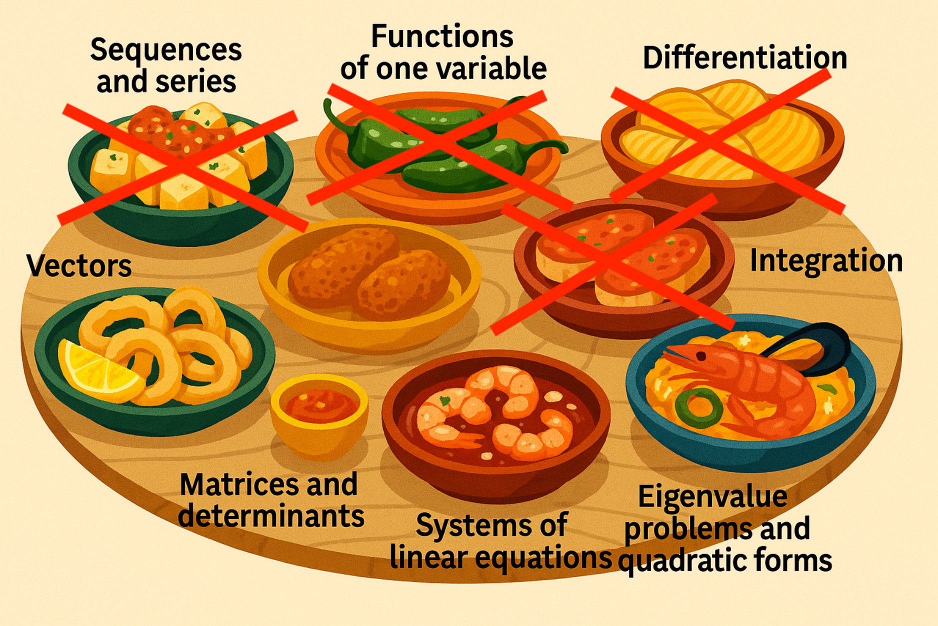
Lecture 3: Calculus and Linear Algebra

Chiara Amorino

Mathematics Brush-up

Barcelona School of Economics

Course content (math tapas)



Orthogonality and linear independence

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: $\boldsymbol{u} \perp \boldsymbol{v}$ if and only if $\|\boldsymbol{u} - \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$.

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 if and only if $\|\boldsymbol{u} - t\boldsymbol{v}\|^2 = 0$ 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

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

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

Say $\boldsymbol{v}_1 + \boldsymbol{v}_2 = \boldsymbol{v}_3$. Show that $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ are NOT linearly independent.

Orthogonal complements and other subspaces

Subspaces, span, and dimension

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

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

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 $\boldsymbol{X}$.

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

Suppose $n = 2$ and $V = \mathrm{span}\{\boldsymbol{u}\}$. Take any $\boldsymbol{y} \in \mathbb{R}^2$.

We have

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

The projection $\boldsymbol{y}$ must satisfy

- $\boldsymbol{y} - \hat{\boldsymbol{y}} \in V^\perp$ or, equivalently, $\langle \boldsymbol{y}, \boldsymbol{u} \rangle = \langle \hat{\boldsymbol{y}}, \boldsymbol{u} \rangle$.
- $\hat{\boldsymbol{y}} \in V$ or, equivalently, $\hat{\boldsymbol{y}} = \lambda \boldsymbol{u}$.

Using these two properties, we get $\hat{\boldsymbol{y}} = \lambda \boldsymbol{u}$, where

$$\lambda \;=\; \frac{\langle \boldsymbol{y}, \boldsymbol{u} \rangle}{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}$$

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

Chapter 6: Matrices and determinants

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

Matrix basics

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**: $\boldsymbol{X} \in \mathbb{R}^{n\times d}$, $\boldsymbol{y} \in \mathbb{R}^n$

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

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

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

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

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

(We still need to show that $\boldsymbol{X}^\top\boldsymbol{X}$ is invertible.)

Matrix inverse

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

$$AA^{-1} = I_n.$$

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. $A\boldsymbol{x} = \mathbf{0}_n$ only for $\boldsymbol{x} = \mathbf{0}_n$ if and only if columns of A are linearly independent.
2. n independent vectors in $\mathbb{R}^n$ form a basis and so $\forall i = 1, \dots, n\ \exists \boldsymbol{b}_i$ 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 15 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 $\boldsymbol{x} = \mathbf{0}$.

Side remark

We defined the inverse of A as the matrix A^{-1} such that $AA^{-1} = I_n$.

But of course, equivalently, $A^{-1}A = I_n$.

Indeed, if $AA^{-1} = I_n$ then $AA^{-1}A = A$ or, equivalently,

$$A(A^{-1}A - I_n) \;=\; \mathbf{0}_{n\times n}.$$

Using the fact that the columns of $A\boldsymbol{x} = \mathbf{0}_n$ if and only if $\boldsymbol{x} = \mathbf{0}_n$, we can conclude that each column of $A^{-1}A - I_n$ is zero, and so

$$A^{-1}A - I_n \;=\; 0$$

as claimed.

Matrix subspaces and decompositions

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) $\ker(A) = \{\boldsymbol{x} : \ A\boldsymbol{x} = \boldsymbol{0}\} \subset \mathbb{R}^n$
- 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


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

$$\begin{matrix}1&14&14&4\\11&7&6&9\\8&10&10&5\\13&2&3&15\end{matrix}\,\mathbf{1}_4 = 33\cdot\mathbf{1}_4 \quad \text{and} \quad \mathbf{1}_4^\top\begin{matrix}1&14&14&4\\11&7&6&9\\8&10&10&5\\13&2&3&15\end{matrix} = 33\cdot\mathbf{1}_4^\top.$$

$$\operatorname{trace}\left(\begin{matrix}1&14&14&4\\11&7&6&9\\8&10&10&5\\13&2&3&15\end{matrix}\right) = 33, \qquad \det\left(\begin{matrix}1&14&14&4\\11&7&6&9\\8&10&10&5\\13&2&3&15\end{matrix}\right) = -1485 = 45\cdot 33.$$

$$\operatorname{rank}\left(\begin{matrix}1&14&14&4\\11&7&6&9\\8&10&10&5\\13&2&3&15\end{matrix}\right) = 4 \qquad \ker\left(\begin{matrix}1&14&14&4\\11&7&6&9\\8&10&10&5\\13&2&3&15\end{matrix}\right) = \{\mathbf{0}_4\}.$$

Row/column space are both equal to $\mathbb{R}^4$

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:

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

Singular value decomposition (SVD)

Singular value decomposition

For every $A \in \mathbb{R}^{m \times n}$ there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ and a "diagonal" matrix $\Lambda \in \mathbb{R}^{m \times n}$ such that

$$A = U \Lambda V^\top.$$

If $m \leq n$ then Λ has m "diagonal" entries.

We can assume $\Lambda_{11} \geq \cdots \geq \Lambda_{mm} \geq 0$.

These diagonal entries are called singular values of A.

Column/row reductions and determinant

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.

To see what is happening, 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

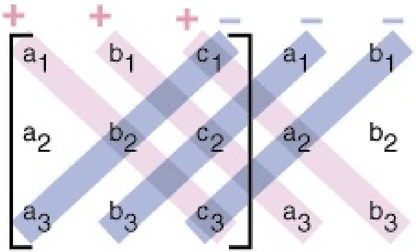
$$\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j}), \quad \det([a_{11}]) = a_{11}.$$

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 determinant unchanged
7. $\det(A) = 0$ if and only if 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 the sign encodes orientation).

2D example:

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \quad \det(A) = 1.$$

The unit square is mapped to a parallelogram of the same area but tilted. If $\det(A) = -1$, the square would be flipped (orientation reversed).

Linear mappings and their inverse

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:

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

Chapter 7: Systems of Linear Equations

Many problems in economics and data science can be modeled as a system of linear equations.

In this chapter we recall core facts and general solution procedures.

Read Chapter 8 of Werner–Sotskov; Chapter 7 of Simon–Blume.

Exercises 8.5, 8.6, 8.8 (Werner–Sotskov).

System of linear equations

Systems of linear equations

Let A be an $m \times n$ matrix, $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$ unknown, and $\mathbf{b} \in \mathbb{R}^m$ given. The system

$$A\mathbf{x} = \mathbf{b}$$

has the vector form

$$\sum_{j=1}^{n} x_j \, \mathbf{a}_j = \mathbf{b},$$

where $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are the columns of A.

If $\mathbf{b} = \mathbf{0}$ the system is homogeneous; otherwise it is inhomogeneous.

An inhomogeneous system can be inconsistent (no solution) or consistent (at least one solution).

A homogeneous system always has $\mathbf{x} = \mathbf{0}$ as a solution.

Motivating example: ranking the web (PageRank)

Modern search engines rank pages by modeling a random surfer: a user who clicks links from page to page.

Let P be the transition matrix where P_{ij} is the probability of moving from page i to j. In the simplest model, the long-run probabilities $\mathbf{x}$ satisfy

$$\mathbf{x} = P^\top \mathbf{x},$$

so $\mathbf{x}$ is the stationary distribution of the Markov chain.

Issue: If a page has no outgoing links, or if the web splits into parts, this system can break down (no unique stationary distribution).

Teleportation and damping

To fix these problems, PageRank introduces two ingredients:

- **Teleportation vector v**: with some probability, the surfer jumps to any page instead of following links. Often chosen uniform: $\mathbf{v} = \frac{1}{n}(1, \ldots, 1)$ so every page is reachable.
- **Damping factor** α: probability of *following links*. Typical choice: $\alpha = 0.85$ (85% follow a link, 15% teleport).

The modified PageRank system is:

$$\left(I - \alpha P^{\top}\right) \mathbf{x} = (1 - \alpha) \mathbf{v}.$$

Result: This ensures a unique ranking vector **x** for any web structure.

Computing the rank with Gaussian elimination

Elementary row or column operations do not change rank(A).

$$\begin{pmatrix} 1 & 2 & 0 \\ 4 & 6 & 2 \\ 3 & 2 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 0 \\ 0 & -2 & 2 \\ 0 & -4 & 4 \end{pmatrix} \sim \begin{pmatrix} 1 & 2 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

The pivot in each nonzero row is the first nonzero entry from the left.

There are 2 pivots, hence $\operatorname{rank}(A) = 2$.

To find a linear relation among columns, solve the homogeneous system for $(\lambda_1, \lambda_2, \lambda_3)$:

$$\lambda_1 + 2\lambda_2 = 0, \qquad -\lambda_2 + \lambda_3 = 0.$$

Thus $(\lambda_1, \lambda_2, \lambda_3) = (-2\lambda_3, \lambda_3, \lambda_3)$ and, e.g., for $\lambda_3 = 1$,

$$-2\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3 = \mathbf{0}.$$

Note that this relation holds for all three matrices above. Why?