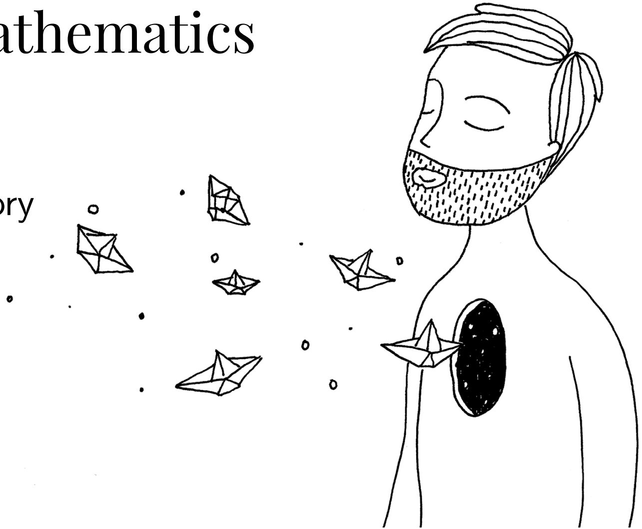
4509 - Bridging Mathematics

Lecture 3: Linear Algebra, Projection and Spectral Theory



PAULO FAGANDINI

Objectives

- Formalize vector spaces, subspaces, linear independence, basis and dimension.
- Master the geometry of inner-product spaces and **orthogonal projection**, the mathematical core of OLS and Frisch-Waugh-Lovell.
- Understand the spectral decomposition of symmetric matrices, eigenvalues, and the definiteness of quadratic forms.

The spine: OLS is not algebra, it is geometry. $\hat{\beta} = (X^\top X)^{-1} X^\top y$ is just the perpendicular dropped from y onto the span of the regressors; the Hessian's eigenvalues are how we read curvature.

Vector spaces and the four subspaces

A **subspace** W of a vector space V is a subset closed under linear combinations: $u, v \in W$ and scalars $\alpha, \beta \Rightarrow \alpha u + \beta v \in W$. **Span**, **linear independence** ($\sum \alpha_i v_i = 0 \Rightarrow$ all $\alpha_i = 0$), **basis** (independent spanning set) and **dimension** carry over as usual.

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

- Column space $\mathrm{Im}(A) = \{Ax\} \subseteq \mathbb{R}^m$, $\dim = \mathrm{rank}(A)$.
- Null space $\mathrm{Ker}(A) = \{x : Ax = 0\} \subseteq \mathbb{R}^n$, $\dim = \mathrm{nullity}(A)$.

Theorem (Rank-nullity)

$\mathrm{rank}(A) + \mathrm{nullity}(A) = n$.

$\mathrm{Ker}(A)$ is the directions A flattens to zero; $\mathrm{Im}(A)$ is where A can reach. Rank-nullity is conservation of dimension: what you lose in the kernel you keep in the image.

Full column rank and OLS

Full column rank of the design matrix X means:

$$\text{no perfect multicollinearity} \iff \mathrm{Ker}(X) = \{0\} \iff X^\top X \text{ invertible.}$$

That is exactly the condition for OLS to have a unique solution, which we derive next.

Inner products and orthogonality

On $\mathbb{R}^n$: $\langle x, y\rangle = x^\top y = \sum_i x_i y_i$, with norm $\|x\| = \sqrt{\langle x, x\rangle}$.

- **Cauchy-Schwarz:** $|\langle x, y\rangle| \leq \|x\|\,\|y\|$, equality iff collinear. Since $\cos\theta = \frac{\langle x,y\rangle}{\|x\|\|y\|}$, the inequality just says a cosine cannot exceed 1.
- **Orthogonal:** $x \perp y \iff \langle x, y\rangle = 0$.
- **Pythagoras:** $x \perp y \Rightarrow \|x + y\|^2 = \|x\|^2 + \|y\|^2$.

Cauchy-Schwarz is where $|\operatorname{corr}(X, Y)| \leq 1$ comes from: on Friday you re-read it with $\langle X, Y\rangle = E[XY]$ and get exactly the correlation bound.

The projection theorem

Problem. Given a subspace $S \subseteq \mathbb{R}^n$ and a point y, find $\hat{y} \in S$ minimizing $\|y - s\|$ over $s \in S$.

Theorem (Orthogonal projection)

*Let S be a closed subspace of a Hilbert space H. For any $y \in H$ there exists a unique $\hat{y} \in S$, the **orthogonal projection** of y onto S, minimizing the distance to y:*

$$\|y - \hat{y}\| = \inf_{s \in S} \|y - s\|.$$

*This minimizer is fully characterized by the **orthogonality condition**:*

$$(y - \hat{y}) \perp S \quad (i.e., \langle y - \hat{y}, s \rangle = 0 \ \ \forall s \in S).$$

From geometry to the OLS formula

Let $S = \mathrm{col}(X)$, $X \in \mathbb{R}^{n\times k}$ full column rank, $\hat{y} = X\beta$. Orthogonality says the residual is perpendicular to every column of X:

$$X^\top(y - X\beta) = 0 \implies \underbrace{X^\top X\,\beta = X^\top y}_{\text{normal equations}} \implies \boxed{\hat{\beta} = (X^\top X)^{-1}X^\top y.}$$

Define the **hat matrix** $P = X(X^\top X)^{-1}X^\top$ and the **residual maker** $M = I - P$:

- P is symmetric ($P^\top = P$) and idempotent ($P^2 = P$): projecting twice equals projecting once;
- M is symmetric idempotent, $PM = 0$, and $\hat{y} = Py$, $e = My$.

P casts the shadow; M keeps the part that could not be explained.

Why projection is the whole linear model

$y = \hat{y} + e$ is "data = fit + residual." Orthogonality turns this into

$$\|y\|^2 = \|\hat{y}\|^2 + \|e\|^2 \qquad (\mathsf{TSS} = \mathsf{ESS} + \mathsf{RSS}),$$

the ANOVA decomposition, and $R^2 = \|\hat{y}\|^2 / \|y\|^2$ is the cosine^2 of the angle between y and the regressor space.

Econometrics is applied projection geometry.

Frisch-Waugh-Lovell

To get the coefficient on x_1 **controlling for** x_2:

1. regress y on x_2, keep residuals $M_2 y$;
2. regress x_1 on x_2, keep residuals $M_2 x_1$;
3. regress $M_2 y$ on $M_2 x_1$.

You recover the *same* $\hat{\beta}_1$ as the full regression.

"Controlling for x_2" literally means "project it out first, then look at what is left." Partialling-out is orthogonal projection onto the orthogonal complement of x_2, the mental model behind every "holding constant" in applied micro.

Eigenvalues and the spectral theorem

$Av = \lambda v$, $v \neq 0$; $\det(A - \lambda I) = 0$. Trace $= \sum \lambda_i$, det $= \prod \lambda_i$.

Theorem (Spectral theorem, symmetric case)

If $A = A^\top \in \mathbb{R}^{n\times n}$ then all λ_i are real, eigenvectors of distinct eigenvalues are orthogonal, and

$$A = Q\Lambda Q^\top, \qquad Q^\top Q = I, \quad \Lambda = \mathrm{diag}(\lambda_i).$$

A symmetric matrix is just a stretch along n perpendicular axes: Q names the axes, Λ gives the stretch factors.

Quadratic forms and definiteness

$Q(x) = x^\top A x$ (take A symmetric). In the eigenbasis $x = Qz$: $Q(x) = \sum_i \lambda_i z_i^2$. So the sign of the form is dictated by the signs of the eigenvalues:

- **positive definite** $A \succ 0$: all $\lambda_i > 0$ (so $Q(x) > 0 \ \forall x \neq 0$);
- **negative definite** $A \prec 0$: all $\lambda_i < 0$;
- **semidefinite:** allow $= 0$; **indefinite:** mixed signs.

Only the symmetric part matters: for $x^\top A x$, replace A by $\frac{1}{2}(A + A^\top)$.

The curvature bridge to optimization

Second-order Taylor of a C^2 function f:

$$f(x_0 + h) = f(x_0) + \nabla f(x_0)^\top h + \tfrac{1}{2} h^\top H(f, x_0) h + o(\|h\|^2).$$

At a stationary point $\nabla f = 0$, so the **Hessian alone** decides max/min/saddle through its eigenvalues, exactly the quadratic form above:

$$H \prec 0 \;\Rightarrow\; \text{local max}, \quad H \succ 0 \;\Rightarrow\; \text{local min}, \quad H \text{ indefinite} \;\Rightarrow\; \text{saddle}.$$

Curvature $=$ eigenvalues of the Hessian. Next lecture: if H is negative semidefinite *everywhere*, f is concave, and for a concave f the first-order condition is sufficient for a global max, which is why economics assumes concavity. The same reading, applied to a Jacobian, classifies dynamic steady states later this week.

References

- A. de la Fuente, *Mathematical Methods and Models for Economists*, Ch. 3.
- C. P. Simon and L. Blume, *Mathematics for Economists*, Ch. 10, 11 and 16.
- B. E. Hansen, *Econometrics*, Ch. 2 and 3 (linear algebra of OLS).