ELEC 3035, Lecture 1: Review of linear algebra

Ivan Markovsky

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

Notation

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

Linear functions

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

  Interpretation of $y = f(u)$: u given input, y corresponding output of a static system defined by f

  $$u \longrightarrow \boxed{\;f\;} \longrightarrow y$$

  $\mathtt{m}$ — number of inputs, $\mathtt{p}$— number of output

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

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

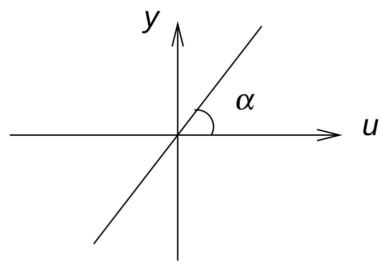
- f is linear $\iff \exists\, A \in \mathbb{R}^{\mathtt{p}\times\mathtt{m}}$, such that $f(u) = Au$, for all $u \in \mathbb{R}^{\mathtt{m}}$

  A is a matrix representing the linear function f

Examples of linear functions

- Scalar function of a scalar argument

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



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

  $$I_{\mathtt{m}} := \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \in \mathbb{R}^{\mathtt{m}\times\mathtt{m}}$$

Matrix–vector multiplication

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

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \vdots \\ a_{p1} & \cdots & a_{pm} \end{bmatrix} = \begin{bmatrix} | & & | \\ c_1 & \cdots & c_m \\ | & & | \end{bmatrix} = \begin{bmatrix} - & r_1 & - \\ & \vdots & \\ - & r_p & - \end{bmatrix}$$

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

$$\begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} a_{1j}u_j \\ \vdots \\ \sum_{j=1}^{m} a_{pj}u_j \end{bmatrix} = \sum_{j=1}^{m} c_j u_j = \begin{bmatrix} r_1 u \\ \vdots \\ r_p u \end{bmatrix}$$

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

Linearlization at a point

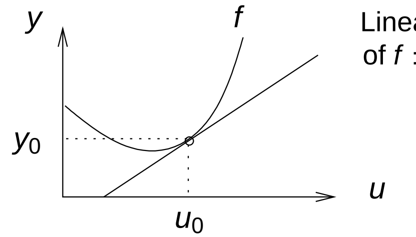
Consider a differentiable function $f : \mathbb{R}^m \to \mathbb{R}^p$. Then for given $u_0 \in \mathbb{R}^m$

$$y = f(u_0 + \tilde{u}) \approx \underbrace{f(u_0)}_{y_0} + A\tilde{u} \qquad \text{where} \quad A = [a_{ij}] = \left[\left. \frac{\partial f_i}{\partial u_j} \right|_{u_0} \right].$$

When the input deviation $\tilde{u} = u - u_0$ is "small", the output deviation

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

is approximately a linear function of $\tilde{u}$, $\tilde{y} = A\tilde{u}$



Linear approximation of $f : u \mapsto y$ at $u = u_0$

Rank of a matrix and inversion

- the set of vectors $y^{(1)}, \dots, y^{g}$ is independent if

$$\alpha_1 y^{(1)} + \cdots + \alpha_g y^{g} = 0 \quad \text{only if} \quad \alpha_1 = \cdots = \alpha_g = 0$$

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

  Interpretation: A not f.r.r. — there are redundant outputs

- Inversion problem:
  given $y \in \mathbb{R}^p$ and $A \in \mathbb{R}^{p\times m}$, find $u \in \mathbb{R}^m$, such that $y = Au$.

  Interpretation: design an input that achieves a desired output for a given system.

- When is the inversion problem solvable? Is the solution unique?

Inversion problem

Given $y \in \mathbb{R}^p$, find u, such that $y = Au$.

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: u is a control input, y is a desired outcome
- Estimation: u 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 $\mathtt{p} = \mathtt{m} = \text{rank}(A)$, then there exists a matrix A^{-1}, such that

$$AA^{-1} = A^{-1}A = I_{\mathtt{p}}.$$

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

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

The inversion problem is solvable and the solution is unique.

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: $\|u\|_2 := \sqrt{u_1^2 + \cdots + u_{\mathtt{m}}^2} = \sqrt{x^\top x}$, for all $u \in \mathbb{R}^{\mathtt{m}}$
- Frobenius matrix norm: $\|A\|_{\mathrm{F}} := \sqrt{\sum_{i=1}^{\mathtt{p}} \sum_{j=1}^{\mathtt{m}} a_{ij}^2}$, for all $A \in \mathbb{R}^{\mathtt{p}\times\mathtt{m}}$

Least squares solution

Assumption $\mathtt{p} \geq \mathtt{m} = \text{rank}(A)$, *i.e.*, $A \in \mathbb{R}^{\mathtt{p}\times\mathtt{m}}$ is full column rank.
The inversion problem has infinitely many solution.

The least squares solution

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

minimises the approximation error

$$\|\underbrace{y - Au}_{e}\|_2 := \sqrt{e_1^2 + \cdots + e_{\mathtt{p}}^2} = \sqrt{e^\top e}.$$

The matrix

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

is called pseudo-inverse of A.

Derivation of the least squares solution

Assumption $\mathtt{p} \geq \mathtt{m} = \text{rank}(A)$, *i.e.*, $A \in \mathbb{R}^{\mathtt{p}\times\mathtt{m}}$ is full column rank.

To minimise the norm of the residual e

$$\|e\|_2^2 = \|y - Au\|_2^2 = (y - Au)^\top (y - Au) = u^\top A^\top Au - 2y^\top Au + y^\top y$$

over u, set the gradient with respect to u equal to zero

$$\nabla_u \|e\|_2^2 = \nabla_u (u^\top A^\top Au - 2y^\top Au + y^\top y) = 2A^\top Au - 2A^\top y = 0.$$

This gives the linear equation $A^\top Au = A^\top y$ in u, called normal equation.

A full column rank, implies that $A^\top A$ is nonsingular, so that

$$u_{\mathrm{ls}} = (A^\top A)^{-1}A^\top y$$

is the unique least squares approximate solution.

Notes

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

$$y = Au_0 + e$$

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

  Then $u_{\rm ls}$ is the best linear unbiased estimator for u_0.

Least norm solution

Assumption $\mathtt{m} \geq \mathtt{p} = \text{rank}(A)$, *i.e.*, $A \in \mathbb{R}^{\mathtt{p}\times\mathtt{m}}$ is full row rank.
The inversion problem has infinitely many solution.

The least norm solution

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

minimises the 2-norm of the solution u, *i.e.*,

$$\min_u \|u\|_2 \quad \text{subject to} \quad Au = y$$

The matrix

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

is called pseudo-inverse of A.

Set of all solutions

$$\{\, u \mid Au = y \,\} = \{\, u_0 + z \mid Az = 0 \,\}$$

where u_0 is a particular solution, *i.e.*, $Au_0 = y$.

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

$$Au_{\rm ln} = (AA^\top)(AA^\top)^{-1}y = y.$$

Moreover, $u_{\rm ln}$ is the minimum 2-norm solution.

Function composition and matrix–matrix multiplication

- Consider two functions $f : \mathbb{R}^{\mathtt{m}} \to \mathbb{R}^{\mathtt{n}}$ and $g : \mathbb{R}^{\mathtt{n}} \to \mathbb{R}^{\mathtt{p}}$
- The composition of f and g (in general, the order matters) is the function $h : \mathbb{R}^{\mathtt{m}} \to R^{\mathtt{p}}$, defined by

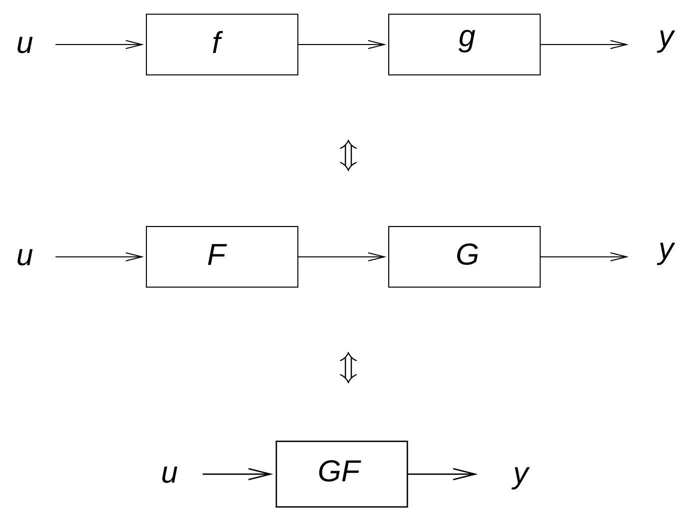
$$h = (gf)(u) := g\big(f(u)\big), \qquad \text{for all } u \in \mathbb{R}^{\mathtt{m}}$$

- Let $F \in \mathbb{R}^{\mathtt{m}\times\mathtt{n}}$ and $G \in \mathbb{R}^{\mathtt{n}\times\mathtt{p}}$ be matrices that represent f and g
- Then the matrix–matrix product $H = GF$

$$H = \big[h_{ij}\big] = GF := \big[\textstyle\sum_{k=1}^{\mathtt{n}} g_{ik} f_{kj}\big]$$

  represents the function composition $h = gf$. (Verify this.)

Function composition and matrix–matrix multiplication



Angle between vectors

The angle between the vectors $u, v \in \mathbb{R}^m$ is defined as

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

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

Subspace, basis, and dimension

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

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

- The set $\{u^{(1)}, \ldots, u^{(m)}\}$ is a basis of a vector space $\mathscr{U}$ if
  - $u^{(1)}, \ldots, u^{(m)}$ span $\mathscr{U}$, *i.e.*,

    $$\mathscr{U} = \text{span}(u^{(1)}, \ldots, u^{(m)}) := \{\alpha_1 u^{(1)} + \cdots + \alpha_m u^{(m)} \mid \alpha_1, \ldots, \alpha_m \in \mathbb{R}\}$$

  - $\{u^{(1)}, \ldots, v^{(m)}\}$ is an independent set of vectors.
- $\dim(\mathscr{U})$ — number of basis vectors (doen't depend on the basis)

Null space of a matrix (kernel)

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

$$\ker(A) := \{u \in \mathbb{R}^m \mid Au = 0\}$$

- $y = A(u+v)$, for all $v \in \ker(A)$

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

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

- $\ker(A) = \{0\} \quad \Longleftrightarrow \quad f(u) := Au$ 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 $f(u) := Au$

$$\text{image}(A) := \{ Au \mid u \in \mathbb{R}^m \}$$

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

Change of basis

- standard basis vectors in $\mathbb{R}^m$ — the columns $e^{(1)}, \dots, e^{(m)}$ of I_m
- Elements of $u \in \mathbb{R}^m$ are coordinates of x w.r.t. standard basis.
- A new bases is given by the columns $v^{(1)}, \dots, v^{(m)}$ of $V \in \mathbb{R}^{m \times m}$.
- The coordinates of u in the new basis are $\tilde{u}_1, \dots, \tilde{u}_m$, such that

$$u = \tilde{u}_1 v^{(1)} + \cdots + \tilde{u}_m v^{(m)} = V\tilde{u} \quad \Longrightarrow \quad \tilde{u} = V^{-1}u$$

- V^{-1} transforms standard basis coordinates u into V-coordinates

Similarity transformation

- Consider linear operator $f : \mathbb{R}^m \to \mathbb{R}^m$, given by $f(u) = Au$, $A \in \mathbb{R}^{m \times m}$.
- Change standard basis to basis defined by columns of $V \in \mathbb{R}^{m \times m}$.
- The matrix representation of f changes to $V^{-1}AV$:

$$u = V\tilde{u}, \quad y = V\tilde{y} \qquad \Longrightarrow \qquad \tilde{y} = (V^{-1}AV)\tilde{u}$$

- $A \mapsto V^{-1}AV$ — similarity transformation of A