Introduction to mathematics for economists

Exercise Sheet #2 — Linear Algebra

Alexandre Grellet

2026-08-27

During the tutorials, we will solve exercises 1, 3, 5, 8, 10, and 12. So you must prioritize these. You will be provided a full answer sheet.

Vector Spaces, Span, and Independence

Subspace

A subset $W \subseteq \mathbb{R}^n$ is a **linear subspace** if it is nonempty and closed under the two linear operations:

$$\mathbf{u}, \mathbf{v} \in W \implies \mathbf{u} + \mathbf{v} \in W, \qquad \mathbf{u} \in W,\ c \in \mathbb{R} \implies c\,\mathbf{u} \in W.$$

Equivalently: W contains $\mathbf{0}$ and every linear combination of its elements.

Exercise 1 : Subspaces, cones, and why "linear" earns its name

1. For each set below, decide whether it is a subspace of their definition space. Prove or disprove by giving a counter-example. If it is a subspace, give a basis and its dimension.
   a. $W_1 = \{(x, y) \in \mathbb{R}^2 : x + 2y = 0\}$.
   b. $W_2 = \{(x, y) \in \mathbb{R}^2 : x + 2y = 3\}$.
   c. $W_3 = \{(x, y) \in \mathbb{R}^2 : x \geq 0,\ y \geq 0\}$ (the non-negative quadrant).
   d. $W_4 = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$.
2. The set of feasible consumption bundles is the non-negative quadrant $\mathbb{R}^n_+ = \{x : x_i \geq 0\ \forall i\}$. Explain in one or two sentences why this is a **cone** (i.e. only multiplications by positive scalars allowed) but not a vector space, and why a budget line $\{x : p \cdot x = m\}$ (with $m > 0$) is not one either.
3. Let U, W be subspaces of $\mathbb{R}^n$. Show that $U \cap W$ is always a subspace, but that $U \cup W$ generally is **not**. Give a concrete counterexample for the union in $\mathbb{R}^2$.

Exercise 2 : Spanning, independence, and complete markets

Consider a one-period economy with two future **states of the world**, s_1 and s_2. An asset is described by the vector of its payoffs across states: an asset paying θ_1 in state s_1 and θ_2 in state s_2 is the vector $(\theta_1, \theta_2) \in \mathbb{R}^2$. A **portfolio** is a linear combination of assets (holding c_j units of asset j), and its payoff vector is the same linear combination of payoff vectors. A market is called **complete** if every conceivable payoff profile (θ_1, θ_2) can be manufactured by some portfolio.

Take the risk-free bond $a_1 = (1, 1)$ (pays 1 in each state) and the Arrow security $a_2 = (2, 0)$ (pays 2 in state s_1, nothing in state s_2).

1. Are a_1 and a_2 linearly independent? Do they span $\mathbb{R}^2$? Conclude whether this two-asset market is complete.
2. Now add a third asset $a_3 = (0, 3)$. Explain, without computing, why $\{a_1, a_2, a_3\}$ must be linearly dependent. Then find an explicit non-trivial dependence relation $\alpha a_1 + \beta a_2 + \gamma a_3 = 0$.
3. Find the portfolio in a_1 and a_2 that **replicates** a_3. In one sentence, relate this to the idea that a redundant asset carries no new insurance possibilities.
4. Give a basis for $\mathbb{R}^2$ drawn from these assets and state $\dim \mathbb{R}^2$. State in one line what "complete market" means in the vocabulary of span and dimension.

Orthogonality and Projections

Dot Product, Orthogonality, and Projection

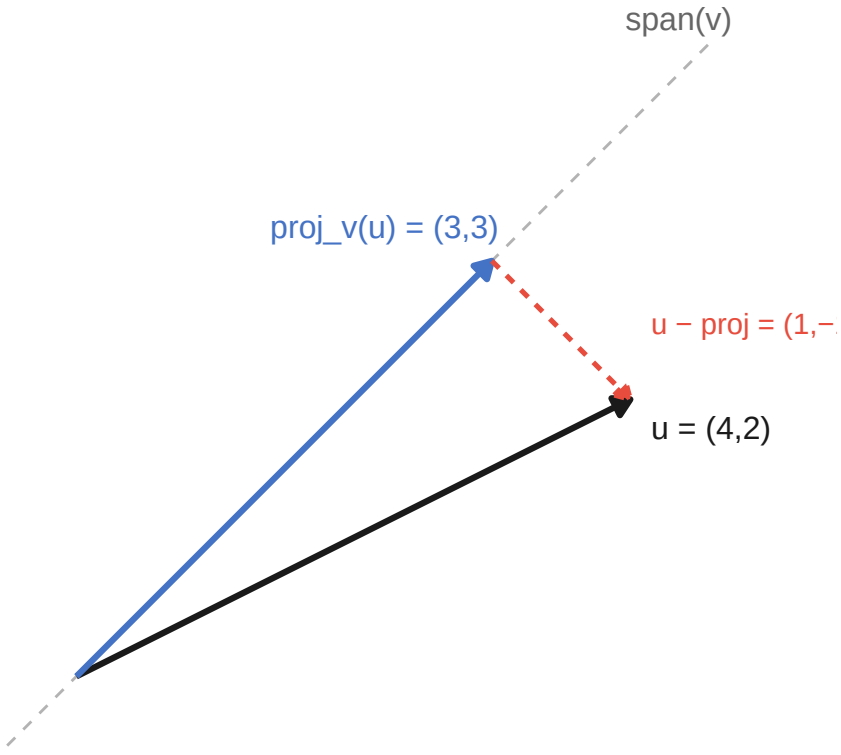
For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, the **dot product** is $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \sum_i u_i v_i$, and $\mathbf{u}, \mathbf{v}$ are **orthogonal** iff $\mathbf{u} \cdot \mathbf{v} = 0$.
The **orthogonal projection** of u onto the line spanned by $v \neq 0$ is

$$\text{proj}_v(u) = \frac{u \cdot v}{v \cdot v} v.$$

More generally, for a subspace $S = \text{col}(A)$ the projection of u onto S is the unique $\hat{u} \in S$ whose residual is orthogonal to all of S; solving $A^\top(u - Ac) = 0$ gives $\hat{u} = A(A^\top A)^{-1}A^\top u = P_S\, u$, with $P_S = A(A^\top A)^{-1}A^\top$.

Exercise 3 : From a line to a subspace to the projection matrix

1. Compute $\text{proj}_v(u)$ for $u = (4, 2)$ and $v = (1, 1)$. Verify explicitly that the residual $u - \text{proj}_v(u)$ is orthogonal to v.
2. Derive the projection formula from scratch. Try not to check the lecture notes: this is just to remind you that the formula is not really worth remembering because it's derived easily :)
3. Now move up to a plane in $\mathbb{R}^3$. Project $u = (1, 2, 0)$ onto $S = \text{span}\{a_1, a_2\}$ with $a_1 = (1, 0, 1)$, $a_2 = (0, 1, 1)$. Use $A = [a_1\ a_2]$, solve the normal equations $A^\top A\, c = A^\top u$, and report $\hat{u}$ and the residual. Check the residual is orthogonal to **both** a_1 and a_2. *Hint: you will need to invert a certain matrix.*
4. Form the projection matrix $P_S = A(A^\top A)^{-1}A^\top$ for the previous part. Verify that (i) P_S is symmetric, (ii) $P_S u$ reproduces the $\hat{u}$ you found, and (iii) $\text{tr}(P_S) = \dim S = 2$.



Exercise 4 : Properties of projection matrices

Let $S = \mathrm{col}(A) \subseteq \mathbb{R}^n$ with A of full column rank, and $P = A(A^\top A)^{-1}A^\top$.

1. Prove P is **idempotent**: $P^2 = P$. (Work directly from the formula.)
2. Prove P is **symmetric**: $P^\top = P$.
3. Let $M = I - P$. Show M is also symmetric and idempotent, and that M projects onto the orthogonal complement $S^\perp$.
4. **The eigenvalues of** P, and why $\mathrm{tr}(P) = \dim S$. We build this up in small steps.
   a. Suppose $Pv = \lambda v$ for some $v \neq 0$. Apply P once more to both sides, use $P^2 = P$ from part 1, and deduce that $\lambda^2 = \lambda$. Conclude that the **only** possible eigenvalues of P are 0 and 1.
   b. Show that all vectors of S have eigenvalue 1, and that if a vector has eigenvalue 1 then it must belong to S. Show similarly that the vectors with eigenvalue 0 are precisely the vectors of $S^\perp$. *(Hint: use part 3, and the fact that P fixes every vector already in S and sends every vector of $S^\perp$ to 0.)*
   c. You may use the following fact: **every symmetric matrix is diagonalisable** (this is the *spectral theorem*). Combining this with (a) and (b), explain why P has exactly $\dim S$ eigenvalues equal to 1 and $n - \dim S$ eigenvalues equal to 0.
   d. Use (c) to conclude
   $$\mathrm{tr}(P) = \mathrm{rank}(P) = \dim S.$$

Linear Maps, Rank, and the Rank-Nullity Theorem

Image, Kernel, Rank

For a linear map $f(x) = Ax$ with $A \in \mathbb{M}_{m\times n}$:

$$\mathrm{Im}(f) = \mathrm{span}\{c_1(A), \dots, c_n(A)\} \subseteq \mathbb{R}^m, \qquad \ker(f) = \{x \in \mathbb{R}^n : Ax = 0\} \subseteq \mathbb{R}^n,$$

and $\mathrm{Rank}(A) = \dim \mathrm{Im}(f)$ = number of linearly independent columns.

Rank–Nullity

For $f : \mathbb{R}^n \to \mathbb{R}^m$,

$$\underbrace{\mathrm{Rank}(f)}_{\dim \mathrm{Im}(f)} + \underbrace{\mathrm{Null}(f)}_{\dim \ker(f)} = n = \dim(\text{domain}).$$

Every independent constraint the map imposes costs one degree of freedom; the nullity is whatever freedom survives.

Exercise 5 : Image, kernel, rank — and the degrees of freedom in OLS

Let

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 5 \end{pmatrix} : \mathbb{R}^3 \to \mathbb{R}^3.$$

1. Find a basis for $\mathrm{Im}(A)$ and state $\mathrm{Rank}(A)$. *(Hint: start by checking dependence.)*
2. Find a basis for $\ker(A)$ and state the nullity.

3. Verify the rank–nullity theorem for A.
4. Is $f(x) = Ax$ injective? Surjective onto $\mathbb{R}^3$? For a given b, how many solutions can $Ax = b$ have — 0, 1, or ∞ — and on what does the count depend?

Exercise 6 : The Leontief input–output economy

An economy has two sectors, **Agriculture** (1) and **Manufacturing** (2). Producing one unit of agricultural output requires 0.1 units of agriculture and 0.3 units of manufacturing as inputs; producing one unit of manufacturing output requires 0.4 units of agriculture and 0.2 units of manufacturing. Households demand 50 units of agriculture and 60 of manufacturing as final goods. Writing a_{ij} for the units of good i needed to make one unit of good j,

$$A = \begin{pmatrix} 0.1 & 0.4 \\ 0.3 & 0.2 \end{pmatrix}, \qquad d = \begin{pmatrix} 50 \\ 60 \end{pmatrix}.$$

1. Write the material-balance equation $x = Ax + d$ and explain, in one or two sentences, why the presence of x on both sides means you cannot solve one sector before the other.
2. Solve for total output $x = (I - A)^{-1}d$. Compute $(I - A)^{-1}$ with the 2×2 inverse formula.
3. Verify that your x satisfies $x - Ax = d$ coordinate by coordinate.
4. The $(1, 1)$ entry of $(I - A)^{-1}$ is the **output multiplier** for agriculture. Read it off, and explain why a one-unit rise in final demand for agriculture raises total agricultural output by *more* than one unit.

Determinants, Invertibility, and Linear Systems

Determinant — Key Facts

The determinant $\det : \mathbb{M}_{n\times n} \to \mathbb{R}$ satisfies $\det(AB) = \det A \det B$, $\det(A^\top) = \det A$, $\det(cA) = c^n \det A$, and equals the product of the diagonal for triangular matrices. Geometrically $|\det A|$ is the volume-scaling factor of the map, and its sign encodes orientation. The central equivalence:

$$\det A \neq 0 \iff \text{Rank}(A) = n \iff A^{-1} \text{ exists} \iff Ax = b \text{ has a unique solution for every } b.$$

Exercise 7 : Computing and reading determinants

We know this exercise is painful (computing co-factors is boring for all of us), but it's good to have done it a few times before you begin using your computer :)

1. Compute $\det B$ for $B = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}$ by cofactor expansion along the first row.
2. Suppose A is 3×3 with $\det A = 5$. Using only the properties above (no entries needed), compute $\det(2A)$, $\det(A^\top)$, $\det(A^{-1})$, and $\det(A^2)$.
3. For the map $A = \begin{pmatrix} 3 & 1 \\ 0 & 2 \end{pmatrix}$, compute $\det A$, state the factor by which it scales area, and say whether it preserves or reverses orientation.
4. Prove that if two columns of a square matrix are equal, then its determinant is 0.

Exercise 8 : Linear Systems, Cramer's Rule, and IS–LM Equilibria

A stylised macroeconomy is summarised by two market-clearing conditions in the two unknowns: output Y and the interest rate i. The goods market (IS, representing the investment-saving tradeoff) and the money market (LM, representing the liquidity preference–money supply tradeoff) yield the following system:

$$\text{IS:} \quad Y + \beta i = \alpha, \qquad\qquad \text{LM:} \quad kY - hi = \frac{M}{P}$$

where α (which rises with government spending G), $\beta, k, h > 0$, and $M/P > 0$ are structural parameters.

Part I: The Linear System

1. Write the system in matrix form $Az = b$ with $z = (Y, i)^\top$, b the outputs of the goods and money market and A the linear map that relates the goods z to the outputs.
2. Compute $\det A$. With the given sign restrictions ($\beta, k, h > 0$), show that $\det A$ can never be zero. Explain geometrically why this guarantees a unique equilibrium, and what $\det A = 0$ would have meant.
3. Solve for the equilibrium output Y^* and interest rate i^* using **Cramer's Rule**.
4. **Comparative statics.** An expansionary fiscal policy raises α. Compute the multiplier vector $\frac{\partial z^*}{\partial \alpha}$. Interpret the signs.

Part II: The Non-Linear Generalization

Now consider the more general, non-linear IS-LM model:

$$Y - C(Y) - I(i) - G = 0$$

$$L(Y, i) - \frac{M}{P} = 0$$

Assume standard economic conditions: the marginal propensity to consume $0 < C_Y < 1$, investment is strictly decreasing in the interest rate $I_i < 0$, money demand increases with output $L_Y > 0$, and money demand decreases with the interest rate $L_i < 0$.

5. Totally differentiate the system with respect to the endogenous variables (Y, i) and the exogenous variable G. Set it up in matrix form: $J\begin{pmatrix} dY \\ di \end{pmatrix} = \begin{pmatrix} dG \\ 0 \end{pmatrix}$, where J is the Jacobian matrix.
6. Sign the determinant of the Jacobian ($|J|$). Does a unique local equilibrium exist?
7. **Solving the non-linear system.** You can now *solve* the differentiated system exactly as you solved $Az = b$ in Part I.

   a. Apply **Cramer's Rule** to $J\begin{pmatrix} dY \\ di \end{pmatrix} = \begin{pmatrix} dG \\ 0 \end{pmatrix}$ to obtain $\frac{dY}{dG}$ and $\frac{di}{dG}$ in terms of C_Y, I_i, L_Y, L_i.

   b. Sign both derivatives. Compare them to the signs you found in Part I.

   c. Show that
   $$\frac{dY}{dG} < \frac{1}{1 - C_Y},$$
   and identify the term in $|J|$ responsible for the gap. What happens to the gap if $I_i = 0$ (investment does not respond to the interest rate)?

   d. Now let the money supply move instead: replace the right-hand side by $\begin{pmatrix} 0 \\ d(M/P) \end{pmatrix}$ and compute $\frac{dY}{d(M/P)}$ and $\frac{di}{d(M/P)}$. Sign them, and say in one sentence how monetary and fiscal expansion differ in their effect on the interest rate.

   e. Take the first row of J and divide it by $(1 - C_Y)$. Compare the result to the matrix A from part 1. What are β, k, and h in terms of the derivatives C_Y, I_i, L_Y, L_i? Conclude that Part I was the special case of Part II all along.

Eigenvalues, Diagonalisation, and Dynamic Stability

Eigenvalues, Diagonalisation

A nonzero v with $Av = \lambda v$ is an **eigenvector** with **eigenvalue** λ; the λ are the roots of $\det(A - \lambda I) = 0$. If A has n independent eigenvectors it is **diagonalisable**, $A = PDP^{-1}$ with $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$, and then $A^k = PD^kP^{-1}$. Also $\det A = \prod_i \lambda_i$ and $\mathrm{tr}\, A = \sum_i \lambda_i$.

Stability of a Linear System

For $x_{t+1} = Ax_t$ (**discrete time**), the origin is stable iff every eigenvalue satisfies $|\lambda_i| < 1$. For $\dot{x} = Ax$ (**continuous time**), it is stable iff every eigenvalue satisfies $\mathrm{Re}(\lambda_i) < 0$. Mixed cases (some roots stable, some not) give a **saddle**.

Exercise 9 : Diagonalisation and matrix powers

Let $A = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$.

1. Find the eigenvalues of A, then an eigenvector for each.
2. Write $A = PDP^{-1}$ explicitly (give P, D, and P^{-1}), and verify $P^{-1}AP = D$.
3. Use the eigenvalues to read off $\det A$ and $\mathrm{tr}\, A$, and write the general formula for A^k. Explain the computational payoff of diagonalisation for large k.
4. Explain in one or two sentences why any 2×2 matrix with two **distinct** eigenvalues is automatically diagonalisable.

Exercise 10 : A saddle-path economy and the Blanchard–Kahn logic

This exercise is entirely self-contained: you do not need any macroeconomics to do it.. We build up the economic story in small steps before asking you to touch a single matrix.

Setting the scene. Imagine an economy described at each date $t = 0, 1, 2, \dots$ by two numbers stacked in a vector $x_t = (x_{1,t}, x_{2,t})^\top$:

- $x_{1,t}$ is the **capital stock**. Capital is a physical quantity of machines and buildings: it is whatever was inherited from yesterday's investment decisions, and it cannot jump. In particular, $x_{1,0} = k_0$ is simply *given* to you as data — a number fixed by history, not something anyone chooses at date 0. Economists call this a **predetermined** variable.
- $x_{2,t}$ is some **price** (think of it as an asset price, or an exchange rate). Unlike capital, a price can move instantly: nothing physical stops it from taking any value at date 0. Its initial value $x_{2,0}$ is a **free choice** — economists call it a **jump variable**.

The economy evolves according to the linear law of motion $x_{t+1} = Ax_t$ with

$$A = \begin{pmatrix} 1.25 & 0.75 \\ 0.75 & 1.25 \end{pmatrix}.$$

Once $x_0 = (x_{1,0}, x_{2,0})$ is fixed, every future x_t is completely determined by repeatedly applying A. The economic question is: **which value of the free jump variable** $x_{2,0}$ should we expect the market to pick? We will see that linear algebra answers this question, and that (perhaps surprisingly) *almost every* choice of $x_{2,0}$ leads to nonsense.

To simplify the quantities, both entries are normalized as deviations from a long-run steady state, so $x_{i,t} = 0$ means "at its steady-state value" and negative entries are perfectly sensible.

1. Before touching A, consider the trivial one-dimensional example $y_{t+1} = 1.5\, y_t$ with $y_0 = 1$. Compute y_1, y_2, y_3, y_{10}. What happens to y_t as $t \to \infty$? Now do the same for $z_{t+1} = 0.5\, z_t$ with $z_0 = 1$. Contrast the two behaviours in one sentence.

2. **Diagonalise** A. Find the eigenvalues λ_1, λ_2 of A and an eigenvector for each. You should get one eigenvalue bigger than 1 in absolute value and one smaller.

3. **Classify the two directions.** In discrete time, a component of the dynamics that moves along an eigenvector with $|\lambda| > 1$ grows without bound; a component along an eigenvector with $|\lambda| < 1$ shrinks to zero. Using this, label which of your two eigenvectors spans the **stable arm** (deviations die out) and which spans the **unstable arm** (deviations explode). Is the origin of this system stable, unstable, or a saddle?

4. **Write down the general path.** Any initial condition x_0 can be written as a combination of the two eigenvectors, $x_0 = c_1 v_1 + c_2 v_2$. Show that this implies

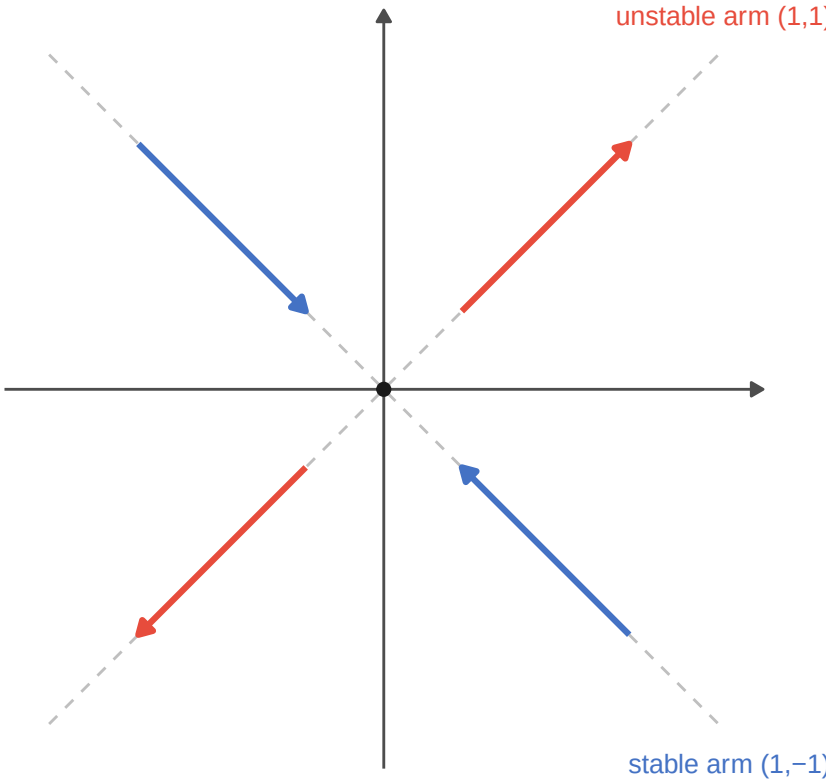
$$x_t = c_1\, \lambda_1^t\, v_1 + c_2\, \lambda_2^t\, v_2.$$

   (Hint: apply A to x_0 once, using $Av_i = \lambda_i v_i$, and spot the pattern; or simply note $x_t = A^t x_0$ and diagonalise as in Exercise 9.)

5. **Why almost every starting point is economically absurd.** Suppose $c_1 \neq 0$ (the coefficient on the *unstable* eigenvector). What happens to $\|x_t\|$ as $t \to \infty$? Explain briefly why no sensible equilibrium can look like this. Conclude that a bounded, economically sensible path requires $c_1 = 0$, i.e. x_0 must lie **exactly on the stable arm**.

6. **Solving for the jump variable.** The capital stock $x_{1,0} = k_0$ is fixed by history — you do not get to choose it. But being on the stable arm pins down a *specific* relationship between $x_{1,0}$ and $x_{2,0}$. Use the stable eigenvector you found in Question 3 to solve for the unique $x_{2,0}$ (as a function of k_0) that places x_0 on the stable arm. This is the value the jump variable "jumps to" at date 0.

7. **Putting the pieces together.** In one or two sentences, explain why the argument above delivers a **unique** economically sensible equilibrium: one predetermined variable, one unstable eigenvalue, and a jump variable that has exactly one degree of freedom to spend. (This equality between "number of jump variables" and "number of unstable roots" is known as the **Blanchard–Kahn condition** for an economy to follow a stable course)



Exercise 11 : A Markov chain and the long-run unemployment rate

Workers move each period between **Employed** (E) and **Unemployed** (U). An employed worker stays employed with probability 0.9 and loses the job with probability 0.1; an unemployed worker finds a job with probability 0.6 and stays unemployed with probability 0.4.

Setting up the bookkeeping. At date t, let π_t^E be the *share* of the whole workforce that is currently employed, and π_t^U the share currently unemployed, so $\pi_t^E + \pi_t^U = 1$. Stack them into $\pi_t = (\pi_t^E, \pi_t^U)^\top$. Each period, a slice of the employed pool becomes unemployed and a slice of the unemployed pool becomes employed, so tomorrow's shares are a weighted blend of today's:

$$\pi_{t+1}^E = 0.9\,\pi_t^E + 0.6\,\pi_t^U, \qquad \pi_{t+1}^U = 0.1\,\pi_t^E + 0.4\,\pi_t^U.$$

In matrix form this is $\pi_{t+1} = P\pi_t$ with

$$P = \begin{pmatrix} 0.9 & 0.6 \\ 0.1 & 0.4 \end{pmatrix}.$$

We want to know: if we let this process run forever, does the unemployment rate settle down to some fixed number, and if so, which number — regardless of where we started?

1. **Get your hands dirty first.** Suppose the whole workforce starts out employed: $\pi_0 = (1, 0)$. Compute $\pi_1 = P\pi_0$, then $\pi_2 = P\pi_1$, then $\pi_3 = P\pi_2$ by direct matrix multiplication (no eigenvalues yet). What do you notice about $\pi_1^U, \pi_2^U, \pi_3^U$ — are they moving towards some number? *(You don't need to guess the exact limit yet — just observe the trend.)*

2. **Read the matrix.** Confirm that each column of P sums to 1 (this property is called **column-stochastic**), and explain in words what column 1 and column 2 each represent in terms of where a worker goes next period.

3. **Why $\lambda = 1$ is special here.** A **stationary distribution** π^* is a distribution that reproduces itself: $P\pi^* = \pi^*$. Recognise that this is exactly the eigenvector equation $P\pi^* = \lambda\pi^*$ with

$\lambda = 1$. So finding the long-run outcome (if it exists) reduces to finding an eigenvector of P with eigenvalue 1.

4. $\lambda = 1$ is guaranteed to exist. Before computing anything, show that $\lambda = 1$ is *always* an eigenvalue of a column-stochastic matrix like P. *(Hint: compute* $\mathbf{1}^\top P$ where $\mathbf{1} = (1, 1)^\top$, using the column-sums-to-1 property from Question 2. What does this tell you about $\mathbf{1}$ as a left eigenvector of P? You may use, without proof, the fact that P and $P^\top$ share the same eigenvalues.)

5. **Solve for the stationary distribution.** Solve $(P - I)\pi = 0$ for π, then normalise so that $\pi^E + \pi^U = 1$ (a distribution's shares must sum to 1). Call the result π^*. Compare $\pi^{*,U}$ to the trend you spotted numerically in Question 1.

6. **The second eigenvalue.** Find the other eigenvalue λ_2 of P (using $\operatorname{tr} P$ and $\det P$, as in Exercise 9). Confirm $|\lambda_2| < 1$.

7. **Why any starting point converges — step by step.** Any initial distribution π_0 can be written as a combination of the two eigenvectors, $\pi_0 = a\,\pi^* + b\,v_2$, where v_2 is an eigenvector for λ_2.

   a. Using $P\pi^* = \pi^*$ and $Pv_2 = \lambda_2 v_2$, write down $\pi_t = P^t\pi_0$ in terms of a, b, π^*, v_2, λ_2, and t.
   b. What happens to the v_2 term as $t \to \infty$, given $|\lambda_2| < 1$? *(Compare to the warm-up reasoning from Exercise 10, Question 1.)*
   c. It remains to pin down a. Show that the entries of v_2 must sum to 0. *(Hint: apply* $\mathbf{1}^\top$ to both sides of $Pv_2 = \lambda_2 v_2$, use $\mathbf{1}^\top P = \mathbf{1}^\top$ from Question 4, and solve for $\mathbf{1}^\top v_2$ given that $\lambda_2 \neq 1$.)
   d. Every π_t is a distribution, so its two entries must sum to 1 for **every** t, including $t = 0$. Using part (c) and $\pi^{*,E} + \pi^{*,U} = 1$, show this forces $a = 1$.
   e. Conclude: what is $\lim_{t\to\infty} \pi_t$? Does it depend on the starting point π_0?

8. **Speed of convergence.** Explain, in one sentence, why a *smaller* $|\lambda_2|$ means the economy forgets its starting unemployment rate *faster*. Roughly how many periods does it take for the gap $\|\pi_t - \pi^*\|$ to shrink by a factor of 10, given $\lambda_2 = 0.3$? *(A rough order of magnitude is fine — you may use* $\log_{10}$.)

OLS as a Projection

OLS and the Hat Matrix

For $Y = X\beta + \varepsilon$ with $X \in \mathbb{M}_{N\times K}$ of full column rank, the OLS estimator minimises $\|Y - Xb\|^2$ and equals

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y, \qquad \hat{Y} = P_X Y, \qquad \hat{\varepsilon} = M_X Y,$$

with $P_X = X(X^\top X)^{-1}X^\top$ and $M_X = I - P_X$. Residuals are orthogonal to every regressor: $X^\top \hat{\varepsilon} = 0$.

Frisch–Waugh–Lovell

In the partitioned model $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, with $M_1 = I - P_{X_1}$,

$$\hat{\beta}_2 = (X_2^\top M_1 X_2)^{-1} X_2^\top M_1 Y.$$

Equivalently: partial X_1 out of both Y and X_2, then regress the residuals — the coefficient on X_2 is identical to its value in the full regression.

Exercise 12 : Deriving OLS and the orthogonal residual

Motivation

Suppose you've collected data on three students. For each one you know how many **hours they studied** and what **exam score** they got:

student	hours studied	exam score
A	1	1
B	2	3
C	3	4

You suspect score goes up roughly linearly with hours, so you want to draw the "best" straight line score $\approx \beta_1 + \beta_2 \cdot$ (hours) through these three points. Here β_1 is the intercept (predicted score for zero hours) and β_2 is the slope (extra score per extra hour of study). The catch: no straight line passes through all three points exactly. Whatever line you pick, it misses each point by some amount: call those three misses the **prediction errors** ε. **OLS** ("ordinary least squares") is about following a simple rule: *pick the line that makes the total of the squared misses as small as possible.*
In this exercise, we will make clear that the best line is actually an orthogonal projection.

Collect the scores in a vector $Y = (1, 3, 4)$ and stack the inputs in a matrix X whose first column is all 1's (that column represents the fact that individuals would have had some test score, even in the absence of study time) and whose second column is the hours $(1, 2, 3)$:

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \qquad Y = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix}.$$

1. **Find the best coefficients.** We want the b minimising the total squared error $L(b) = \|Y - Xb\|^2$.

a. Expand $L(b) = (Y - Xb)^\top (Y - Xb)$ into three terms in b.

b. To minimise, differentiate with respect to b, set the gradient to zero, and solve to obtain the **normal equations** $X^\top X\, b = X^\top Y$ and $\hat{\beta} = (X^\top X)^{-1} X^\top Y$. *Hint: you can either do the derivatives directly with the matrices if you are comfortable, or expand into 2 equations and derivate each equation separately.*

c. Solving needs $X^\top X$ to be invertible. Look at the two columns of X and explain why they're independent. Explain why linear independence of columns of X is sufficient to have invertibility of $X^\top X$. *Hint: think about what linear independence means for the equations* $X^\top Xb = 0$ *and* $Xb = 0$.

d. Suppose a careless research assistant adds a third column recording study time **in minutes**, so X becomes 3×3 with columns $(1, 1, 1)$, $(1, 2, 3)$ and $(60, 120, 180)$. Find a nonzero b with $Xb = 0$. What does exercise 5, Question 4 then say about how many b solve the normal equations?

e. Even so, the *predictions* are fine. Explain why $\mathrm{col}(X)$ is the same set with or without the minutes column, and conclude that $\hat{Y}$ is unchanged. Which of $\hat{\beta}$ and $\hat{Y}$ is the fragile one, and why?

f. A subtler failure, and the one you will actually meet: to a regression with an intercept, add a dummy D_M equal to 1 for male students and 0 otherwise, *and* a dummy D_F equal to 1 for female students and 0 otherwise. Show the columns of X are dependent. Conclude.

2. Can you explain why the orthogonal projection solves the minimization problem?

3. **Crunch the numbers.**

   a. Compute $X^\top X$ and $X^\top Y$.
   b. Invert $X^\top X$ and get $\hat{\beta} = (\beta_1, \beta_2)$. State, in words, what your intercept and slope mean in this exercise.
   c. Compute the fitted scores $\hat{Y} = X\hat{\beta}$ and the residuals $\hat{\varepsilon} = Y - \hat{Y}$.
   d. Check that $X^\top \hat{\varepsilon} = 0$ (both entries). The second entry vanishing is exactly "errors uncorrelated with hours studied."