Lecture 1: Review of linear algebra

- Linear functions and linearization
- Inverse matrix, least-squares and least-norm solutions
- Subspaces, basis, and dimension
- Change of basis and similarity transformations
- Eigenvalues and eigenvectors

Notation

- $\mathbb{R}$ — real numbers, $\mathbb{Z}$ — integers, $\mathbb{N}$ — natural numbers
- $\mathbb{R}^n$ — n-dimensional real vector space
- $\mathbb{R}^{m\times n}$ — space of real $m \times n$ matrices
- LHS $:=$ RHS — the LHS is defined by the RHS
- $A^\top$ — the transposed of A

Linear functions

- $f : \mathbb{R}^n \to \mathbb{R}^m$ — function mapping vectors in $\mathbb{R}^n$ to vectors in $\mathbb{R}^m$

  Interpretation of $y = f(x)$: x given input, y corresponding output

$$x \longrightarrow \boxed{f} \longrightarrow y$$

- f is a linear function if and only if superposition holds:

$$f(\alpha x + \beta v) = \alpha f(x) + \beta f(v), \quad \text{for all } \alpha, \beta \in \mathbb{R},\ x, v \in \mathbb{R}^n$$

- f is linear $\iff \exists\, A \in \mathbb{R}^{m\times n}$, such that $f(x) = Ax$, for all $x \in \mathbb{R}^n$

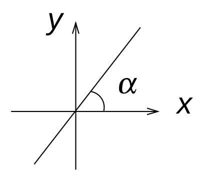
  A is a matrix representing the linear function f

Q: How can you find a matrix representation of a linear function f, if you are allowed to evaluate f at points $x \in \mathbb{R}^n$ of your choice?

Examples of linear functions

- Scalar function of a scalar argument

$$y = \tan(\alpha)x, \qquad \text{where} \quad \alpha \in [0, 2\pi)$$



- Identity function $x = f(x)$, for all $x \in \mathbb{R}^n$
  is a linear function represented by the identity matrix

$$I_n := \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

Matrix–vector multiplication

Partition $A \in \mathbb{R}^{m\times n}$ elementwise, column-wise, and row-wise

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} | & & | \\ c_1 & \cdots & c_n \\ | & & | \end{bmatrix} = \begin{bmatrix} — & r_1 & — \\ & \vdots & \\ — & r_m & — \end{bmatrix}$$

The matrix–vector product $y = Ax$ can be written as

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{bmatrix} = \sum_{j=1}^n c_j x_j = \begin{bmatrix} r_1 x \\ \vdots \\ r_m x \end{bmatrix}$$

Interpretation: a_{ij} gain factor from the jth input x_j to the ith output y_i.
(*e.g.*, $a_{ij} = 0$ means that jth input has no influence on ith output.)

Linearlization

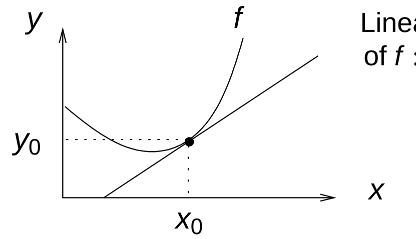
Consider a differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$. Then for given $x_0 \in \mathbb{R}^n$

$$y = f(x_0 + \widetilde{x}) \approx \underbrace{f(x_0)}_{y_0} + A\widetilde{x} \qquad \text{where} \quad A = [a_{ij}] = \left[\left. \frac{\partial f_i}{\partial x_j} \right|_{x_0} \right].$$

When the input deviation $\widetilde{x} = x - x_0$ is "small", the output deviation

$$\widetilde{y} := y - y_0$$

is approximately a linear function of $\widetilde{x}$, $\widetilde{y} = A\widetilde{x}$



Linear approximation of $f : x \mapsto y$ at $x = x_0$

Rank of a matrix and inversion

- the set of vectors $\{v_1, \ldots, v_n\}$ is independent if

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0 \quad \text{only if} \quad \alpha_1 = \cdots = \alpha_n = 0$$

- rank of a matrix — number of lin. indep. columns (or rows)
- $A \in \mathbb{R}^{m\times n}$ is full row rank (f.r.r.) if $\operatorname{rank}(A) = m$

  Interpretation: A not f.r.r. — there are redundant outputs
- Inversion problem: given $y \in \mathbb{R}^m$, find x, such that $y = Ax$.

  Interpretation: design an input that achieves a desired output.
- When is the inversion problem solvable? Is the solution unique?

Inversion problem

Given $y \in \mathbb{R}^m$, find x, such that $y = Ax$.

Solution may not exist, be unique, or there may be ∞ many solutions.
(Why it is not possible to have a finite number of solutions?)

Interpretations:

- Control: x is a control input, y is a desired outcome
- Estimation: x is a vector of parameters, y is a set of measurements

Typically

in control, the solution is nonunique and we aim to find the "best" one.

in estimation, there is no solution and we aim to find the "best" approximation.

Inverse of a matrix

If $m = n = \text{rank}(A)$, then there exists a matrix A^{-1}, such that

$$AA^{-1} = A^{-1}A = I_m.$$

Then for all $y \in \mathbb{R}^m$

$$y = \underbrace{(AA^{-1})}_{I} y = A \underbrace{(A^{-1}y)}_{x} = Ax.$$

The inversion problem is solvable and the solution is unique.

Q: Can you find a matrix representation of a linear function f, from given values $y_1, \ldots, y_n$ of f at given points $x_1, \ldots, x_n$? If so, how?

Vector and matrix norms

Mathematical formalisation of the geometric notion of size or distance.

Norm is a function $\|x\| : x \mapsto \mathbb{R}$ that satisfies the following properties:

- Nonnegativity: $\|x\| \geq 0$ for all x
- Definiteness: $\|x\| = 0 \iff x = 0$
- Homogeneity: $\|\alpha x\| = |\alpha| \|x\|$ for all x and α
- Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$

Examples:

- Vector 2-norm: $\|x\|_2 := \sqrt{x_1^2 + \cdots + x_n^2} = \sqrt{x^\top x}$, for all $x \in \mathbb{R}^n$
- Frobenius matrix norm: $\|A\|_{\text{F}} := \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$, for all $A \in \mathbb{R}^{m \times n}$

Unit ball: $\mathscr{U} = \{ x \mid \|x\| \leq 1 \}$

Least-squares solution

Assumption $m \geq n = \text{rank}(A)$, *i.e.*, A is full column rank.
The inversion problem typically has no solution.

The least-squares solution

$$x_{\text{ls}} = (A^\top A)^{-1} A^\top y =: A^+ y$$

minimizes the approximation error

$$\|\underbrace{y - Ax}_{e}\|_2 := \sqrt{e_1^2 + \cdots + e_m^2} = \sqrt{e^\top e}.$$

The matrix

$$A^+ := (A^\top A)^{-1} A^\top \qquad (\text{if } m > n = rank(A))$$

is called pseudo-inverse of A.

Notes

- x_{ls} is a linear function of y (given by the pseudo inverse matrix A^+)
- If A is square $x_{\text{ls}} = A^{-1}y$ (in other words $A^+ = A^{-1}$)
- x_{ls} is an exact solution if $Ax = y$ has an exact solution
- $\widehat{y} = Ax_{\text{ls}} = A(A^\top A)^{-1} A^\top y$ is a least-squares approximation of y
- Statistical interpretation: assume that

  $$y = Ax_0 + e$$

  where e is zero mean Gaussian random vector with covariance $\sigma^2 I$

  Then x_{ls} is the best linear unbiased estimator for x_0.

Least-norm solution

Assumption $n \geq m = \text{rank}(A)$, *i.e.*, A is full row rank.
The inversion problem has infinitely many solution.

The least-norm solution

$$x_{\text{ln}} = A^\top (AA^\top)^{-1} y =: A^+ y$$

minimizes the 2-norm of the solution x, *i.e.*,

$$\text{minimize } \|x\|_2 \quad \text{subject to} \quad Ax = y$$

The matrix

$$A^+ := A^\top (AA^\top)^{-1} \qquad (\text{if } n > m = \text{rank}(A))$$

is called pseudo-inverse of A.

Set of all solutions

$$\{\, x \mid Ax = y \,\} = \{\, x_{\text{p}} + z \mid Az = 0 \,\}$$

where x_{p} is a particular solution, *i.e.*, $Ax_{\text{p}} = y$.

Note that $x_{\text{ln}} = A^\top (AA^\top)^{-1} y$ is a particular solution

$$Ax_{\text{ln}} = (AA^\top)(AA^\top)^{-1} y = y.$$

Moreover, x_{ln} is the minimum 2-norm solution.

Inner product

- The inner product of two vectors $a, b \in \mathbb{R}^n$ is defined as

$$\langle a, b \rangle := a^\top b = \sum_{k=1}^{n} a_k b_k.$$

- Matrix–matrix product $H = GF$, $F : \mathbb{R}^{p \times n}$, $G : \mathbb{R}^{n \times m}$ gives pm inner products between the rows of G and the columns of F

$$H = GF = \begin{bmatrix} - & g_1 & - \\ & \vdots & \\ - & g_m & - \end{bmatrix} \begin{bmatrix} | & & | \\ f_1 & \cdots & f_p \\ | & & | \end{bmatrix} = \begin{bmatrix} \langle g_1, f_1 \rangle & \cdots & \langle g_1, f_p \rangle \\ \vdots & & \vdots \\ \langle g_m, f_1 \rangle & \cdots & \langle g_m, f_p \rangle \end{bmatrix}$$

- The Gram matrix of the vectors $f_1, \ldots, f_m$ is defined by

$$\begin{bmatrix} f_1^\top \\ \vdots \\ f_m^\top \end{bmatrix} \begin{bmatrix} f_1 & \cdots & f_m \end{bmatrix}$$

Cauchy-Schwarz inequality

$$|x^\top y| \leq \|x\| \|y\|$$

equality holds if and only if $x = \alpha y$, for some $\alpha \in \mathbb{R}$ or $x = 0$.

Application: optimization of a linear function over the unit ball

Given $y \in \mathbb{R}^n$

$$\text{maximize} \quad x^\top y \quad \text{subject to} \quad \|x\| \leq 1$$

The solution follows from the Cauchy-Schwarz inequality

$$x_{\text{opt}} = \frac{y}{\|y\|}$$

Angle between vectors

The angle between the vectors $x, y \in \mathbb{R}^n$ is defined as

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

- $x \neq 0$ and y are aligned if $y = \alpha x$, for some $\alpha \geq 0$
  In this case, $\angle(x,y) = 0$.
- $x \neq 0$ and y are opposite if $y = -\alpha x$, for some $\alpha \geq 0$
  In this case, $\angle(x,y) = \pi$.
- x and y are orthogonal (denoted $x \perp y$) if $x^\top y = 0$
  In this case, $\angle(x,y) = \pi/2$.

Q: Given $y \in \mathbb{R}^n$, which x minimize $|x^\top y|$ subject to $\|x\| \geq 1$?

Subspace, basis, and dimension

- $\mathscr{V} \subset \mathbb{R}^n$ is a subspace of a vector space $\mathbb{R}^n$ if $\mathscr{V}$ is a vector space

$$v, w \in \mathscr{V} \quad \Longrightarrow \quad \alpha v + \beta w \in \mathscr{V}, \quad \text{for all } \alpha, \beta \in \mathbb{R}$$

- The set $\{v_1, \ldots, v_n\}$ is a basis of $\mathscr{V}$ if
  - $v_1, \ldots, v_n$ span $\mathscr{V}$, *i.e.*,

$$\mathscr{V} = \text{span}(v_1, \ldots, v_n) := \{\alpha_1 v_1 + \cdots + \alpha_n v_n \mid \alpha_1, \ldots, \alpha_n \in \mathbb{R}\}$$

  - $\{v_1, \ldots, v_n\}$ is an independent set of vectors.
- $\dim(\mathscr{V})$ — number of basis vectors (does not depend on the basis)

Null space of a matrix (kernel)

- kernel of A — the set of vectors mapped to zero by $f(x) := Ax$

$$\ker(A) := \{x \in \mathbb{R}^n \mid Ax = 0\}$$

- $y = A(x + \tilde{x})$, for all $\tilde{x} \in \ker(A)$

  Interpretation: $\ker(A)$ is the uncertainty in finding x, given y.

  Interpretation: $\ker(A)$ is the freedom in the x's that achieve y.

- $\ker(A) = \{0\} \quad \Longleftrightarrow \quad f(x) := Ax$ is one-to-one
- $\ker(A) = \{0\} \quad \Longleftrightarrow \quad A$ is full column rank

Range of a matrix (image)

- image of A — the set of all vectors obtainable by $f(x) := Ax$

$$\text{image}(A) := \{Ax \mid x \in \mathbb{R}^n\}$$

- $\text{image}(A)$ = span of the columns of A
- $\text{image}(A)$ = set of vectors y for which $Ax = y$ has a solution
- $\text{image}(A) = \mathbb{R}^m \quad \Longleftrightarrow \quad f(x) := Ax$ is onto ($\text{image}(f) = \mathbb{R}^m$)
- $\text{image}(A) = \mathbb{R}^m \quad \Longleftrightarrow \quad A$ is full row rank

Change of basis

- standard basis vectors in $\mathbb{R}^n$ — the columns $e_1, \dots, e_n$ of I_n
- Elements of $x \in \mathbb{R}^n$ are coordinates of x w.r.t. standard basis.
- A new bases is given by the columns $t_1, \dots, t_n$ of $T \in \mathbb{R}^{n \times n}$.
- The coordinates of x in the new basis are $\widetilde{x}_1, \dots, \widetilde{x}_n$, such that

$$x = \widetilde{x}_1 t_1 + \cdots + \widetilde{x}_n t_n = T\widetilde{x} \quad \Longrightarrow \quad \widetilde{x} = T^{-1}x$$

- T^{-1} transforms standard basis coordinates x into T-coordinates

Similarity transformation

- Consider linear operator $f : \mathbb{R}^n \to \mathbb{R}^n$, given by $f(x) = Ax$, $A \in \mathbb{R}^{n \times n}$.
- Change standard basis to basis defined by columns of $T \in \mathbb{R}^{n \times n}$.
- The matrix representation of f changes to $T^{-1}AT$:

$$x = T\widetilde{x}, \quad y = T\widetilde{y} \qquad \Longrightarrow \qquad \widetilde{y} = (T^{-1}AT)\widetilde{x}$$

- $A \mapsto T^{-1}AT$ — similarity transformation of A

Eigenvalues and eigenvectors

$$\begin{aligned} \lambda \in \mathbb{C} \text{ is eigenvalue of } A \in \mathbb{R}^{n \times n} \quad &:\Longleftrightarrow \quad \text{there is } v \neq 0, \text{ s.t. } Av = \lambda v \\ &:\Longleftrightarrow \quad \lambda I_n - A \text{ is singular} \end{aligned}$$

Any nonzero $v \in \mathbb{C}^n$ such that $Av = \lambda v$ is called an eigenvector of A associate with the eigenvalue λ.

Meaning of λ and v: the action of A in the direction defined by v is equivalent to scalar multiplication by λ

Characteristic polynomial of A: $p_A(\lambda) := \det(\lambda I_n - A)$, $\deg(p_A) = n$

λ is an eigenvalue of A if and only if λ is a root of p_A

Geometric multiplicity of λ: $\dim\big((\lambda I_n - A)\big)$

Algebraic multiplicity of λ: multiplicity of the root λ of p_A

Eigenvalue decomposition

Suppose $\{v_1, \dots, v_n\}$ is a lin. indep. set of eigenvectors of $A \in \mathbb{R}^{n \times n}$

$$Av_i = \lambda_i v_i, \qquad \text{for } i = 1, \dots, n$$

written in a matrix form is

$$A \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{V} = \underbrace{\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}}_{V} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{\Lambda}$$

V is nonsingular, so that

$$AV = V\Lambda \quad \Longrightarrow \quad V^{-1}AV = \Lambda$$

Diagonalization by similarity transformation

- V is nonsingular since by assumption $\{v_1, \dots, v_n\}$ is lin. indep.
- similarity transformation with $T = V^{-1}$ diagonalizes A

Conversely if there is a nonsingular $V \in \mathbb{C}^{n\times n}$, such that

$$V^{-1}AV = \Lambda$$

then $Av_i = \lambda_i v_i$ and $\{v_1, \dots, v_n\}$ is a lin. indep. set of eigenvectors

A is diagonalizable if

- there is nonsingular T, such that TAT^{-1} is diagonal
- there is a set of n lin. indep. eigenvectors of A

if A is not diagonalizable, it is called defective

Jordan canonical form

Distinct eigenvalues $\implies$ diagonalizable matrix (converse not true)

Prototypical example of a defective matrix: $\begin{bmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$

Defective matrices have an eigenvalue which algebraic multiplicity is higher than the corresponding geometric multiplicity.

Jordan form: generalization of $TAT^{-1} = \Lambda$ for defective matrices

Eigenvalues and eigenvectors of symmetric matrix

Theorem: A symmetric matrix A has real eigenvalues and a full set of eigenvectors, that can be chosen to form an orthonormal set.

Symmetric matrix $\implies$
- real eigenvalues
- orthonormal eigenvectors

Summary

- f is linear if superposition holds $f(\alpha x + \beta v) = \alpha f(x) + \beta f(v)$
- f is linear $\iff$ there is matrix A, such that $f(x) = Ax$
- image (column span, range) of A — $\text{image}(A) := \{Ax \mid x \in \mathbb{R}^n\}$
- kernel (null space) of A — $\ker(A) := \{x \in \mathbb{R}^n \mid Ax = 0\}$
- $\mathscr{V} \subset \mathbb{R}^n$ is subspace of $\mathbb{R}^n$ if $\alpha v + \beta w \in \mathscr{V}$ for all $v, w \in \mathscr{V}$
- basis of a subspace — set of linearly indep. vectors that span $\mathscr{V}$
- dimension of a subspace — the number of basis vectors
- $\text{image}(A)$ and $\ker(A)$ are subspaces

- rank of A — number of linearly independent rows (or columns)
- $\dim\big(\mathrm{image}(A)\big) = \mathrm{rank}(A)$, $\quad \mathrm{coldim}(A) - \dim(\ker(A)) = \mathrm{rank}(A)$
- A is full row rank if $\mathrm{rank}(A) = \mathrm{rowdim}(A)$
- A is full column rank if $\mathrm{rank}(A) = \mathrm{coldim}(A)$
- A is full rank if either full row or column rank
- A is nonsingular if A is square and full rank
- inversion problem: given $y = Ax$, find x
- A^+ is left inverse of A if $A^+A = I$
- solution of the inversion problem: $x = A^+y$, $A^+A = I$

- left inverse exists iff A is full column rank
- least-squares left inverse $A_{\mathrm{ls}} = (A^\top A)^{-1}A^\top$
- A^+ is right inverse of A if $AA^+ = I$
- right inverse exists iff A is full row rank
- least-norm right inverse $A_{\mathrm{ln}} = A^\top(AA^\top)^{-1}$
- A^{-1} is inverse of A if $A^{-1}A = A^{-1}A = I$
- for A to have inverse, A should be square and full rank
- 2-norm of a vector $\|x\| = \sqrt{x^\top x}$, unit ball $\{x \mid \|x\| \le 1\}$
- inner product of $a, b \in \mathbb{R}^n$ — $\langle a, b\rangle := a^\top b$

- Cauchy-Schwarz inequality: $|a^\top b| \le \|a\|\|b\|$
- $a, b \in \mathbb{R}^n$ are orthogonal if $\langle a, b\rangle = 0$
- similarity transformation — $A \mapsto T^{-1}AT$, T nonsingular
- eigenvalue decomposition — $A = T^{-1}\Lambda T$, Λ diagonal
- characteristic polynomial of A — $p_A(\lambda) := \det(\lambda I - A)$
- symmetric matrix $A = A^\top \implies$ real eigenvalues
orthonormal eigenvectors

References

Introductory texts:

- G. Strang, Introduction to linear algebra
- G. Strang, Linear algebra and its applications,
- C. Meyer, Matrix analysis and applied linear algebra, SIAM, 2000

Advanced texts:

- R. Bellman, Introduction to matrix analysis, 1970
- R. Horn & Johnson, Matrix analysis, Cambridge Univ. Press, 1985
- R. Horn & Johnson, Topics in matrix analysis, CUP, 1991