 & 2 \\ 0 & 0 & 0 & 1 & -4 & 3 \end{array}\right]$$

- basic variables are x_1, x_3, x_4 and a particular solution is $x = (1, 0, 2, 3, 0)$
- for $Ax = 0$ expressing the basic variables in terms of free variables x_2, x_5:

$$x_1 = -3x_2 - 3x_5, \quad x_3 = -9x_5, \quad x_4 = 4x_5$$

- so the homogeneous solution has the form

$$\begin{bmatrix} 3x_2 - 3x_5 \\ x_2 \\ -9x_5 \\ 4x_5 \\ x_5 \end{bmatrix} = x_2 \begin{bmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} -3 \\ 0 \\ -9 \\ 4 \\ 1 \end{bmatrix}, \quad x_2, x_5 \in \mathbb{R}$$

References and further readings

- S. Boyd and L. Vandenberghe. *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least Squares.* Cambridge University Press, 2018. (Ch. 1,2,3,6,8,10)

- L. Vandenberghe, *EE133A Lecture Notes*, UCLA.

- E. K.P. Chong, Wu-S. Lu, and S. H. Zak, *An Introduction to Optimization: With Applications to Machine Learning*. John Wiley & Sons, 2023. (Ch. 2, 5.5)