Game Theory

Lecture 4: Joint Policies, Expected Return, and Minimax

László Gulyás

Eötvös University, Department of Artificial Intelligence

2025/26/1

Course textbooks

- Bonanno, G. (2024). *Game Theory (3rd ed.)*. University of California, Davis. Received from: GT Book
- Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books. Received from: Axelrod Article
- Nisan, N., Roughgarden, T., Tardos, É., & Vazirani, V. V. (2007). *Algorithmic Game Theory*. Cambridge University Press. Received from: AGT Book
- Myerson, R. B. (1991). Game Theory: Analysis of Conflict. Harvard University Press. Received from: GT Book 2
- F. Christianos et al., *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches*, 2023. Received from: MARL Book.pdf
- Shoham, Y., & Leyton-Brown, K. (2008). Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations. Cambridge University Press Received from: MARL Book.pdf
- `nashpy` documentation (readthedocs) Link: NashPy Docs

Previously on Lecture 3

- Efficiency lenses: Pareto frontier, Social Welfare, Price of Anarchy
- Beyond NE: **CE/CCE** (obedience via signals), **QRE** (noisy best response)
- **Learning dynamics**: FP, BRD, Replicator - how play moves over time
- Takeaway: Many ways to improve/interpret outcomes **given** a stage game

Lecture Overview

- Previous lecture gave selection/improvement tools (CE/QRE) & dynamic paths.
- This lecture zooms into **strictly competitive** settings:
  - Which joint policies are *guaranteed-safe*?
  - What is the **value** of play per period/discounted?
  - How to compute it fast and verify it's optimal?

Outline

Formal setup: feasible payoffs

- Finite normal-form game with payoff functions $u_i(a)$ for $a \in A = \prod_i A_i$.
- Let X be the set of **joint distributions** on A (mixed/correlated play).
- **Feasible payoff set**:

$$U = \{(\mathbb{E}_x[u_1(a)], \ldots, \mathbb{E}_x[u_n(a)]) : x \in X\}.$$

- U is **compact**; if mixed/correlated are allowed, U is the convex hull of the pure payoff vectors.

Nash social welfare (convex form)

Given baselines $\bar{u}_i$ with feasibility $u_i > \bar{u}_i$, solve

$$\max_{x \in X} \sum_{i=1}^{n} \log \Big(\sum_{a} x(a) u_i(a) - \bar{u}_i \Big)$$

- Concave in x (sum of concave $\log$ of affine functions).
- Yields the **Nash bargaining** point under standard axioms.

Aumann's idea: recommendations you want to obey

- A **mediator** draws a joint action $a = (i, j)$ from a public distribution x on $A_1 \times A_2$, and sends **private** recommendation i to Row and j to Column.
- Each player updates by Bayes:

$$\Pr(j \mid i) = \frac{x_{ij}}{\sum_{k \in A_2} x_{ik}}.$$

- A **Correlated Equilibrium (CE)** is any x such that *obeying the recommendation* is a best response given the **posterior** they infer from their own signal.

CE vs public correlation

- **Private** recommendations are sufficient for CE.
- With a **public** signal only (no private advice), you generally get a **public correlated equilibrium**; this can be weaker (players can infer others' advice and may want to deviate).
- Private messages are key to **obedience** at the individual level.

When CE fails to help

- **Dominance-driven** temptations (PD): obedience to "cooperate" is not credible.
- **Strictly competitive** (zero-sum): value fixed.
- **Miscoordinated posteriors**: your candidate x induces posteriors that make deviation profitable $\to$ not a CE.

CE & learning

- Many **no-regret** learning processes converge to the **CCE** set; with smoothness, their **worst-case welfare** matches PoA bounds.
- Adding **signal devices** (recommendations) can move play from CCE toward **CE** and closer to the **Pareto frontier**.

Why QRE? (Bounded rationality meets equilibrium)

- **Empirics:** Lab play often deviates from Nash but is **payoff-sensitive**.
- **Idea:** Players **don't perfectly best-respond**; they choose better actions more often.
- **QRE (McKelvey–Palfrey):** Replace hard best response with a **smooth**, stochastic choice rule; fix points of these smooth responses are equilibria.

Why dynamics?

- Equilibria say **where** play can end up; dynamics say **how** it might get there.
- Useful for **prediction**, **selection** (which NE), and **algorithmic intuition**.

Fictitious Play (FP): beliefs $\to$ BRs

- Each player best-responds to the **empirical frequency** of opponent's past actions.
- Belief update (row vs column with actions $j \in A_2$):

$$\hat{q}_j^{(t)} = \frac{1}{t} \sum_{s=1}^{t} \mathbf{1}\{a_2^{(s)} = j\}, \quad a_1^{(t+1)} \in \arg\max_i (A\,\hat{q}^{(t)})_i.$$

- **Converges** in 2p zero-sum, potential, and dominance-solvable games.
- **May cycle** in general-sum (e.g., Shapley's game), but time-averages can converge.
- **Intuition:** conservative learning of others' stationary mix.

Quick comparison table

Concept	Info	Randomness	Solve	Welfare
NE	None	independent	supports/LP	baseline
CE	private signals	correlated	LP	often $\uparrow$
QRE	none	stochastic choice	fixed point	behavioral
Learning	history	induced by play	simulation	depends

What is a joint policy? (repeated matrix game, zero-sum)

- **Stage game:** Row payoff matrix $A \in \mathbb{R}^{m \times n}$, Column payoff $-A$.
- **Round $t = 0, 1, 2, \dots$:** players choose pure actions $i_t \in [m]$, $j_t \in [n]$.
- **Instant payoff:** $r_t = A_{i_t j_t}$ to Row (and $-r_t$ to Column).

A **joint policy** $\Pi = (\pi_1, \pi_2)$ specifies, for each round t, a (possibly history-dependent) mixed action for each player:

$$\pi_1(\cdot \mid h_t) \in \Delta^m, \quad \pi_2(\cdot \mid h_t) \in \Delta^n, \qquad h_t = (i_0, j_0, \dots, i_{t-1}, j_{t-1}).$$

- **Stationary (memoryless) policy:** $\pi_1(\cdot \mid h_t) \equiv p \in \Delta^m$, $\pi_2(\cdot \mid h_t) \equiv q \in \Delta^n$ for all t.
- Unless stated otherwise, draws are **independent across players and across time** under a stationary policy.
  Notation. *e_i denotes the i-th standard basis vector.*
  $\mathbf{1}$ denotes an all-ones vector.
  $\Delta^k = \{x \in \mathbb{R}^k : x \geq 0,\ \mathbf{1}^\top x = 1\}$ is the probability simplex.

Discounted return under stationary independent mixing

Fix $\gamma \in (0,1)$. Under stationary **independent** mixing (p,q) each round,

$$\mathbb{E}[r_t] = \sum_{i,j} p_i \, A_{ij} \, q_j \;=\; p^\top A q \quad \text{for all } t,$$

and, by linearity of expectation with i.i.d. draws,

$$J_\gamma(p,q) \;:=\; \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right] \;=\; \sum_{t=0}^{\infty} \gamma^t \, \mathbb{E}[r_t] \;=\; \frac{p^\top A q}{1-\gamma}.$$

Remarks.

- The independence across time is sufficient; no ergodic or martingale machinery is required here.
- If you add a constant c to all entries of A, then J_γ shifts by $c/(1-\gamma)$; optimal mixes are unchanged.

Long-run average reward (Cesàro average)

Under stationary independent mixing (p, q),

$$\bar{J}(p,q) := \lim_{T\to\infty} \frac{1}{T}\,\mathbb{E}\left[\sum_{t=0}^{T-1} r_t\right] = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[r_t] = p^\top A q.$$

Interpretation. The one-shot value $p^\top Aq$ is the **per-period** expected return of the stationary joint policy; discounting just rescales it by $(1-\gamma)^{-1}$.

> ***Caution.*** *If players correlate across time (e.g., contingent punishments), per-period* ***expectation*** *under stationary (p, q) is still $p^\top Aq$, but non-stationary history-dependent strategies can implement different paths. In* ***matrix zero-sum*** *games, however, these paths cannot raise the secure average payoff above the minimax value (see Minimax section).*

Security levels: maximin vs minimax

Row's **maximin** (security guarantee):

$$v^- \;:=\; \max_{p\in\Delta^m}\;\min_{q\in\Delta^n}\; p^\top Aq.$$

Column's **minimax** (Row's worst-case given Column's choice):

$$v^+ \;:=\; \min_{q\in\Delta^n}\;\max_{p\in\Delta^m}\; p^\top Aq.$$

Weak minimax inequality. For any bilinear form over compact convex sets,

$$v^- \;\le\; v^+.$$

Proof sketch: For any p, q, $\min_{q'} p^\top Aq' \le p^\top Aq \le \max_{p'} p'^\top Aq$.
Take $\max_p$ on the left inequality and $\min_q$ on the right inequality.

Interpretation.

- v^-: Row can guarantee at least v^- regardless of Column ($\Rightarrow$ **security level**).
- v^+: Column can hold Row down to at most v^+ regardless of Row.

In finite zero-sum matrix games, the **Minimax Theorem** (next section) states $v^- = v^+ = v$: the **value** of the game.

Side notes & edge cases

- **Finite horizon T:** Under stationary independent (p, q),

$$\mathbb{E}\left[\sum_{t=0}^{T-1} r_t\right] = T\, p^\top A q.$$

- **Time-varying stationary but independent** (piecewise constant p_t, q_t): the per-period mean is $\frac{1}{T}\sum_{t=0}^{T-1} p_t^\top A q_t$.
- **History-dependent** strategies (threats/punishments): In general-sum, these matter (Folk theorems). In **zero-sum matrix** games, they cannot beat the minimax value v in expected average payoff.
- **Nonzero-sum:** Expected returns are bilinear in (p, q) per player; many results above carry but **security equality** (minimax) does not.

Micro-check (2×2 stationary)

Let

$$A = \begin{pmatrix} 2 & -1 \\ -3 & 4 \end{pmatrix}, \quad p = (p, 1-p), \; q = (q, 1-q).$$

Then

$$p^\top A q = p\,[2q - 1(1-q)] + (1-p)\,[-3q + 4(1-q)] = (5p-3)q + (-4p+4).$$

- For fixed q, Row's best response is $\arg\max_p$ of a linear function in p.
- For fixed p, Column's best response is $\arg\min_q$ of the same bilinear expression.

  Indifference equalization yields $p^* = 0.7,\; q^* = 0.5$ (derived later), hence $\bar{J} = p^{*\top} A q^* = 0.05$ and $J_\gamma = 0.05/(1-\gamma)$.

Questions

1. Show $J_\gamma(p,q) = \frac{p^\top A q}{1-\gamma}$ using only independence and linearity of expectation.
2. Prove $v^- \le v^+$ for any compact convex P, Q and continuous bilinear payoff.
3. (Concept) Give a general-sum 2×2 where non-stationary correlation across time changes the *distribution of outcomes* relative to stationary play, even though the stage expectation formula holds under stationarity.

Outline

Statement & Intuition (finite matrix games)

Theorem (von Neumann). For any finite two-player zero-sum matrix game with Row payoff $A \in \mathbb{R}^{m\times n}$,

$$\max_{p\in\Delta^m}\ \min_{q\in\Delta^n} p^\top Aq \;=\; \min_{q\in\Delta^n}\ \max_{p\in\Delta^m} p^\top Aq \;=\; v.$$

There exist optimal mixes $p^* \in \Delta^m,\ q^* \in \Delta^n$ such that

$$p^{*\top} Aq \geq v \ \ \forall q, \qquad p^\top Aq^* \leq v \ \ \forall p.$$

Intuition. Row can **guarantee** at least v (security), Column can **hold** Row down to at most v; equality pins down the **value** and optimal mixed strategies.

Saddle-Point View (equivalent characterization)

$$(p^*, q^*) \text{ is minimax } \iff p^{*\top} A q \ge p^{*\top} A q^* \ge p^\top A q^* \quad \forall p, q.$$

At a saddle point, neither player can profitably deviate; the common value is $v = p^{*\top} A q^*$.

Weak Minimax Inequality (prelude)

For any bilinear payoff over compact convex sets,

$$\max_p \min_q p^\top A q \;\leq\; \min_q \max_p p^\top A q.$$

Proof sketch: For all p, q, $\min_{q'} p^\top A q' \leq p^\top A q \leq \max_{p'} p'^\top A q$. Take $\max_p$ on the left, $\min_q$ on the right.

LP Formulation (Row / "primal")

(Optionally shift A by a constant so entries are nonnegative; mixes are invariant to affine shifts.)

$$\begin{aligned} \max_{p,v} \quad & v \\ \text{s.t.} \quad & A^\top p \;\geq\; v\,\mathbf{1}, \\ & \mathbf{1}^\top p = 1, \quad p \geq 0. \end{aligned}$$

Meaning. Choose p so that **every column** yields at least v.

LP Dual (Column)

$$\begin{aligned} \min_{q,v} \quad & v \\ \text{s.t.} \quad & Aq \ \le \ v\,\mathbf{1}, \\ & \mathbf{1}^\top q = 1, \quad q \ge 0. \end{aligned}$$

Meaning. Choose q so that **every row** yields at most v.

Consequence. LP **strong duality** ⇒ optimal values match ⇒ minimax equality.

Complementary Slackness (support equalization)

At an optimal pair (p^*, q^*, v):

- If $p_i^* > 0$, then the i-th row payoff equals the value: $(Aq^*)_i = v$.
- If $q_j^* > 0$, then the j-th column payoff equals the value: $(A^\top p^*)_j = v$.

Takeaway. Supported pure actions are **equalized at value** v; excluded actions satisfy the corresponding inequality strictly.

Computing by Indifference (support method)

Given supports $I \subseteq [m]$, $J \subseteq [n]$:

1. **Row equalization on I:** $(Aq)_i = v$ for all $i \in I$.
2. **Column equalization on J:** $(A^\top p)_j = v$ for all $j \in J$.
3. **Normalization:** $\sum_{i \in I} p_i = 1,\ p_i \geq 0;\ \sum_{j \in J} q_j = 1,\ q_j \geq 0.$
4. **Verify inequalities:** $(Aq)_i \leq v$ for $i \notin I$; $(A^\top p)_j \geq v$ for $j \notin J$.

If feasible, (p, q, v) is a solution. Otherwise try different supports.

Example 1 (2×2 by indifference)

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

Let $p = \Pr[U]$, $q = \Pr[L]$.

- **Row indiff (equalize Column's payoffs):**

$$2q + (-3)(1-q) = -1 \cdot q + 4(1-q) \;\Rightarrow\; 3q - 1 = -7q + 4 \;\Rightarrow\; q^* = 0.5.$$

- **Column indiff (equalize Row's payoffs):**

$$2p + (-1)(1-p) = -3p + 4(1-p) \;\Rightarrow\; 3p - 1 = -7p + 4 \;\Rightarrow\; p^* = 0.7.$$

- **Value:**

$$v = p^{*\top} A q^* = [0.7 \quad 0.3] \begin{bmatrix} 2 & -1 \\ -3 & 4 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} = 0.05.$$

Example 2 (3×3 RPS)

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

By symmetry, $p^* = q^* = (1/3, 1/3, 1/3)$, $v = 0$. Check: $Aq^* = 0 \cdot \mathbf{1}$, $A^\top p^* = 0 \cdot \mathbf{1}$.

Example 3 (weighted RPS, full support system)

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

Solve

$$Aq = v\,\mathbf{1}, \quad A^\top p = v\,\mathbf{1}, \quad \mathbf{1}^\top p = \mathbf{1}^\top q = 1.$$

Check $p, q \geq 0$; the solution yields full-support mixes and v.

Example 4 (dominance pruning)

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

Row 3 is dominated by a mixture of Rows 1–2. Remove it, solve 2×2 by indifference; verify Row 3 remains unprofitable at the solution.

Geometry & Invariances (quick facts)

- **Hyperplanes.** Equal-payoff sets $(Aq)_i = (Aq)_{i'}$ and $(A^\top p)_j = (A^\top p)_{j'}$ are linear (hyperplanes in the simplexes).
- **Polytopes.** Best-response regions are intersections of half-spaces $\Rightarrow$ polytopes; equilibria are at polytope intersections.
- **Affine transforms.** $A \mapsto \alpha A + c\mathbf{1}\mathbf{1}^\top$: mixes unchanged; value scales by α and shifts by c.

Computational Tools

```
import numpy as np, nashpy as nash

A = np.array([[2,-1],[-3,4]])
G = nash.Game(A)  # zero-sum shorthand
list(G.vertex_enumeration())  # returns (p*, q*)
```

Outline

LP template for Row (code-level concept)

- **Variables:** probabilities on rows $p \in \mathbb{R}^m$ and value $v \in \mathbb{R}$.
- **Constraints:** $A^\top p \geq v\mathbf{1}$, $\mathbf{1}^\top p = 1$, $p \geq 0$.
- **Objective:** maximize v.

The **dual** is Column's problem automatically (minimize v with $Aq \leq v\mathbf{1}$, $\mathbf{1}^\top q = 1$, $q \geq 0$).

Use any LP solver that handles linear inequalities (cvxopt, PuLP, SciPy `linprog`, CVXPy, ...).

Complementary Slackness (quick check on Example 1)

At $(p^*, q^*, v) = (0.7,\ 0.5,\ 0.05)$ for

$$A = \begin{pmatrix} 2 & -1 \\ -3 & 4 \end{pmatrix},$$

- **Row** supported actions U, D both achieve value exactly 0.05 against q^*: $(Aq^*)_U = (Aq^*)_D = v$.
- **Column** supported actions L, R both yield value exactly 0.05 against p^*: $(A^\top p^*)_L = (A^\top p^*)_R = v$.
- In a 2×2 there are no excluded pure actions; feasibility is immediate.

Takeaway: supported actions are *equalized at* v; any excluded actions (in larger games) must satisfy strict inequality.

Invariances: value shifts & scalings

- **Shift:** $A \mapsto A + c\,\mathbf{1}\mathbf{1}^\top$
  mixes p^*, q^* unchanged; value shifts by c.
- **Scale:** $A \mapsto \alpha A$ with $\alpha > 0$
  mixes unchanged; value scales by α.
  Use these to simplify arithmetic (e.g., make entries nonnegative for LP stability).

ε-security (numerical robustness)

For any $\varepsilon > 0$, there exists p_ε s.t.

$$\min_q p_\varepsilon^\top A q \;\geq\; v - \varepsilon,$$

and q_ε s.t.

$$\max_p p^\top A q_\varepsilon \;\leq\; v + \varepsilon.$$

Practice: When you compute $(\hat{p}, \hat{q})$ numerically, report the **deviation incentives**

$$\varepsilon_{\text{row}} = \max_i (A\hat{q})_i - \hat{p}^\top A\hat{q}, \qquad \varepsilon_{\text{col}} = \hat{p}^\top A\hat{q} - \min_j (\hat{p}^\top A)_j,$$

and use $\max(\varepsilon_{\text{row}}, \varepsilon_{\text{col}})$ as a conservative error bound.

Relation to Nash Equilibrium (zero-sum)

In two-player zero-sum games, **minimax strategies** (p^*, q^*) are **exactly** Nash equilibria, and the **equilibrium payoff** equals the **value** v. Conversely, any NE mixed profile is minimax.

Sion's Minimax Theorem (statement)

Let $X \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^n$ be nonempty compact convex sets.
If $f : X \times Y \to \mathbb{R}$ is **quasi-convex and lower semicontinuous** in x for each y, and **quasi-concave and upper semicontinuous** in y for each x, then

$$\min_{y \in Y} \max_{x \in X} f(x, y) \;=\; \max_{x \in X} \min_{y \in Y} f(x, y).$$

This generalizes the matrix-game minimax equality.

Proof Sketch A: LP strong duality

1. Write Row's problem as an LP; Column's is the **dual**.
2. Feasibility & boundedness $\Rightarrow$ **strong duality**: optimal values match.
3. Optimal primal/dual solutions yield (p^*, q^*, v).
4. **Complementary slackness** explains equalization of supported actions.

Proof Sketch B: Fixed-point route (intuition)

1. Mixed strategy spaces are simplexes (compact, convex).
2. Best-response correspondences are nonempty, convex-valued, upper-hemicontinuous (Berge).
3. Existence of NE (Kakutani) in zero-sum $\Rightarrow$ value-attaining equilibrium; Row's secured payoff equals Column's held-down payoff $\Rightarrow$ minimax equality.

Outline

Bridge from last section (what changes in repeated play?)

- In **static** zero-sum matrix games you computed (p^*, q^*, v) by **minimax**.
- In **repeated** play (finite/infinite horizon), if each round is the *same* stage game and players mix **independently** each round, then:
  - Per-period payoff is $p^\top Aq$.
  - Discounted return $J_\gamma(p, q) = \frac{p^\top Aq}{1-\gamma}$.
  - Stationary minimax p^*, q^* **secure** value v each round $\rightarrow$ the repeated game's value is v (per period).

Joint policies in repeated matrix games

- Stage game payoffs: Row $A \in \mathbb{R}^{m \times n}$, Column $-A$ (zero-sum).
- **Joint policy** (possibly history-dependent): mapping from histories $\mathcal{H}_t$ to mixed actions.
- **Stationary independent** policy: fixed $p \in \Delta^m$, $q \in \Delta^n$ each round.

Discounted return (stationary, independent)

$$J_\gamma(p, q) \;=\; \mathbb{E}\Big[\sum_{t=0}^{\infty} \gamma^t \, a_{i_t j_t}\Big] \;=\; \frac{p^\top A q}{1-\gamma}.$$

Average reward (Cesàro)

$$\lim_{T \to \infty} \frac{1}{T} \, \mathbb{E}\Big[\sum_{t=0}^{T-1} a_{i_t j_t}\Big] \;=\; p^\top A q.$$

Security levels in repeated play (why history doesn't help in zero-sum)

Let

$$v^- = \max_p \min_q p^\top A q, \quad v^+ = \min_q \max_p p^\top A q, \quad (v^- \le v^+).$$

By the **minimax theorem** $v^- = v^+ = v$.

- Against any opponent policy (even history-dependent), Row can play p^* i.i.d. each round and **guarantee** at least v per period.
- Symmetrically Column can hold Row to **at most** v.
- Thus the **repeated zero-sum game** (with the same stage game) has **per-period value** v; using history cannot beat v in expectation.

Example 5: uniform security in a centered symmetric game

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} - \frac{2}{3}\mathbf{1}\mathbf{1}^\top.$$

- Each row/column sum is $0 \to$ **uniform** $p = q = (1/3, 1/3, 1/3)$.
- Value $v = 0$. Verify $Aq = \mathbf{0}$ and $A^\top p = \mathbf{0}$.

Example 6: 3×3 with support size 2 (support equalization)

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

Try supports $I = \{U, D\}$, $J = \{L, R\}$. Solve the induced 2×2 by equalizing supported payoffs at value v.
Check the middle actions are **not** profitable; if violated, adjust supports.

Example 7: dominance pruning first

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

Row 3 is dominated by a mixture of Rows 1–2. Remove, solve the 2×2, then **reinsert** Row 3 to confirm it remains unprofitable at (p^*, q^*, v).

Geometry refresher (why intersections matter)

- **Equal-payoff hyperplanes:** $e_i^\top Aq = e_{i'}^\top Aq$ are linear constraints in q.
- **Best response regions:** intersections of halfspaces (polyhedral).
- **Equilibria:** at intersections where both players are indifferent on **their supported actions** and inequalities hold for excluded ones.

Algorithms in practice (what to use when)

- **2×2 / some 3×3:** support guessing + indifference + inequality checks.
- **Small/medium:** vertex enumeration of BR polytopes (e.g., NashPy for zero-sum).
- **Larger:** LP (sparse) or first-order primal-dual methods.
- **Teaching/demo:** NashPy is quick and reliable for small sizes.

Epsilon-security & solver tolerance (diagnostics)

Given numerical $(\hat{p}, \hat{q})$, define

$$\varepsilon_{\text{row}} = \max_i (A\hat{q})_i - \hat{p}^\top A\hat{q}, \quad \varepsilon_{\text{col}} = \hat{p}^\top A\hat{q} - \min_j (\hat{p}^\top A)_j.$$

Report $\max(\varepsilon_{\text{row}}, \varepsilon_{\text{col}})$ as a conservative suboptimality bound.

Exercise 1. - hand computation (2×2)

For

$$A = \begin{pmatrix} 3 & -1 \\ 0 & 2 \end{pmatrix},$$

1) Compute (p^*, q^*, v) by support equalization.
2) Remove any dominated actions if found and recompute.
3) Verify equalization/inequalities.

Exercise 2. - 3×3 support search

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

Enumerate size-2 supports, keep feasible ones. If none feasible, try full support and solve the linear system $Aq = v\mathbf{1},\ A^\top p = v\mathbf{1},\ \mathbf{1}^\top p = \mathbf{1}^\top q = 1$.

Exercise 3. - robustness under perturbations

Add i.i.d. noise $\xi_{ij} \sim$ Unif$[-0.1, 0.1]$ to a 3×3 with known (p^*, q^*, v).
Recompute $(\tilde{p}, \tilde{q}, \tilde{v})$.
Summarize how supports and v change.

Exercise 4. - Exploitability of naive play

Fix Column at a non-equilibrium q_0.
(a) Compute Row's best response $p^{BR}(q_0)$ and value $v(q_0)$.
(b) Define **exploitability** $E(q_0) = v(q_0) - v$.
(c) Repeat for several q_0 to visualize the **exploitability landscape**.

Exercise 5. - Family with a parameter (piecewise supports)

Let

$$A(\theta) = \begin{pmatrix} 2 & -1 \\ -3 & 1+\theta \end{pmatrix}, \quad \theta \in [-1, 2].$$

Derive $(p^*(\theta), q^*(\theta), v(\theta))$ **piecewise** in θ; identify breakpoints where supports change.

Exercise 6. - Regret minimization minimax

Run a no-regret algorithm (e.g., Hedge) for both players on a 3×3 zero-sum game.
Track average plays $\bar{p}_T, \bar{q}_T$ and payoffs $\bar{v}_T$.
Show $\bar{v}_T \to v$ and exploitability $\to 0$.

Exercise 7. - Correlation doesn't help in zero-sum

Propose any correlated device x for a zero-sum game.
Show Row's CE payoff $\leq v$ and Column's $\geq -v$.
Conclude CE **cannot** beat minimax value in zero-sum.

Exercise 8. - Repeated play with discounting

Prove rigorously that stationary (p^*, q^*) yields discounted return $v/(1-\gamma)$. Argue any history-dependent deviation cannot improve the **per-period** value above v.

Common pitfalls

- Equalizing **all** actions instead of **supported** ones only.
- Forgetting normalization $\sum p_i = 1,\ \sum q_j = 1$.
- Mixing Row/Column inequalities' directions.
- Not re-checking excluded actions after solving.

Summary

- In zero-sum matrix games, **minimax = Nash** and yields value v.
- Repeated play with stationary independent minimax achieves **per-period** v; history dependence cannot beat it in expectation.
- **Computation:** supports + indifference for small games; LP/vertex enumeration otherwise.
- **Diagnostics:** complementary slackness and ε-security quantify solution quality.
- **Design insight:** in zero-sum, **correlation** doesn't raise value; learning/no-regret converges to minimax.