MIT 6.S890 — Topics in Multiagent Learning Tue, Oct 17th 2024

Lecture 11

Sequential irrationality and perfect equilibria

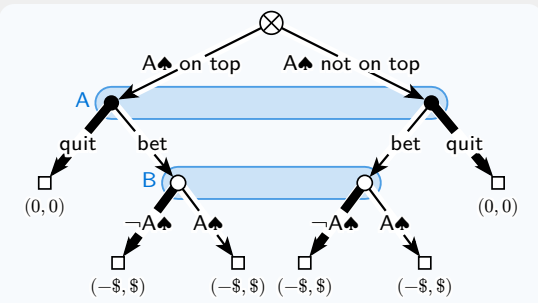
Instructor: Prof. Gabriele Farina (✉ gfarina@mit.edu)*

As we discussed on multiple occasions, Nash equilibrium strategies encode the idea of playing optimally against the strongest possible opponent. Even when the opponent is only close to optimal (for example, in the poker competitions where the opponent were top professional poker players), playing a Nash equilibrium is often the safe choice, as professional players are very quick at exploiting suboptimal strategies, making opponent modeling risky. However, as we reveal today, not all Nash equilibria are equally strong in extensive-form games when playing against players that might make mistakes.

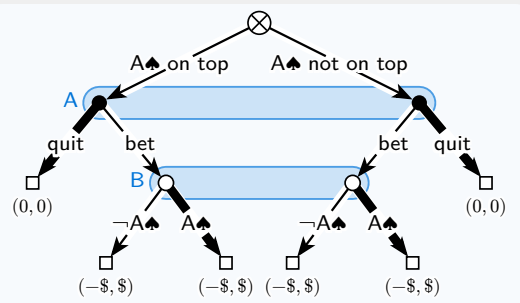
1 Sequential irrationality

Nash equilibrium strategies are only optimized for the strongest possible opponent. Because of that, they are completely indifferent to what happens in parts of the game tree that are reached only if a player makes a mistake.

Example 1.1. To make the discussion more concrete, consider the *Guess-the-Ace* game, introduced by Miltersen, P. B., & Sørensen, T. B. [MS06].



Sensible Nash equilibrium



Questionable Nash equilibrium

In Guess-the-Ace, at the start a standard 52-card deck is perfectly shuffled, face down, by a dealer. Then, Player 1 can decide whether to immediately end the game, at which point no money is transferred between the players, or offer \$1000 to Player 2 if they can correctly guess whether the top card of the shuffled deck is the ace of spaces or not. If Player 2 guesses correctly, the \$1000 get transferred from Player 1 to Player 2; if not, no money is transferred. The game tree is summarized in Example 1.1.

Clearly, the only Nash equilibrium strategy for Player 1 is to quit immediately, or they are guaranteed to *lose* money. Since Player 2 does not get to play, any strategy for Player 2 is a Nash equilibrium strategy.

*These notes are class material that has not undergone formal peer review. The TAs and I are grateful for any reports of typos.

In particular, both highlighted equilibria in Example 1.1 are Nash equilibria. However, the two equilibria are significantly different from a practical point of view. Imagine that Player 2 is a bot playing against opponents in the real world, blindly following the Nash equilibrium strategy it has precomputed. If Player 1 makes a mistake and decides to offer the \$1000 instead of immediately quitting, the Nash equilibrium that bets that the top card is *not* the ace of space has an expected utility of $>$ \$980 whereas the Nash equilibrium that bets that the top card is the ace of spade only has an expected utility of $<$ \$20.

So, while both strategy profiles in Example 1.1 are Nash equilibria, only one of the two is *"sensible"*.

Formalizing this subtle notion of rationality within the set of Nash equilibria has been a major endeavor for the game-theoretic literature in the 70s and 80s. Today, we say that the equilibrium in Example 1.1 (Left) is *sequentially irrational*, while the one on the right is *sequentially rational*. The takeaway lesson is:

Remark 1.1. Not all Nash equilibria are equally "good" when the agents can make mistakes. Specifically, sequentially-irrational Nash equilibria might leave value on the table, by being incapable of capitalizing on opponents' mistakes.

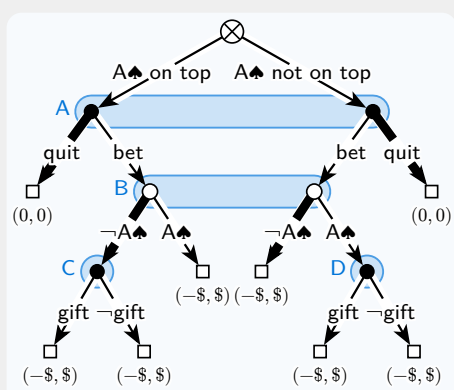
The goal of this lecture is to investigate how one can rule out sequential irrationality and compute a sequentially-rational Nash equilibrium in a two-player zero-sum imperfect-information game.

2 Undomination is not the solution

One might believe that the problem of sequential irrationality is that of picking dominated strategies. So, one might be inclined to look into the problem of finding a Nash equilibrium whose support does not include any (weakly) dominated strategy (the concept is not immediately well defined, but for the purposes of this discussion let's restrict ourselves to Nash equilibria in deterministic strategies).

Unfortunately, domination of strategies is not the root cause of sequential irrationality, and therefore undomination is not its solution. Indeed, as much as undomination *does* get rid of the undesirable behavior of Example 1.1 (Right), since action '$\mathsf{A}\spadesuit$' is strictly dominated by action '$\neg\mathsf{A}\spadesuit$', it does not prevent sequential irrationality in more complex settings, such as Example 2.1.

Example 2.1. Undomination does *not* prevent a player from playing risky actions, hoping for an opponent's mistake. In this example, again due to Miltersen, P. B., & Sørensen, T. B. [MS06], the Guess-the-Ace game is slightly modified in that, when Player 2 guesses wrong, Player 1 can decide whether they still want to give \$1000 to Player 2 out of the kindness of their heart or not. By introducing that possibility, action '$\neg\mathsf{A}\spadesuit$' is not strictly dominating anymore, because Player 2 might still hope that the second gift of \$1000 is given only when the insensible guess '$\mathsf{A}\spadesuit$' is made. So, the second takeaway lesson for today's class is the following.



3 Trembling-hand refinements

The issue of sequential irrationality stems from the fact that some parts of the game tree are unreachable at equilibrium. For those excluded parts of the game tree, any strategy can be picked without affecting the equilibrium. The idea behind trembling-hand refinements is simple: to avoid sequential irrationality, it forces all players to explore the whole game tree. It does so by forcing the players to *tremble*, that is, by constraining them to play all actions at all decision points with a strictly positive lower bound

probability that grows as a function of a hyperparameter $\epsilon > 0$. For each $\epsilon > 0$, a Nash equilibrium subject to the trembling constraints is found. A trembling-hand refinements is then any limit points of such Nash equilibria as $\epsilon \to 0^+$.

Different equilibrium notions differ as to how the lower bounds are set as a function of ϵ. We will see two, which are the two best known: extensive-form perfect equilibrium and quasi-perfect equilibrium.

3.1 Extensive-form perfect equilibrium (EFPE)

Extensive-form perfect equilibrium (EFPE), due to Selten, R. [Sel75], is conceptually the simplest of the two. In an EFPE, the trembles are *behavioral*: given $\epsilon > 0$, the perturbed game simply mandates that every action at every decision point must be picked with probability at least ϵ.

Since our game solving formalism is based around the sequence-form representation of strategies, it is important to check that those behavioral trembling constraints can be expressed in the sequence form. That is the case: asking that action a at decision point j of Player 1 be selected with probability at least ε corresponds to the sequence-form constraint

$$x_{ja} \geq \begin{cases} \epsilon & \text{if } \; p_j = \varnothing \\ \epsilon \cdot x_{p_j} & \text{otherwise.} \end{cases} \tag{1}$$

Collecting all sequence-form trembling constraints (1) constraints across all decision points $j \in \mathcal{J}$ and actions $a \in A_j$ of Player 1, we can express the whole set of trembling constraints in matrix form as $M_1(\epsilon)x \geq m_1(\epsilon)$. (An analogous statement holds for Player 2). So, given any $\epsilon > 0$, and indicating with $F_1 x = f_1, x \geq 0$ and $F_2 y = f_2, y \geq 0$ the polytope of sequence form strategies of Player 1 and Player 2 respectively, a Nash equilibrium strategy for Player 1 under the trembling constraints can be expressed as the saddle point problem

$$\begin{cases} \max\limits_{x} \; \min\limits_{y} \, x^\top \mathrm{U}_1 y \\ \text{s.t.} \; ① \; F_2 y = f_2 \\ \qquad ② \; M_2(\epsilon) y \geq m_2(\epsilon) \\ \qquad ③ \; F_1 y = f_1 \\ \qquad ④ \; M_1(\epsilon) x \geq m_1(\epsilon). \end{cases} \tag{EFPE}$$

We will look into how to compute a limit point of solutions to (EFPE) as $\epsilon \to 0^+$ in Section 4.

3.2 Quasi-perfect equilibrium (QPE)

Quasi-perfected equilibrium (QPE), introduced by van Damme, E. [van84], is a bit more intricate than EFPE. Specifically, while in an EFPE each trembling constraints mandates a lower bound of ϵ on the probability of playing each *action*, in the case of a QPE the lower bounds are given on the probability of each *sequence* of actions. More precisely, for any $\epsilon > 0$ and player $i \in \{1, 2\}$, let $\ell_i : \mathbb{R}_{>0} \to \mathbb{R}_{>0}^{\Sigma_i}$ denote the vector parametrized on ϵ and indexed on the sequences Σ_i of Player i, whose entries are defined as

$$\ell_{i,\sigma}(\epsilon) = \epsilon^{|\sigma|} \qquad \forall \sigma \in \Sigma_i, \tag{3}$$

where $|\sigma|$ denotes the number of actions for Player i in the sequence σ. Miltersen, P. B., & Sørensen, T. B. [MS10] proved that any limit point of the solution to the perturbed optimization problem

$$\begin{cases} \max\limits_{x} \; \min\limits_{y} \, x^\top \mathrm{U}_1 y \\ \text{s.t.} \; ① \; F_2 y = f_2 \\ \qquad ② \; y \geq \ell_2(\epsilon) \\ \qquad ③ \; F_1 y = f_1 \\ \qquad ④ \; x \geq \ell_1(\epsilon) \end{cases} \tag{QPE}$$

is a QPE. (Recently, Gatti, N., Gilli, M., & Marchesi, A. [GGM20] took this construction further, and showed that *any* QPE can be expressed as a limit point of solutions to (QPE), as long as more general vectors of polynomials ℓ_1, ℓ_2 are used than in (3). In this paper we will focus on Miltersen-Sørensen-style perturbation as defined in (3).)

Once again, we will discuss how to compute a limit point of solutions to (QPE) as $\epsilon \to 0^+$ in Section 4.

3.3 Relationship between the equilibria

We already know from Section 2 that undomination does not imply sequential rationality. Interestingly, the converse also is not true in general. So, undomination and sequential rationality are actually incomparable concepts, in the sense that neither implies the other.

At this point, one might naturally wonder whether a refinement that is both undominated and sequentially-rational can be devised. The answer is yes: a nice property of QPE is that not only it is sequentially rational, but it is also undominated! The same cannot be said of EFPE. So, as Mertens, J.-F. [Mer95] noted, a *quasi-perfect equilibrium* is nowadays considered superior to EFCE.

> *"Observe that the "quasi-perfect" equilibria [..] are still sequential—and sequential equilibria have all backward-induction properties (*e.g.*, Kohlberg and Mertens, 1986)—but are at the same time normal form perfect—which can be viewed as the strong version of undominated. (And every proper equilibrium is quasi-perfect.) Thus, by some irony of terminology, the "quasi"-concept seems in fact far superior to the original unqualified perfection itself."*

The relationship between the different refinements is summarized in the Venn diagram of Figure 3.

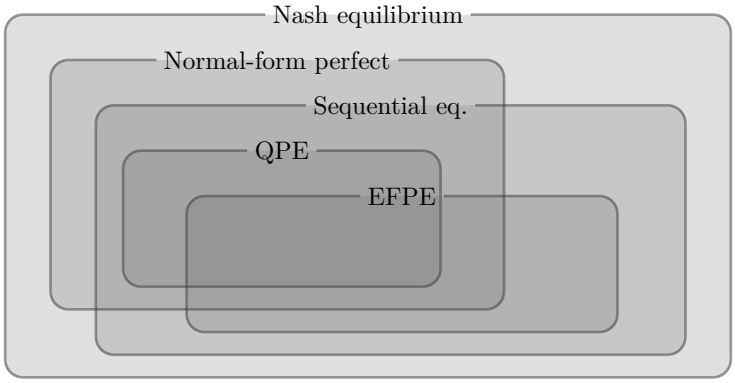


Figure 3: Relationship between the different Nash equilibrium refinements

3.4 Computational complexity

Perhaps surprisingly, finding an EFPE or a QPE in a two-player game is not harder than finding a Nash equilibrium. In particular, in zero-sum games, an EFPE and a QPE can be found in polynomial time in the size of the input game. Table 1 summarizes the computational complexity of computing the Nash equilibrium refinements mentioned so far in two-player games.

Solution concept	General-sum	Zero-sum
Nash equilibrium (NE)	PPAD-complete [DGP09]	FP [Rom62; von96]
Subgame perfect equilibrium (SPE)	PPAD-complete	FP
Quasi perfect equilibrium (QPE)	PPAD-complete [MS10]	FP [MS10]
Extensive-form perfect equilibrium (EFPE)	PPAD-complete [FG17]	FP [FG17]

Table 1: Complexity of computing different Nash equilibrium refinements in two-player games.

4 Trembling linear programs and computation of QPE and EFPE

We can compute a limit point of solutions to (EFPE) and (QPE) using the same machinery. As a first step, just like what we did for the Nash equilibrium, we convert the bilinear saddle-point formulations (EFPE), (QPE) into linear programs by dualizing the internal minimization problems. This gives us a linear program where the constraints matrix and the objective function depend polynomially on ϵ. In particular, for both QPE and EFPE we end up with a linear program of the form

$$P(\epsilon) : \begin{cases} \max\limits_{x} & c(\epsilon)^\top x \\ \text{s.t.} & A(\epsilon)x = b(\epsilon) \\ & x \geq 0. \end{cases}$$

where c, A and b are *polynomial* functions of ϵ with rational coefficients. We will call an object of that form a *trembling linear program (TLP)*, and a limit point of solutions to $P(\epsilon)$ as $\epsilon \to 0^+$ a *limit solution* of the TLP. With this formalism, we can reframe the computation of an EFPE or a QPE as the problem of finding a limit solution to their corresponding TLPs.

We will now discuss the complexity of solving a TLP, and two different computational approaches. Both of them are based on the concept of *basis stability* (Recall that a *basis* of an LP is a subset of the program's variables such that when only those columns of matrix A that correspond to those variables are included in a new matrix A', the new matrix A' is invertible [BT97] (page 55).

Definition 4.1 (Stable basis). Let $P(\epsilon)$ be a TLP. The LP basis B is said to be *stable* if there exists $\overline{\epsilon} > 0$ such that B is optimal for $P(\epsilon)$ for all $\epsilon : 0 < \epsilon \leq \overline{\epsilon}$.

If a stable basis were to be found, from there a limit solution of $P(\epsilon)$ could be computed in polynomial time. As it turns out, a stable basis always exists, and can be computed in polynomial time.

4.1 Negligible Positive Perturbations (NPP)

Farina, G., Gatti, N., & Sandholm, T. [FGS18], extending prior work by Miltersen, P. B., & Sørensen, T. B. [MS10] and Farina, G., & Gatti, N. [FG17], showed the following.

Theorem 4.1 (Farina, G., Gatti, N., & Sandholm, T. [FGS18]). Given as input a TLP $P(\epsilon)$, there exists $\epsilon^* > 0$—called a *negligible positive perturnation (NPP)*—such that for all $0 < \overline{\epsilon} \leq \epsilon^*$, any optimal basis for the numerical LP $P(\overline{\epsilon})$ is stable. Furthermore, such a value ϵ^* can be computed in polynomial time in the input size, assuming that a polynomial of degree d requires $\Omega(d)$ space in the input.[1]

So, at least in principle, a solution to a TLP $P(\epsilon)$ could be computed as follows:

- First, compute the value of the NPP ϵ^* using the constructive proof of Theorem 4.1.
- Then, solve the numerical linear program $P(\epsilon^*)$ to optimality. Since the bit complexity of ϵ^* is polynomial in the size of the TLP, the numerical LP can be solved to optimality in polynomial time, and a basis B can be extracted. From Theorem 4.1, such a basis is stable (Definition 4.1)
- Finally, extract the limit solution to the TLP from the stable basis.

The algorithm just described has polynomial complexity in the TLP size. In the case of the TLP arising form QPE and EFPE, that translates into a polynomial-time algorithm to find an exact EFPE and QPE in a two-player zero-sum game (see also Table 1).

[1] If this were not the case, evaluating a polynomial in an integer n would not be an efficient operation, since it requires $\Omega(d \log n)$ bits to represent the output.

4.2 A significantly more scalable approach

While technically polynomial, the NPP-based algorithm described in the previous subsection is mostly of conceptual interest. In practice, because the value of the NPP is so small, any linear programming solver that wants to have a chance at solving the numerical linear program $P(\epsilon^*)$ must—as a minimum—use rational arithmetic, rendering the algorithm extremely slow.

A significantly more scalable algorithm for solving TLPs, due to Farina, G., Gatti, N., & Sandholm, T. [FGS18], avoids the pessimistically small numerical NPP ϵ^* of Theorem 4.1 by using an efficient stability-checking oracle for checking if a basis is stable or not.

The iterative algorithm repeatedly picks a numerical perturbation $\bar{\epsilon}$, computes an optimal basis for the perturbed LP $P(\bar{\epsilon})$, and queries the basis-stability oracle. If the basis is not stable, the algorithm concludes that the perturbation value $\bar{\epsilon}$ was too optimistic, and a new iteration is performed with a smaller perturbation reduced by a multiplicative constant (for example, divide it by 1000). On the other hand, if the basis is stable, the algorithm takes the limit of the LP solution and returns it as the limit solution of the TLP. Correctness and termination are guaranteed by the following observation.

Remark 4.1. Any value of $\bar{\epsilon}$ in the range $(0, \epsilon^*]$ guarantees termination of the algorithm. Indeed, by Theorem 4.1, any optimal basis for $P(\bar{\epsilon})$ is stable and makes our iterative algorithm terminate. Furthermore, if after every negative stability test the value of $\bar{\epsilon}$ is reduced by a constant multiplicative factor (*e.g.*, halved), then since ϵ^* only has a polynomial number of bits, the algorithm terminates after trying at most a polynomial number of different values for $\bar{\epsilon}$.

The practical algorithm is 3-4 orders of magnitude faster than the conceptual algorithm described in Section 4.1, and is the current state-of-the-art algorithm for computing QPE and EFPE.

Bibliography

[MS06] P. B. Miltersen and T. B. Sørensen, "Computing sequential equilibria for two-player games," in *Symposium on Discrete Algorithms (SODA)*, 2006.

[Sel75] R. Selten, "Reexamination of the perfectness concept for equilibrium points in extensive games," *International journal of game theory*, 1975.

[van84] E. van Damme, "A relation between perfect equilibria in extensive form games and proper equilibria in normal form games," *International Journal of Game Theory*, 1984.

[MS10] P. B. Miltersen and T. B. Sørensen, "Computing a Quasi-Perfect Equilibrium of a Two-Player Game," *Economic Theory*, vol. 42, no. 1, 2010.

[GGM20] N. Gatti, M. Gilli, and A. Marchesi, "A characterization of quasi-perfect equilibria," *Games and Economic Behavior*, vol. 122, pp. 240–255, 2020.

[Mer95] J.-F. Mertens, "Two examples of strategic equilibrium," *Games and Economic Behavior*, vol. 8, no. 2, pp. 378–388, 1995.

[DGP09] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The Complexity of Computing a Nash Equilibrium," *Commun. ACM*, vol. 52, no. 2, pp. 89–97, 2009.

[Rom62] I. Romanovskii, "Reduction of a Game with Complete Memory to a Matrix Game," *Soviet Mathematics*, vol. 3, 1962.

[von96] B. von Stengel, "Efficient Computation of Behavior Strategies," *Games and Economic Behavior*, vol. 14, no. 2, pp. 220–246, 1996.

[FG17] G. Farina and N. Gatti, "Extensive-Form Perfect Equilibrium Computation in Two-Player Games," in *AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

[BT97] D. Bertsimas and J. Tsitsiklis, "Introduction to linear optimization," 1997.

[FGS18] G. Farina, N. Gatti, and T. Sandholm, "Practical Exact Algorithm for Trembling-Hand Equilibrium Refinements in Games," in *Neural Information Processing Systems (NeurIPS)*, 2018.

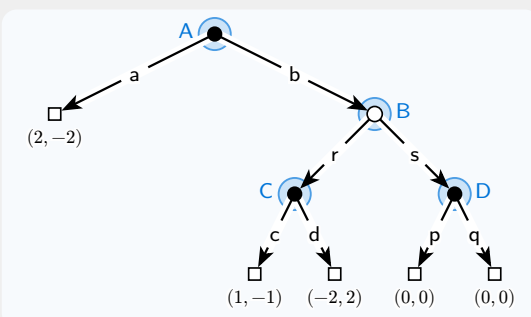
[KW82] D. M. Kreps and R. Wilson, "Sequential Equilibria," *Econometrica*, vol. 50, no. 4, pp. 863–894, 1982.

A Why not uniform lower bounds in QPE?

Not all vanishing perturbations $\ell_1(\epsilon), \ell_2(\epsilon)$ in the QPE formulation (QPE) lead to a sequentially-rational equilibrium. For example, it is natural to wonder whether it is *really* necessary to consider lower bounds of the form ϵ^σ instead of, for example, the uniform lower bound ϵ for all sequences. After all, %the exponential dependence on the length of the sequences is only %detrimental to the practical performance of QPE solvers, and surely a uniform lower bound of ϵ would still force the whole game to be explored, wouldn't it? While appealing on the surface, such a uniform lower bound might result in a solution that is not even subgame perfect, much less sequentially rational!

We illustrate this point with an example.

Example 1.1. Consider the following simple game.



Player	Action	Probability
Player 1 (black)	a	$1-4\epsilon$
Player 1 (black)	b	4ϵ
Player 1 (black)	c, d, p, q	$1/2$
Player 2 (white)	r	$1-\epsilon$
Player 2 (white)	s	ϵ

For any choice of $\epsilon \in \left[0, \frac{1}{4}\right]$, we now argue that the only Nash equilibrium of the perturbed game assigns probability $1-\epsilon$ to action r of Player 2, and probability $\frac{1}{2}$ to actions c and d of Player 1. Indeed, action a strictly dominates b, since all payoffs for the black player (Player 1) are strictly lower in the subtree rooted at b. Hence, the black player must minimize the probability mass put on the sequences that contain action b, compatibly with lower bounds. Because we are using uniform lower bounds ϵ on the probability of each sequence, the black player will need to put at least probability ϵ on the four sequences bc, bd, bp, bq. This can be achieved when c, d, p, q are each selected with probability $1/2$ and action b with probability 4ϵ. From the point of view of the white player (Player 2), information set C guarantees an expected utility of $-1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = \frac{1}{2}$, while information set D guarantees and expected utility of 0. So, it is rational for the white player to put as much probability mass as allowed by the lower bounds to action r. This is achieved when action r is selected with probability $1-\epsilon$, and action s with probability ϵ.

So, as $\epsilon \to 0^+$, any limit point sees Player 2 pick action r with probability 1 and Player 1 randomizing uniformly between actions c and d, despite action d being strictly dominated. Thus, both players will act irrationally (with Player 1 not even playing a best response in the subtree rooted at C) should Player 1 make the mistake of picking action b instead of a at the root A. The resulting equilibrium is not sequentially rational. (In fact, it's not even subgame perfect, which is even stronger [KW82].)