Lecture 3: Calculus and Linear Algebra

Piotr Zwiernik

Mathematics Brush-up

Barcelona School of Economics

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).

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 solving a linear system. Build a row-stochastic transition matrix P where P_{ij} is the probability to move from page i to page j. The long-run visit probabilities $\mathbf{x}$ solve

$$\left(I - \alpha P^\top\right)\mathbf{x} = (1-\alpha)\,\mathbf{v}, \quad 0 < \alpha < 1,$$

with $\mathbf{v}$ a teleport vector (often uniform).

Mini-web example: pages A, B, C, D with links $A \to \{B, C\}$, $B \to C$, $C \to A$, $D \to \{A, C\}$. Then

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}, \quad \alpha = 0.85,\ \mathbf{v} = \tfrac{1}{4}(1,1,1,1)^\top.$$

Solving gives

$$\mathbf{x} \approx (0.380, 0.199, 0.384, 0.038),$$

so C and A rank highest.

Node D is rarely visited because nobody links to it (only teleportation contributes).

Motivating example: deblurring a photo

Phone photos often suffer from motion blur. Blurring is (approximately) linear:

$$b = A\,x + \text{noise},$$

where x is the unknown sharp image (vectorized), b is the observed blurred image, and A encodes the blur kernel (convolution).

- Ideal case: noise-free, known $A \Rightarrow$ solve the linear system $Ax = b$.
- Realistic case: noisy or overdetermined $\Rightarrow$ solve a least-squares problem

  $$\min_x \; \|Ax - b\|^2 + \lambda \|x\|^2,$$

  which leads to the normal equations $(A^\top A + \lambda I_n)x = A^\top b$.

Same linear algebra ideas: vector form, consistency, rank, and algorithms to solve large systems.

Rank and existence/uniqueness

Theorem (Column rank = Row rank)

The maximum number of linearly independent columns of A equals the maximum number of linearly independent rows. This common number is the rank of A, denoted $\operatorname{rank}(A)$.

Fact: $\operatorname{rank}(A)$ equals the largest order of a minor (determinant of a square submatrix) that is nonzero.

Example:

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 4 & 6 & 2 \\ 3 & 2 & 4 \end{pmatrix}$$

Here $\det(A) = 0$, so $\operatorname{rank}(A) \leq 2$. Since $\det(A_{33}) = \det\begin{pmatrix} 1 & 2 \\ 4 & 6 \end{pmatrix} = -2 \neq 0$, we get $\operatorname{rank}(A) = 2$.

Note $2\mathbf{a}_1 - \mathbf{a}_2 = \mathbf{a}_3$, but no single column is a multiple of another. Any two of the three columns form a basis of the column space.

Computing the rank with Gaussian elimination

Elementary row or column operations do not change $\operatorname{rank}(A)$.

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

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 $\lambda_3 = 1$,

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

Augmented matrix; consistency

The augmented matrix of the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is $A_{\mathbf{b}} = \begin{bmatrix} A & \mathbf{b} \end{bmatrix} \in \mathbb{R}^{m\times(n+1)}$.

Remark: Either $\operatorname{rank}(A_{\mathbf{b}}) = \operatorname{rank}(A)$ or $\operatorname{rank}(A_{\mathbf{b}}) = \operatorname{rank}(A) + 1$.

Theorem (Rouch–Capelli): $A\mathbf{x} = \mathbf{b}$ is consistent $\iff \operatorname{rank}(A_{\mathbf{b}}) = \operatorname{rank}(A)$ (equivalently $\mathbf{b} \in \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$).

Moreover:

1. If $\operatorname{rank}(A) = \operatorname{rank}(A_{\mathbf{b}}) = n$, the solution is unique. Q: What is it?
2. If $\operatorname{rank}(A) = \operatorname{rank}(A_{\mathbf{b}}) < n$, there are infinitely many solutions; choose $n - \operatorname{rank}(A)$ variables free.

Triangularization (Gaussian elimination)

Theorem

Applying elementary row operations to $A_{\mathbf{b}}$ transforms any system into an equivalent upper-triangular one.

This follows essentially from the fact that $A\boldsymbol{x} = \boldsymbol{b} \Leftrightarrow A_b \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix} = \mathbf{0} \Leftrightarrow CA_b \begin{bmatrix} \mathbf{x} \\ -1 \end{bmatrix} = \mathbf{0}$ for any invertible C

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 2 \\ 4x_1 + 6x_2 - x_3 &= 9 \end{aligned} \qquad \Longleftrightarrow \qquad \begin{aligned} x_1 + x_2 + x_3 &= 3 \\ -2x_2 + x_3 &= -1 \\ -4x_3 &= -4 \end{aligned}$$

Row ops: $r_2 \leftarrow r_2 - r_1$, $r_3 \leftarrow r_3 - 4r_1$, $r_3 \leftarrow r_3 + r_2$.

We have $\text{rank}(A) = \text{rank}(A_{\mathbf{b}}) = 3 \Rightarrow$ unique solution $\mathbf{x} = (1, 1, 1)$.

A homogeneous example

Consider

$$\begin{aligned} a+b+c+d+e+f &= 0, \\ 2a+2b+2c+2d-e-f &= 0, \\ 3a+3b-c-d-e-f &= 0. \end{aligned}$$

Row-reducing the coefficient matrix gives

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

so there are 3 pivots, rank $= 3$, and the solution space has dimension $6 - \text{rank} = 3$:

$$(a, -a, c, -c, e, -e).$$

A basis is $\{(1, -1, 0, 0, 0, 0), (0, 0, 1, -1, 0, 0), (0, 0, 0, 0, 1, -1)\}$.

General solution structure

Theorem: Every solution of $A\mathbf{x} = \mathbf{b}$ can be written as

$$\mathbf{x} = \mathbf{x}_h + \mathbf{x}_p,$$

where $\mathbf{x}_h$ is the general solution of $A\mathbf{x} = \mathbf{0}$ and $\mathbf{x}_p$ is any particular solution of $A\mathbf{x} = \mathbf{b}$.

Corollary: If $A\mathbf{x} = \mathbf{b}$ is consistent, it has as many solutions as $A\mathbf{x} = \mathbf{0}$.

Example: Three identical equations

$$x_1 + x_2 + x_3 = 1$$

yield $\operatorname{rank}(A) = \operatorname{rank}(A_{\mathbf{b}}) = 1$. Choose two free variables:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \lambda_1 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R}.$$

Two quick examples

1. For $A = \begin{pmatrix} 1 & 2 & 0 \\ 4 & 6 & 2 \\ 3 & 2 & 4 \end{pmatrix}$ we found $\operatorname{rank}(A) = 2$. Thus $\dim\ker(A) = 3 - 2 = 1$ and

$$\ker(A) = \{\lambda(-2, 1, 1) : \ \lambda \in \mathbb{R}\}.$$

2. Consider

$$\begin{aligned} x_1 + x_2 + x_3 &= -1, \\ x_1 + 2x_2 + 4x_3 &= 2. \end{aligned}$$

Row reduction gives $x_1 + x_2 + x_3 = -1, \quad x_2 + 3x_3 = 3$, so $\operatorname{rank}(A) = \operatorname{rank}(A_{\mathbf{b}}) = 2$. One free variable:

$$\{\, \lambda(2, -3, 1) + (-4, 3, 0) : \ \lambda \in \mathbb{R} \,\}.$$

Finding a basis for a span

Let

$$\mathbf{v}_1 = \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix}.$$

Form the 3×3 matrix with these as columns:

$$\begin{pmatrix} 2 & -1 & 0 \\ 3 & 0 & -1 \\ 0 & 3 & -2 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & -1/3 \\ 0 & 1 & -2/3 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus rank $= 2$ and $3\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}$. A convenient basis of $\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ -1/3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -2/3 \end{pmatrix} \right\}$$

(or, equivalently, two of the original vectors that are independent).

Standing on the shoulders of giants

"If I have seen further, it is by standing on the shoulders of giants." **Isaac Newton (1675)**.

Some of the giants of linear/matrix algebra:

- Gauss (Gaussian elimination, least squares, determinants)
- Cauchy (CauchySchwarz, eigenvalues, determinants)
- Laplace (Laplace expansion of determinants)
- Cayley (CayleyHamilton theorem, matrix theory)
- Sylvester (matrix rank, invariant theory)
- Jordan (Jordan canonical form)
- Frobenius (matrix factorizations, linear algebra foundations)

(pictured here as the base of a *castell*)

Chapter 8: Eigenvalues and Quadratic Forms

Eigenvalues/eigenvectors appear in growth models, Markov chains, PCA, and stability of dynamical systems.

Quadratic forms drive optimization via second-order conditions.

Read Chapter 10 of Werner–Sotskov (optional: Simon–Blume, Ch. 7).

Exercises 10.2; 10.5 A,B,C; 10.6(a,b).

Eigenvalues and eigenvectors: migration example

Consider a population split between two regions: city (x_t^C) and countryside (x_t^R) at time t. Suppose each year a fraction of people move between the two regions:

$$\begin{pmatrix} x_{t+1}^C \\ x_{t+1}^R \end{pmatrix} = \begin{pmatrix} 0.8 & 0.4 \\ 0.2 & 0.6 \end{pmatrix} \begin{pmatrix} x_t^C \\ x_t^R \end{pmatrix} = A \begin{pmatrix} x_t^C \\ x_t^R \end{pmatrix}.$$

Here, 0.8 means that 80% of city dwellers stay in the city, while 0.4 means that 40% of rural dwellers move to the city in the next period.

If the urban/rural ratio stabilizes, the next state must be a constant multiple of the current state:

$$\begin{pmatrix} x_{t+1}^C \\ x_{t+1}^R \end{pmatrix} = \lambda \begin{pmatrix} x_t^C \\ x_t^R \end{pmatrix}.$$

Thus, the long-run stable distribution of the population is given by an eigenvector of A, and λ is the associated eigenvalue describing the growth rate.

Eigenvalues and eigenvectors: definitions

For $A \in \mathbb{R}^{n\times n}$, $\lambda \in \mathbb{R}$ is an eigenvalue if there exists $\mathbf{x} \neq \mathbf{0}$ with

$$A\mathbf{x} = \lambda\mathbf{x} \quad \Longleftrightarrow \quad (A - \lambda I_n)\mathbf{x} = \mathbf{0}.$$

Then $\mathbf{x}$ is an eigenvector for λ.

Characterization:

$$\lambda \text{ eigenvalue} \iff \operatorname{rank}(A - \lambda I_n) < n \iff \det(A - \lambda I_n) = 0.$$

The characteristic polynomial $P(\lambda) = \det(A - \lambda I_n)$ has degree n; the real roots (at most n) are the eigenvalues.

Procedure:

1. Solve $\det(A - \lambda I_n) = 0$ to get eigenvalues.
2. For each λ, solve $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$ for eigenvectors.

Cohort-mix example (continued)

With $A = \begin{pmatrix} 0.8 & 0.4 \\ 0.3 & 0.9 \end{pmatrix}$,

$$(0.8 - \lambda)(0.9 - \lambda) - 0.12 = 0 \iff \lambda^2 - 1.7\lambda + 0.6 = 0.$$

The solutions are $\lambda_1 = 1.2$ and $\lambda_2 = 0.5$.

For $\lambda_1 = 1.2$, eigenvectors are multiples of $(1, 1)$. For $\lambda_2 = 0.5$, eigenvectors are multiples of $(1, -\frac{3}{4})$.

For a population-growth interpretation we keep the dominant eigenpair: equal cohort proportions grow by 20% per step.

Basic spectral facts

If $A \in \mathbb{R}^{n \times n}$,

1. Eigenvectors associated with distinct eigenvalues are linearly independent.
2. If A is symmetric, then all eigenvalues are real and eigenvectors for distinct eigenvalues are orthogonal (Spectral Theorem).

Example:

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

Then $P(\lambda) = \lambda^2 - 2\lambda - 3$, so $\lambda_1 = 3$, $\lambda_2 = -1$ with eigenvectors along $(1, 1)$ and $(-1, 1)$, respectively.

Repeated eigenvalues

Consider

$$A = \begin{pmatrix} 0 & -1 & 1 \\ -7 & 0 & 5 \\ -5 & -2 & 5 \end{pmatrix}.$$

The characteristic polynomial is

$$p_A(\lambda) = \lambda^3 - 5\lambda^2 + 8\lambda - 4 = (\lambda - 2)^2(\lambda - 1).$$

Hence the eigenvalues are $\lambda = 2$ (algebraic multiplicity 2) and $\lambda = 1$ (algebraic multiplicity 1).

Eigenspaces:

$$E_2 = \ker(A - 2I) = \operatorname{span}\{(1, -1, 1)\}, \qquad E_1 = \ker(A - I) = \operatorname{span}\{(2, 1, 3)\}.$$

Since $\dim E_2 = 1 < 2$, A is defective (not diagonalizable).

Application: Principal Component Analysis (PCA)

We observe n data points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$.

Step 1 (Centering): Arrange the data in a matrix $\boldsymbol{X} \in \mathbb{R}^{n \times d}$ with rows $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Define the centering matrix $H = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ and then the centered data is $H\boldsymbol{X}$.

Step 2 (Sample covariance):

$$S = \frac{1}{n}(H\boldsymbol{X})^\top (H\boldsymbol{X}).$$

Step 3 (Spectral decomposition):

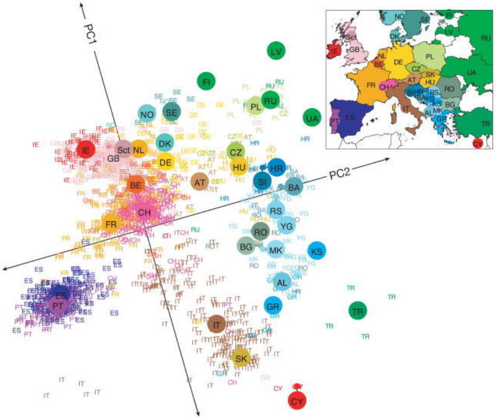
$$S = U\Lambda U^\top \quad \text{and define } \boldsymbol{Y} = H\boldsymbol{X}U \quad \text{(rotate the centered data).}$$

The sample covariance of $\boldsymbol{Y}$ is $\frac{1}{n}\boldsymbol{Y}^\top \boldsymbol{Y} = U^\top S U = \Lambda$ (diagonal!).

The columns of U are principal directions and diagonal entries of Λ are the explained variances.

Sorting eigenvalues in decreasing order ranks components by variance captured.

PCA: why it matters



Dimensionality reduction: compress high-dimensional data into a few directions that preserve most variability.

Noise filtering: small eigenvalues $\rightarrow$ directions with little signal.

Visualization: projecting onto first 2 PCs often reveals hidden structure.

Nice example: genetic variation in Europeans plotting individuals on the first two PCs of SNP data *recovers the geographic map of Europe*!

Numerical toy example

Covariance matrix:

$$S = \begin{pmatrix} 72.9 & 87.5 & 6.875 \\ 87.5 & 108.3 & 8.75 \\ 6.875 & 8.75 & 0.73 \end{pmatrix}.$$

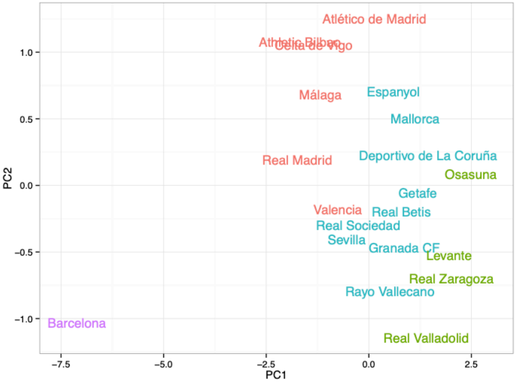
Spectral decomposition:

$$S = U\Lambda U^\top, \quad \Lambda = \mathrm{diag}(180,\ 1.38,\ 0.005).$$

$\Rightarrow$ The first component explains $> 99\%$ of the variance; the last can be safely ignored.

PCA & Unique football styles

"*Mille viae ducunt Barcinonem*"
(A thousand roads lead to Barcelona)



"Searching for a Unique Style in Soccer" (Gyarmati, Kwak & Rodriguez, 2014), studies sequential pass patterns in a team's passing.

They defined "flow motifs" as statically significant pass subsequences (e.g., $A \to B \to C \to A$).

Applied motif analysis to compare styles across teams and leagues. The data are projected to the first two principal components.

Quadratic forms

A quadratic form is $Q(\mathbf{x}) = \mathbf{x}^\top A\mathbf{x} = \sum_{i,j=1}^n a_{ij}x_ix_j$ with $A \in \mathbb{R}^{n\times n}$ symmetric.

Remark: Since the coefficient of x_ix_j is $a_{ij} + a_{ji}$, every quadratic form has a unique symmetric representative (replace A by $\frac{1}{2}(A + A^\top)$).

Example: $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ and $\begin{pmatrix} 1 & 5/2 \\ 5/2 & 4 \end{pmatrix}$ define the same $Q(x, y) = x^2 + 5xy + 4y^2$.

Signs of quadratic forms

For symmetric A:

1. Positive definite: $\mathbf{x}^\top A\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
2. Negative definite: $\mathbf{x}^\top A\mathbf{x} < 0$ for all $\mathbf{x} \neq \mathbf{0}$.
3. Positive semidefinite: $\mathbf{x}^\top A\mathbf{x} \geq 0$ for all $\mathbf{x}$.
4. Negative semidefinite: $\mathbf{x}^\top A\mathbf{x} \leq 0$ for all $\mathbf{x}$.
5. Indefinite: takes both positive and negative values.

Example:

$$\mathbf{x}^\top \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \mathbf{x} = (x-y)^2 \geq 0$$

so the matrix is positive semidefinite.

Eigenvalue and minor tests

Spectral test (symmetric A):

- PD $\iff$ all eigenvalues > 0; ND $\iff$ all < 0;
- PSD $\iff$ all ≥ 0; NSD $\iff$ all ≤ 0;
- Indefinite $\iff$ eigenvalues of different signs.

Sylvesters criterion (leading principal minors D_k):

- PD $\iff$ $D_k > 0$ for $k = 1, \ldots, n$;
- ND $\iff$ $(-1)^k D_k > 0$ for $k = 1, \ldots, n$.

Example: $A = \begin{pmatrix} -3 & 2 & 0 \\ 2 & -3 & 0 \\ 0 & 0 & -5 \end{pmatrix}$ has $D_1 = -3$, $D_2 = 5$, $D_3 = -25 \Rightarrow$ negative definite.

Examples

1. $A = \begin{pmatrix} 1 & 4 & 6 \\ 4 & 2 & 1 \\ 6 & 1 & 6 \end{pmatrix}$: $D_1 = 1,\ D_2 = -14,\ D_3 = -109 \Rightarrow$ indefinite.
2. $A = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 3 \end{pmatrix}$: $D_1 = 3,\ D_2 = 5,\ D_3 = 12 \Rightarrow$ positive definite.
3. The symmetric matrix of $Q(x, y, z) = xy + yz$ is indefinite since $D_1 = D_3 = 0$ and $D_2 = -\frac{1}{4}$.