Lecture 2 : Calculus and Linear Algebra

Piotr Zwiernik

Mathematics Brush-up

Barcelona School of Economics

Often a function f is given and we are looking for F whose derivative is f.

Example: the marginal cost function $C'(x)$ is known (how cost changes with production x). We want the cost $C(x)$ itself.

Idea: **integration** reverses differentiation: it accumulates small changes.

Read Chapter 5 of Werner-Sotskov   Exercises 5.1(a)-(b), 5.2(a)-(b), 5.3(c)

Indefinite integrals

A differentiable F is an antiderivative of f if $F'(x) = f(x)$ on a common domain.

Fact: all antiderivatives differ by a constant: if $F'(x) = f(x)$, then any $\tilde{F}(x) = F(x) + c$ also satisfies $\tilde{F}'(x) = f(x)$.

Definition: the indefinite integral is

$$\int f(x)\,dx \;=\; F(x) + c.$$

Linearity:

1. $\int (f+g)\,dx = \int f\,dx + \int g\,dx$
2. $\int c\,f\,dx = c \int f\,dx$

Indefinite integrals you should know

Templates:

1. $\displaystyle\int x^n \, dx = \frac{x^{n+1}}{n+1} + c, \ n \neq -1$
2. $\displaystyle\int \frac{1}{x} \, dx = \log|x| + c$
3. $\displaystyle\int e^x \, dx = e^x + c$
4. $\displaystyle\int \sin x \, dx = -\cos x + c$
5. $\displaystyle\int \cos x \, dx = \sin x + c$
6. $\displaystyle\int a^x \, dx = \frac{a^x}{\log a} + c \quad (a > 0, a \neq 1)$

Always check by differentiating the right-hand side.

Integration by substitution

Theorem (Substitution)

If $t = g(x)$ and F is an antiderivative of f, then

$$\int f(g(x))\, g'(x)\, dx = \int f(t)\, dt = F(t) + c = F(g(x)) + c.$$

Examples:

$$1.\quad \int (ax+b)^n\, dx = \frac{1}{a}\int t^n\, dt = \frac{(ax+b)^{n+1}}{a(n+1)} + c.$$

($t = ax + b$, so $dt = a\, dx$.)

$$2.\quad \int \frac{e^x}{\sqrt[3]{1+e^x}}\, dx = \int t^{-1/3}\, dt = \frac{3}{2} t^{2/3} + c = \frac{3}{2}(1+e^x)^{2/3} + c.$$

($t = 1 + e^x$, so $dt = e^x\, dx$.)

Integration by parts

Theorem (Integration by parts)

For differentiable u, v,

$$\int u(x)v'(x)\,dx = u(x)v(x) - \int u'(x)v(x)\,dx.$$

Proof: Differentiate $u(x)v(x)$ and integrate.

Example: with $u = \log x$, $v' = 1$,

$$\int \log x\,dx = x\log x - \int 1\,dx = x(\log x - 1) + c.$$

A combo: substitution then parts

Compute $\int \sin\sqrt{x}\,dx$.

Substitute $t = \sqrt{x}$, so $x = t^2$ and $dx = 2t\,dt$:

$$\int \sin\sqrt{x}\,dx = 2\int t\sin t\,dt.$$

Now parts with $u = t$, $v' = \sin t$:

$$2\int t\sin t\,dt = 2\Big(-t\cos t + \int \cos t\,dt\Big) = 2(-t\cos t + \sin t) + c.$$

Back-substitute $t = \sqrt{x}$:

$$\boxed{\int \sin\sqrt{x}\,dx = 2\big(-\sqrt{x}\cos\sqrt{x} + \sin\sqrt{x}\big) + c}.$$


A combo: substitution then parts

Definite integral as area and accumulation

For continuous $f : [a, b] \to \mathbb{R}$ with $f \geq 0$, the definite integral

$$\int_a^b f(x)\,dx$$

is the area under f between a and b. More generally it accumulates signed change.

Properties:

1. $\int_b^a f = -\int_a^b f$
2. $\int_a^b cf = c\int_a^b f$
3. If $c \in [a, b]$, then $\int_a^b f = \int_a^c f + \int_c^b f$
4. $\left|\int_a^b f\right| \leq \int_a^b |f|$
5. If $f \leq g$ on $[a, b]$, then $\int_a^b f \leq \int_a^b g$

Some uses: cumulative revenue from a known marginal revenue curve, energy used by a device with power draw $P(t)$, or probability mass from a density.

Fundamental Theorem of Calculus

Fundamental Theorem of Calculus

If f is continuous on $[a, b]$ and F is an antiderivative of f, then

$$\int_a^b f(x)\,dx = F(b) - F(a).$$

Moreover, $G(t) = \int_a^t f(x)\,dx$ is differentiable and $G'(t) = f(t)$.

Marginal to total: If $C'(x)$ is marginal cost, then the change in total cost from $x = 300$ to $x = 400$ is

$$C(400) - C(300) = \int_{300}^{400} C'(x)\,dx.$$

Example: $C'(x) = 6 - \dfrac{60}{x+1}$ for $x \in [0, 1000]$:

$$\int_{300}^{400} \left(6 - \frac{60}{x+1}\right) dx = \Big(6x - 60 \log |x+1|\Big)_{300}^{400} \approx 582.79.$$

Application: proving $e = \lim_{n\to\infty}(1+\frac{1}{n})^n$ via integrals

Define $\log x = \displaystyle\int_1^x \frac{1}{t}\,dt$. Let e^x be the inverse of $\log x$; then $1 = \int_1^e \frac{1}{t}\,dt$.

For $t \in [1, 1+\frac{1}{n}]$,

$$\frac{1}{n+1} = \int_1^{1+\frac{1}{n}} \frac{1}{1+\frac{1}{n}}\,dt \le \int_1^{1+\frac{1}{n}} \frac{1}{t}\,dt \le \int_1^{1+\frac{1}{n}} 1\,dt = \frac{1}{n}.$$

So

$$\frac{1}{n+1} \le \log\left(1+\frac{1}{n}\right) \le \frac{1}{n}.$$

Exponentiate and rearrange to obtain

$$\frac{e}{1+\frac{1}{n}} \le \left(1+\frac{1}{n}\right)^n \le e,$$

and let $n \to \infty$.

Application: Taylor with integral remainder (up to $n = 2$)

Apply FTC repeatedly:

$$
\begin{aligned}
f(x) &= f(x_0) + \int_{x_0}^{x} f'(t_1)\, dt_1 \\
&= f(x_0) + f'(x_0)(x - x_0) + \int_{x_0}^{x} \int_{x_0}^{t_1} f''(t_2)\, dt_2\, dt_1 \\
&= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \int_{x_0}^{x} \int_{x_0}^{t_1} \int_{x_0}^{t_2} f^{(3)}(t_3)\, dt_3\, dt_2\, dt_1.
\end{aligned}
$$

By the intermediate value theorem there is y between x_0 and x with

$$
\int\int\int f^{(3)}(t_3)\, dt_3\, dt_2\, dt_1 = \frac{f^{(3)}(y)}{3!}(x - x_0)^3.
$$

Chapter 5: Vectors

A vector is an ordered n-tuple of real numbers: $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$.

Why care in econ/data:

- a bundle of n goods (quantities),
- a user's features or a product's attributes,
- a portfolio's weights across n assets,
- a document's word counts or an embedding.

Read Chapter 6 of Werner-Sotskov; Simon-Blume Chs. 10-11.

Exercises 6.2, 6.3, 6.4, 6.6, 6.7, 6.8

Definition and notation

A vector **v** is an ordered n-tuple $(v_1, \ldots, v_n)$ of real numbers called coordinates.

Notation: $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{R}^n$.
The zero vector is $\mathbf{0} = (0, \ldots, 0)$.
The i-th unit vector is $\boldsymbol{e}_i$ (a 1 in position i, zeros elsewhere).

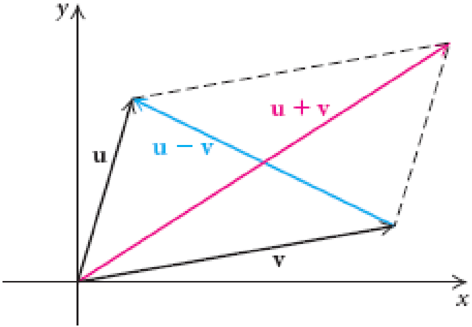
Operations:

$$\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, \ldots, u_n + v_n), \quad \lambda \boldsymbol{v} = (\lambda v_1, \ldots, \lambda v_n).$$

These satisfy the usual commutative, associative, and distributive laws.

Sum and difference of two vectors

Note $\boldsymbol{u} - \boldsymbol{v} \;=\; \boldsymbol{u} + (-1)\boldsymbol{v}$.



Inner product and norm

The inner product (dot product) of $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ is

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \sum_{i=1}^{n} u_i v_i .$$

The Euclidean norm is $\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$.

Properties:

1. $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$
2. $\langle \boldsymbol{u}, \boldsymbol{v} + \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{u}, \boldsymbol{w} \rangle$
3. Cauchy-Schwarz: $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \, \|\boldsymbol{v}\|$
4. Triangle inequality: $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$

Example (platform revenue): hours watched per genre $\boldsymbol{u} = (500, 200, 50)$ and euro-per-hour rates $\boldsymbol{v} = (2, 3, 5)$ yield total revenue $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 2150$.

The Law of Cosines and angles

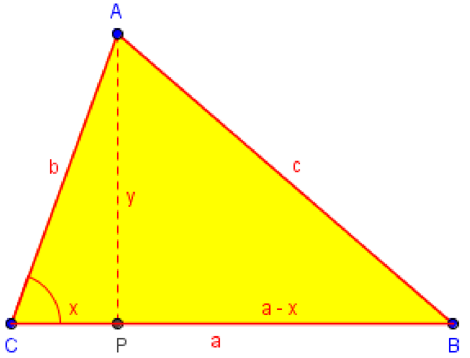
$$\|\boldsymbol{u}-\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 - 2\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|\cos(\angle(\boldsymbol{u},\boldsymbol{v})).$$

Equivalently,

$$\cos(\angle(\boldsymbol{u},\boldsymbol{v})) = \frac{\langle \boldsymbol{u},\boldsymbol{v}\rangle}{\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|}.$$

We prove that $c^2 = a^2 + b^2 - 2ab\cos(C)$:

By Pythagoras, $x^2+y^2 = b^2$ and $(a-x)^2+y^2 = c^2$. Subtract to eliminate y^2 to get $c^2 = a^2+b^2-2ax$. Use $\cos C = x/b$.



Orthogonality

Definition: $\boldsymbol{u} \perp \boldsymbol{v}$ if $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 0$.

Angle example: For $\boldsymbol{u} = (2, -1, 3)$ and $\boldsymbol{v} = (5, -4, -1)$,

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = 11, \quad \cos \angle(\boldsymbol{u}, \boldsymbol{v}) = \frac{11}{\sqrt{14}\sqrt{42}} \approx 0.4537,$$

so the angle is about 63°.

Geometric test: $\boldsymbol{u} \perp \boldsymbol{v}$ iff $\|\boldsymbol{u} - \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$.

Cauchy-Schwarz via completing the square

Assume nonzero $\boldsymbol{u}, \boldsymbol{v}$. For any real t,

$$0 \leq \|\boldsymbol{u} - t\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 - 2t\langle \boldsymbol{u}, \boldsymbol{v}\rangle + t^2\|\boldsymbol{v}\|^2.$$

Complete the square in t:

$$\|\boldsymbol{u}\|^2 - 2t\langle \boldsymbol{u}, \boldsymbol{v}\rangle + t^2\|\boldsymbol{v}\|^2 = \|\boldsymbol{v}\|^2\left(t - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{v}\|^2}\right)^2 + \left(\|\boldsymbol{u}\|^2 - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle^2}{\|\boldsymbol{v}\|^2}\right).$$

Since the left side is ≥ 0 for all t, the second term must be ≥ 0:

$$\langle \boldsymbol{u}, \boldsymbol{v}\rangle^2 \leq \|\boldsymbol{u}\|^2\,\|\boldsymbol{v}\|^2.$$

Equality condition: equality holds iff $\|\boldsymbol{u} - t\boldsymbol{v}\|^2 = 0$ for $t = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{v}\|^2}$, i.e., $\boldsymbol{u} = t\,\boldsymbol{v}$ (collinear vectors).

Linear dependence, independence, and bases

Linear combination: $\boldsymbol{u} = \sum_{i=1}^{m} \lambda_i \boldsymbol{v}_i$.

Linear independence: $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ is linearly independent if

$$\sum_{i=1}^{m} \lambda_i \boldsymbol{v}_i = \mathbf{0} \quad \Rightarrow \quad \lambda_1 = \cdots = \lambda_m = 0.$$

Basis: Any n linearly independent vectors in $\mathbb{R}^n$ form a basis. Then every $\boldsymbol{u} \in \mathbb{R}^n$ has a unique representation $\boldsymbol{u} = \sum_{i=1}^{n} \lambda_i \boldsymbol{v}_i$.

Standard basis: $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$.

Subspaces, span, and dimension

For $E = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$, the span is

$$\operatorname{span}(E) = \left\{ \sum_{i=1}^{k} \lambda_i \boldsymbol{v}_i : \ \lambda_i \in \mathbb{R} \right\}.$$

A subspace $V \subset \mathbb{R}^n$ is any span. A basis of V is a linearly independent set that spans V. The dimension of V is the size of any basis.

Exercise: Basis of $V = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : \ x_1 + x_2 + x_3 = 0\}$.

Orthogonal complement and projection

Orthogonal complement

For a subspace $V \subset \mathbb{R}^n$,

$$V^\perp = \{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{v} \rangle = 0 \ \forall \boldsymbol{v} \in V\}.$$

Projection theorem

For any $\boldsymbol{y} \in \mathbb{R}^n$ and subspace V, there is a unique $\hat{\boldsymbol{y}} \in V$ with $\boldsymbol{y} - \hat{\boldsymbol{y}} \in V^\perp$. We call $\hat{\boldsymbol{y}}$ the orthogonal projection of $\boldsymbol{y}$ onto V.

Least squares view: With design matrix $\boldsymbol{X} \in \mathbb{R}^{n \times d}$ and response $\boldsymbol{y}$, the LS fit $\hat{\boldsymbol{y}}$ is the projection of $\boldsymbol{y}$ onto the column space of $\boldsymbol{X}$.

$$\hat{\boldsymbol{y}} = \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}.$$

SMARTNEWS *Keeping you current*

This Guy Learned Linear Algebra in Ten Days, And You Can Too

At MIT you can take a ton of science classes online. And, in true MIT fashion, someone just did them all at an extraordinary speed

By **Rose Eveleth**
SMITHSONIAN.COM
OCTOBER 26, 2012

A matrix $A \in \mathbb{R}^{m \times n}$ is a table with m rows, n columns. The (i, j) entry is a_{ij}.

We recall basic facts about matrices.

Why care: matrices express linear maps, data tables, network flows, input-output models, regressions, transformations, and more.

Read Werner-Sotskov Ch. 7; Simon-Blume Chs. 8-9.

Exercises 7.6, 7.9(b,c,d), 7.12, 7.14(a), 7.16, 7.18

Matrices

A table of numbers with m rows and n columns: $A \in \mathbb{R}^{m \times n}$. The (i,j)-th entry is denoted by a_{ij}.

Special matrices:

- zero matrix: $\mathbf{0}_{m \times n} \in \mathbb{R}^{m \times n}$
- identity matrix: $\mathbb{I}_n \in \mathbb{R}^{n \times n}$.
- transposition: $A^T \in \mathbb{R}^{n \times m}$, $(A^T)_{ij} = a_{ji}$.
- square matrix: $m = n$.
- symmetric matrix: square matrix such that $A = A^T$.
- diagonal matrix: $a_{ij} = 0$ if $i \neq j$.
- lower (upper) trianglar: $a_{ij} = 0$ if $i < j$ $(i > j)$.
- $\boldsymbol{v} \in \mathbb{R}^n$ is treated as $n \times 1$ matrix, $\mathbb{R}^n \simeq \mathbb{R}^{n \times 1}$.

Basic matrix operations

Addition and scalar multiplication are entrywise:

$$(A+B)_{ij} = a_{ij} + b_{ij}, \qquad (\lambda A)_{ij} = \lambda a_{ij}.$$

They satisfy the usual commutative, associative, and distributive laws.

Matrix product: If $A \in \mathbb{R}^{m\times p}$ and $B \in \mathbb{R}^{p\times n}$, then $C = AB \in \mathbb{R}^{m\times n}$ with

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}.$$

Example:

$$\begin{pmatrix} 2 & 3 & 4 & 1 \\ 7 & -1 & 0 & 4 \end{pmatrix} \begin{pmatrix} 2 & 7 \\ 3 & -1 \\ 4 & 0 \\ 1 & 4 \end{pmatrix} = \begin{pmatrix} 30 & 15 \\ 15 & 66 \end{pmatrix}.$$

Algebraic properties

Facts:

1. $(AB)C = A(BC)$
2. $A(B+C) = AB + AC$ and $(A+B)C = AC + BC$
3. Generally **not** commutative: $AB \neq BA$
4. $AI_n = I_m A = A$ (dimensions must match)
5. $(A^T)^T = A$, $(A+B)^T = A^T + B^T$, $(\lambda A)^T = \lambda A^T$, $(AB)^T = B^T A^T$

Remark: AA^T is always symmetric.

Matrix times vector

We treat $\mathbb{R}^n$ as column vectors $\mathbb{R}^{n\times 1}$. If $A \in \mathbb{R}^{m\times n}$ and $\boldsymbol{x} \in \mathbb{R}^n$ then $A\boldsymbol{x} \in \mathbb{R}^m$ is a linear combination of the columns of A with coefficients x_i:

$$A\boldsymbol{x} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n.$$

e.g. **Another look at the LS method**: $\boldsymbol{X} \in \mathbb{R}^{n\times d}$, $\boldsymbol{y} \in \mathbb{R}^n$

$$\text{minimize}_{\beta\in\mathbb{R}^d} \quad \|\boldsymbol{y} - \boldsymbol{X}\beta\|^2$$

Finds the closest point to $\boldsymbol{y}$ in the space spanned by the columns of $\boldsymbol{X}$.

Orthogonal projection

Theorem: Given a vector $\boldsymbol{y} \in \mathbb{R}^n$ and a subspace $V \subset \mathbb{R}^n$ there exists a unique $\hat{\boldsymbol{y}} \in V$ such that $\boldsymbol{y} - \hat{\boldsymbol{y}} \in V^\perp$. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d$ be a basis of $V \subset \mathbb{R}^n$.

$\boldsymbol{X} \in \mathbb{R}^{n \times d}$ with columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_d$.

$\hat{\boldsymbol{y}} \in V$ means $\quad \hat{\boldsymbol{y}} = \boldsymbol{X}\boldsymbol{\lambda} \quad$ for some $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$.

$\boldsymbol{y} - \hat{\boldsymbol{y}} \in V^\perp \quad$ means $\quad \boldsymbol{X}^\top(\boldsymbol{y} - \hat{\boldsymbol{y}}) = \boldsymbol{0}$.

The unique solution: $\quad \boldsymbol{\lambda} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$.

Matrix inverse

A square matrix A is invertible if there exists a matrix A^{-1} such that

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

We want to show:

$$A \in \mathbb{R}^{n\times n} \text{ is invertible} \quad \Longleftrightarrow \quad A\boldsymbol{x} = \mathbf{0}_n \text{ only for } \boldsymbol{x} = \mathbf{0}_n.$$

Two observations:

1. $A\boldsymbol{x} = \mathbf{0}$ only for $\boldsymbol{x} = \mathbf{0}$ iff columns of A are lin. independent.
2. n independent vectors in $\mathbb{R}^n$ form a basis and so $\forall i = 1, \ldots, n \; \exists \boldsymbol{b}_i$ such that $A\boldsymbol{b}_i = \boldsymbol{e}_i$.
3. This gives $B \in \mathbb{R}^{n\times n}$ such that $AB = \mathbb{I}_n$.

This is enough to show that the matrix $\boldsymbol{X}^\top \boldsymbol{X}$ on slide 29 is invertible.

Note: If $(\boldsymbol{X}^\top \boldsymbol{X})\boldsymbol{x} = \mathbf{0}$ then $\boldsymbol{x}^\top(\boldsymbol{X}^\top \boldsymbol{X})\boldsymbol{x} = \mathbf{0}$ but this only possible if $\boldsymbol{x} = \mathbf{0}$.

Important spaces and rank

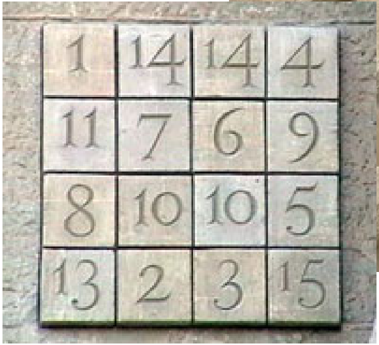
For $A \in \mathbb{R}^{m \times n}$:

- Column space (image) $\mathrm{Im}(A) = \{A\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\} \subset \mathbb{R}^m$
- Row space $\mathrm{Im}(A^T) \subset \mathbb{R}^n$
- Kernel (null space) $\ker(A) = \{\boldsymbol{x} : \ A\boldsymbol{x} = \mathbf{0}\} \subset \mathbb{R}^n$
- Rank $\mathrm{rank}(A) = \dim \mathrm{Im}(A) = \dim \mathrm{Im}(A^T)$

Orthogonality: Row space is orthogonal to $\ker(A)$.

Rank-nullity: $\mathrm{rank}(A) + \dim \ker(A) = n$.

Try: $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}$.


1 14 14 4
11 7 6 9
8 10 10 5
13 2 3 15

Orthogonal matrices

An $n \times n$ matrix A is called orthogonal if $AA^\top = I_n$.

Remark: If A is orthogonal, its row vectors (and also its column vectors) are pairwise orthogonal unit vectors.
Proof: Let $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ be the i-th and j-th rows of A. The (i, j) entry of $AA^\top$ is the scalar product $\boldsymbol{r}_i^\top \boldsymbol{r}_j$. If $AA^\top = I$, then $\boldsymbol{r}_i^\top \boldsymbol{r}_j = 0$ for $i \neq j$ (orthogonality) and $\boldsymbol{r}_i^\top \boldsymbol{r}_i = \|\boldsymbol{r}_i\|^2 = 1$ (unit length). The same holds for columns using $A^\top A = I_n$.

Example:

$$\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Elementary matrix operations

Column or row operations:

1. Swap two columns (or rows)
2. Scale a column (or row) by $\lambda \neq 0$
3. Add a multiple of one column (or row) to another

Each is implemented by multiplying by a suitable elementary matrix on the right (for column ops) or left (for row ops). Useful for Gaussian elimination and determinant computation.

Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and multiply from the right by $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $\begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix}$, or $\begin{bmatrix} 1 & 0 \\ \lambda & 1 \end{bmatrix}$.

Determinants: definition

For $A \in \mathbb{R}^{n\times n}$, let A_{ij} be the submatrix with row i and column j removed. Define recursively

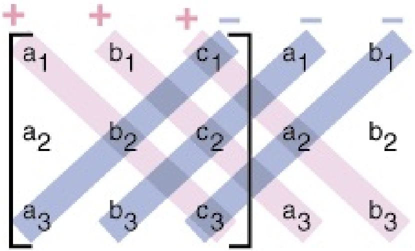
$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}), \quad \det([a_{11}]) = a_{11}.$$

For $n = 2$: $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$.

For $n = 3$:

$$\det(A) = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

A memory aid for $n = 3$



$$\det A = (a_1 b_2 c_3 + b_1 c_2 a_3 + c_1 a_2 b_3) - (a_3 b_2 c_1 + b_3 c_2 a_1 + c_3 a_2 b_1)$$

Cofactor expansion and properties

Cofactor expansion: expand by any row i or any column j:

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

Properties:

1. $\det(A) = \det(A^T)$
2. If A is triangular, $\det(A)$ is the product of diagonal entries
3. $\det(AB) = \det(A)\det(B)$
4. Swapping two rows (or columns) flips the sign of det
5. Scaling a row (or column) by λ scales det by λ
6. Adding a multiple of one row to another leaves det unchanged
7. $\det(A) = 0$ iff rows (or columns) are linearly dependent

Determinants by elimination

Use row operations (keeping track of determinant changes) to reach triangular form.

$$\begin{vmatrix} 1 & 2 & 0 \\ 3 & 0 & 1 \\ 1 & 2 & 3 \end{vmatrix} \to \begin{vmatrix} 1 & 2 & 0 \\ 0 & -6 & 1 \\ 1 & 2 & 3 \end{vmatrix} \to \begin{vmatrix} 1 & 2 & 0 \\ 0 & -6 & 1 \\ 0 & 0 & 3 \end{vmatrix} = (-6) \cdot 3 = -18.$$

First: $r_2 \leftarrow r_2 - 3r_1$. Then: $r_3 \leftarrow r_3 - r_1$.

Geometric meaning: $|\det(A)|$ is the area/volume scaling of the linear map $x \mapsto Ax$ (and its sign encodes orientation).

Linear systems and Cramer's rule

A system $A\boldsymbol{x} = \boldsymbol{b}$ with $A \in \mathbb{R}^{n\times n}$, unknown $\boldsymbol{x} \in \mathbb{R}^n$, and data $\boldsymbol{b}$. If A is nonsingular ($\det A \neq 0$), the solution is unique.

Cramer's rule: Let $A_j(\boldsymbol{b})$ be A with column j replaced by $\boldsymbol{b}$. Then

$$x_j = \frac{\det A_j(\boldsymbol{b})}{\det A}, \quad j = 1, \ldots, n.$$

Note: Great for theory, not used for large-scale computation.

Example:

$$A = \begin{pmatrix} 2 & 3 & 5 \\ 1 & 0 & 2 \\ -1 & -4 & 2 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ 7 \\ 4 \end{pmatrix} \Rightarrow \boldsymbol{x} = \left(\tfrac{75}{8}, -\tfrac{63}{16}, -\tfrac{19}{16}\right).$$

Linear mappings

A mapping $A : \mathbb{R}^n \to \mathbb{R}^m$ is linear if

$$A(\boldsymbol{x}_1 + \boldsymbol{x}_2) = A(\boldsymbol{x}_1) + A(\boldsymbol{x}_2), \qquad A(\lambda \boldsymbol{x}) = \lambda A(\boldsymbol{x}).$$

Then there exists an $m \times n$ matrix (also denoted A) with $A(\boldsymbol{x}) = A\boldsymbol{x}$.

Columns as images: $A(\boldsymbol{e}_i) = \boldsymbol{a}_i$ (the i-th column), and $A(\boldsymbol{x}) = \sum_i x_i A(\boldsymbol{e}_i)$.

Examples: scalings, rotations, reflections, projections, feature maps in ML, Leontief input-output in economics.

Two simple linear maps

1. Reflection across the y-axis:

$$A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix}.$$

2. Rotation by 45° counterclockwise:

$$A = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Check the images of $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$ to see the action.

Inverse matrix and properties

A square A is invertible if A^{-1} exists with $AA^{-1} = A^{-1}A = I_n$.

Properties:

1. $(A^{-1})^{-1} = A$
2. $(AB)^{-1} = B^{-1}A^{-1}$
3. $(A^T)^{-1} = (A^{-1})^T$
4. $(\lambda A)^{-1} = \lambda^{-1}A^{-1}$ for $\lambda \neq 0$
5. $\det(A^{-1}) = 1/\det(A)$

Solve $A\boldsymbol{x} = \boldsymbol{b}$ by $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ when A is invertible.

Computing inverses

Cofactor formula: If A is nonsingular,

$$A^{-1} = \frac{1}{\det A} C(A)^T, \quad C(A)_{ij} = (-1)^{i+j} \det(A_{ij}).$$

For 2×2:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

For larger n, numerical methods use elimination (LU), not cofactors.

Example: Input-output model

Let a_{ij} be units of good i needed to produce 1 unit of good j. Put $A = (a_{ij})$. Let $\boldsymbol{x}$ be total output and $\boldsymbol{y}$ the vector of final demand.

Accounting identity: output = internal demand + final demand

$$\boldsymbol{x} = A\boldsymbol{x} + \boldsymbol{y} \quad \Leftrightarrow \quad (I_n - A)\boldsymbol{x} = \boldsymbol{y} \quad \Rightarrow \quad \boldsymbol{x} = (I_n - A)^{-1}\boldsymbol{y}$$

provided $I_n - A$ is invertible.

Interpretation: $(I - A)^{-1} = I + A + A^2 + \cdots$ accumulates direct, indirect, and higher-order input needs when it converges.

A triangular example

Theorem: If A is strictly upper triangular (zeros on and below diagonal), then $A^n = 0$ and

$$(I_n - A)^{-1} = I_n + A + A^2 + \cdots + A^{n-1}.$$

Check $(I - A)(I + A + \cdots + A^{n-1}) = I - A^n = I$.

Example:

$$A = \begin{pmatrix} 0 & 3 & 5 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} 1 \\ 7 \\ 4 \end{pmatrix} \Rightarrow \boldsymbol{x} = (66, 15, 4).$$