CS364A: Algorithmic Game Theory
Lecture #16: Best-Response Dynamics*

Tim Roughgarden†

November 13, 2013

1 Do Players Learn Equilibria?

In this lecture we segue into the third part of the course, which studies the following questions.

1. Do we expect strategic players do reach an equilibrium of a game, even in principle?

2. If so, will it happen quickly? As we'll see, theoretical computer science is well suited to contribute both positive and negative answers to this question.

3. If so, how does it happen? Which learning algorithms quickly converge to an equilibrium?

Affirmative answers to these questions are important because they justify equilibrium analysis. Properties of equilibria, such as a near-optimal objective function value, are not obviously relevant when players fail to find one. More generally, proving that natural learning algorithms converge quickly to an equilibrium lends plausibility to the predictive power of an equilibrium concept.

To reason about the three questions above, we require a behavioral model — "dynamics" — for players when not at an equilibrium. Thus far, we've just assumed that equilibria persist and that non-equilibria don't. This lecture focuses on variations of "best-response dynamics," while the next two lectures study dynamics based on regret-minimization.

No concrete model of learning will be fully convincing. We aspire toward simple and natural learning algorithms that lead to concrete and non-trivial predictions. Ideally, we would like to reach the same conclusion via multiple different learning processes. Then, even though we might not literally believe in any of the specific learning algorithms, we can have some confidence that the conclusion is robust, and not an artifact of the details of a particular learning process.

†Department of Computer Science, Stanford University, 462 Gates Building, 353 Serra Mall, Stanford, CA 94305. Email: tim@cs.stanford.edu.

2 Best-Response Dynamics

Best-response dynamics is a straightforward procedure by which players search for a pure Nash equilibrium (PNE) of a game. Specifically:

- While the current outcome $\mathbf{s}$ is not a PNE:
  - Pick an arbitrary player i and an arbitrary beneficial deviation s_i' for player i, and move to the outcome $(s_i', \mathbf{s}_{-i})$.

There might be many choices for the deviating player i and for the beneficial deviation s_i'. We leave both underspecified for the moment, specializing it later as needed.[1]

Best-response dynamics can only halt at a PNE — it cycles in any game without one. It can also cycle in games that have a PNE (see the Exercises).

Best-response dynamics is a perfect fit for *potential games*, discussed in passing in Lectures 13 and 15. Recall that a potential game admits a real-valued function Φ, defined on the outcomes of the game, such that for every outcome $\mathbf{s}$, every player i, and every unilateral deviation s_i' by i,

$$\Phi(s_i', \mathbf{s}_{-i}) - \Phi(\mathbf{s}) = C_i(s_i', \mathbf{s}_{-i}) - C_i(\mathbf{s}), \tag{1}$$

where C_i denotes player i's cost (or negative payoff) function. That is, a potential function tracks deviators' cost changes under unilateral deviations. Routing games (Lecture 12), location games (Lecture 14), and network cost-sharing games (Lecture 15) are all potential games.

We already know that potential games have PNE — the potential function minimizer is one. Best-response dynamics offer a more constructive proof of this fact.

Proposition 2.1 ([3]) *In a finite potential game, from an arbitrary initial outcome, best-response dynamics converges to a PNE.*

Proof: In every iteration of best-response dynamics, the deviator's cost strictly decreases. By (1), the potential function strictly decreases. Thus, no cycles are possible. Since the game is finite, best-response dynamics eventually halts, necessarily at a PNE. ■

Proposition 2.1 gives an affirmative answer to the first question of Section 1 for potential games — there is a natural procedure by which players can reach a PNE. Next we turn to the second question — how fast does this happen?

We consider three notions of "fast convergence," from strongest to weakest. The best-case scenario would be that best-response dynamics converges to a PNE in a polynomial number of iterations.[2] This strong conclusion is true when the potential function Φ can take on only

[1] This procedure is sometimes called "better-response dynamics," with the term "best-response dynamics" reserved for the version in which s_i' is chosen to maximize i's payoff (given $\mathbf{s}_{-i}$).

[2] Meaning polynomial in the number of players and the total number of players' strategies. The number of outcomes is exponential in the number of players — if there are n players each with two strategies, there are 2^n outcomes. Thus, the easy fact that the number of iterations of best-response dynamics is at most the number of outcomes of a potential game is not interesting in games with many players.

polynomially many distinct values (e.g., if it is integer-valued and polynomially bounded). In general, however, the potential function can take on exponentially many different values and best-response dynamics can decrease the potential function very slowly, requiring an exponential number of iterations to converge (see also Lecture 19).

3 Fast Convergence to ϵ-PNE in Symmetric Routing Games

The second notion of "fast convergence" settles for an approximate Nash equilibrium.

Definition 3.1 (ϵ-Pure Nash Equilibrium) For $\epsilon \in [0,1]$, an outcome $\mathbf{s}$ of a cost-minimization game is an *ϵ-pure Nash equilibrium (ϵ-PNE)* if, for every player i and deviation $s'_i \in S_i$,

$$C_i(s'_i, \mathbf{s}_{-i}) \geq (1-\epsilon) \cdot C_i(\mathbf{s}). \tag{2}$$

This is essentially the same definition we used in Lecture 14, reparametrized for convenience. An ϵ-PNE in this lecture corresponds to a $\frac{1}{1-\epsilon}$-PNE in Lecture 14.

We next study *ϵ-best response dynamics*, in which we only permit moves that result in "significant" improvements. This is the key to converging much faster than under standard best-response dynamics. Precisely:

- While the current outcome $\mathbf{s}$ is not a ϵ-PNE:
  - Pick an arbitrary player i that has an *ϵ-move* — a deviation s'_i with $C_i(s'_i, \mathbf{s}_{-i}) < (1-\epsilon)C_i(\mathbf{s})$ — and an arbitrary such move for the player, and move to the outcome $(s'_i, \mathbf{s}_{-i})$.

ϵ-best response dynamics can only halt at an ϵ-PNE, and it eventually converges in every finite potential game. But how quickly?

Theorem 3.2 (Convergence of ϵ-Best Response Dynamics [2]) *Consider an atomic selfish routing game where:*

1. *All players have a common source vertex and a common sink vertex.*
2. *Cost functions satisfy the "α-bounded jump condition," meaning $c_e(x+1) \in [c_e(x), \alpha \cdot c_e(x)]$ for every edge e and positive integer x.*
3. *The MaxGain variant of ϵ-best-response dynamics is used: in every iteration, among players with an ϵ-move available, the player who can obtain the biggest absolute cost decrease moves to its minimum-cost deviation.*

Then, an ϵ-PNE is reached in $(\frac{k\alpha}{\epsilon} \log \frac{\Phi(\mathbf{s}^0)}{\Phi_{\min}})$ iterations.

Non-trivial constructions show that the first two hypotheses are necessary — if either is dropped, all variants of best-response dynamics take an exponential number of iterations in the worst case [5]; see also Lecture 19. In the third hypothesis, it is important that ϵ-best-response dynamics is used instead of standard best-response dynamics, but the result continues to hold for all natural variations of ϵ-best-response dynamics (see Exercises). Essentially, the only requirement is that every player is given the opportunity to move sufficiently often [2].

The plan for proving Theorem 3.2 is to strengthen quantitatively the proof of Proposition 2.1 — to prove that every iteration of ϵ-best-response dynamics decreases the potential function by *a lot*. We need two lemmas. The first one guarantees the existence of a player with high cost; if this player is chosen to move in an iteration, then the potential function decreases significantly. The issue is that some other player might be chosen to move by ϵ-best-response dynamics. The second lemma, which is the one that needs the hypotheses in Theorem 3.2, proves that the player chosen to move has cost within an α factor of that of any other player. This is good enough for fast convergence.

Lemma 3.3 *In every outcome* $\mathbf{s}$*, there is a player* i *with* $C_i(\mathbf{s}) \geq \Phi(\mathbf{s})/k$*.*

Proof: Recall from Lecture 12 that in a selfish routing game, the objective function is

$$\mathrm{cost}(\mathbf{s}) = \sum_{e \in E} f_e \cdot c_e(f_e),$$

where f_e is the number of players that choose a strategy including the edge e, while the potential function is

$$\Phi(\mathbf{s}) = \sum_{e \in E} \sum_{i=1}^{k} c_e(i).$$

Since cost functions are nondecreasing, $\Phi(\mathbf{s}) \leq \mathrm{cost}(\mathbf{s})$ for every outcome $\mathbf{s}$.

Since $\mathrm{cost}(\mathbf{s}) = \sum_{i=1}^{k} C_i(\mathbf{s})$ in a selfish routing game, some player has cost at least as large as the average, and

$$\max_{i=1}^{k} C_i(\mathbf{s}) \geq \frac{\mathrm{cost}(\mathbf{s})}{k} \geq \frac{\Phi(\mathbf{s})}{k},$$

as claimed. ■

The next lemma effectively relates the cost of the deviating player in ϵ-best-response dynamics to those of the other players.

Lemma 3.4 *Suppose player* i *is chosen in the outcome* $\mathbf{s}$ *by MaxGain* ϵ*-best-response dynamics, and takes the* ϵ*-move* s_i'*. Then*

$$C_i(\mathbf{s}) - C_i(s_i', \mathbf{s}_{-i}) \geq \frac{\epsilon}{\alpha} C_j(\mathbf{s}) \tag{3}$$

for every other player j*.*

Note that the definition of an ϵ-move is that $C_i(\mathbf{s}) - C_i(s'_i, \mathbf{s}_{-i}) \geq \epsilon C_i(\mathbf{s})$. Thus (3) states that, up to a factor of α, the cost decrease enjoyed by i is at least as large as that of any other player taking an ϵ-move (whether that player has an ϵ-move available or not).

Proof of Lemma 3.4: Fix the player j. If j has an ϵ-move — which decreases player j's cost by at least $\epsilon C_j(\mathbf{s})$ — then (3) holds, even without the α, simply because i was chosen over j in MaxGain dynamics.

The trickier case is when the player j has no ϵ-move available. We use here that all players have the same strategy set: if s'_i is such a great deviation for player i, why isn't it for player j as well? That is, how can it be that

$$C_i(s'_i, \mathbf{s}_{-i}) \leq (1-\epsilon)C_i(\mathbf{s}) \tag{4}$$

while

$$C_j(s'_i, \mathbf{s}_{-j}) \geq (1-\epsilon)C_j(\mathbf{s})? \tag{5}$$

A key observation is that the outcomes $(s'_i, \mathbf{s}_{-i})$ and $(s'_i, \mathbf{s}_{-j})$ have at least $k-1$ strategies in common — s'_i, played by i in the former outcome and by j in the latter, and the $k-2$ fixed strategies played by players other than i and j. Since the two outcomes differ in only one chosen strategy, the load on every edge differs by at most one in the two outcomes. By the α-bounded jump condition in Theorem 3.2, the cost of every edge differs by at most a factor of α in the two outcomes. In particular:

$$C_j(s'_i, \mathbf{s}_{-j}) \leq \alpha \cdot C_i(s'_i, \mathbf{s}_{-i}). \tag{6}$$

Note that both sides of inequality (6) reference the player that chooses strategy s'_i (j on the left-hand side, i on the right).

The inequalities (4)–(6) are compatible only if $C_j(\mathbf{s}) \leq \alpha \cdot C_i(\mathbf{s})$. Combining this with (4) yields $C_i(\mathbf{s}) - C_i(s'_i) \geq \epsilon \cdot C_i(\mathbf{s}) \geq \frac{\epsilon}{\alpha} \cdot C_j(\mathbf{s})$, as required. ■

Lemma 3.3 guarantees that there is always a player that, if chosen to make an ϵ-move, rapidly decreases the potential function. Lemma 3.4 extends the conclusion to the player that is actually chosen to move. It is now a simple matter to upper bound the number of iterations required for convergence.

Proof of Theorem 3.2: In an iteration of ϵ-best-response dynamics where player i switches to the (ϵ-move) s'_i,

$$\begin{aligned} \Phi(\mathbf{s}) - \Phi(s'_i, \mathbf{s}_{-i}) &= C_i(\mathbf{s}) - C_i(s'_i, \mathbf{s}_{-i}) && (7) \\ &\geq \frac{\epsilon}{\alpha} \cdot \max_{j=1}^{k} C_j(\mathbf{s}) && (8) \\ &\geq \frac{\epsilon}{\alpha k} \cdot \Phi(\mathbf{s}), && (9) \end{aligned}$$

where equation (7) follows from the definition of a potential function (Lecture 14) and inequalities (8) and (9) follow from Lemmas 3.4 and 3.3, respectively.

The upshot is that every iteration of ϵ-best-response dynamics decreases the potential function by at least a factor of $(1 - \frac{\epsilon\alpha}{k})$. Thus, every $\frac{k}{\epsilon\alpha}$ iterations decrease the potential function by a constant factor.[3] Since the potential function begins at the value $\Phi(\mathbf{s}^0)$ and cannot drop lower than $\Phi_{\min}$, ϵ-best-response dynamics converges in $O(\frac{k}{\epsilon\alpha} \log \frac{\Phi(\mathbf{s}^0)}{\Phi_{\min}})$ iterations. ■

Theorem 3.2 is good justification for performing equilibrium analysis in atomic selfish routing games with a shared source and sink: many variations of the natural ϵ-best-response dynamics converge quickly to an approximate equilibrium. Unfortunately, Theorem 3.2 cannot be extended much further. In atomic routing games with multiple sources and sinks, for example, ϵ-best-response dynamics can require an exponential number of iterations to converge, no matter how the deviating player and deviation in each iteration are chosen [5].

4 Fast Convergence to Low-Cost Outcomes in Smooth Potential Games

This section explores our third and final notion of "fast convergence": quickly reaching outcomes with objective function value *as good as if* players had successfully converged to an (approximate) equilibrium. This is a weaker guarantee — it does not imply convergence to an approximate equilibrium — but is still quite compelling. In situations where the primary reason for equilibrium analysis is a performance (i.e., POA) bound, this weaker guarantee is a costless surrogate for equilibrium convergence.

Weakening our notion of fast convergence enables positive results with significantly wider reach. The next result applies to all potential games that are smooth (Lecture 14), including routing games (with arbitrary sources, sinks, and cost functions) and location games.

Theorem 4.1 ([1, 4]) *Consider a (λ, μ)-smooth cost-minimization game[4] with a positive potential function Φ that satisfies $\Phi(\mathbf{s}) \leq cost(\mathbf{s})$ for every outcome s. Let $\mathbf{s}^0, \ldots, \mathbf{s}^T$ be a sequence generated by MaxGain best-response dynamics, $\mathbf{s}^*$ a minimum-cost outcome, and $\eta > 0$ a parameter. Then all but*

$$O\left(\frac{k}{\eta(1-\mu)} \log \frac{\Phi(\mathbf{s}^0)}{\Phi_{\min}}\right) \tag{10}$$

outcomes $\mathbf{s}^t$ satisfy

$$cost(\mathbf{s}^t) \leq \left(\frac{\lambda}{1-\mu} + \eta\right) \cdot cost(\mathbf{s}^*),$$

where $\Phi_{\min}$ is the minimum potential function value of an outcome and k is the number of players.

[3] To see this, note that $(1-x)^{1/x} \leq (e^{-x})^{1/x} = \frac{1}{e}$ for $x \in (0, 1)$.

[4] Recall from Lecture 14 that this means that $\sum_{i=1}^{k} C_i(s_i^*, \mathbf{s}_{-i}) \leq \lambda \cdot \text{cost}(\mathbf{s}^*) + \mu \cdot \text{cost}(\mathbf{s})$ for every pair $\mathbf{s}, \mathbf{s}^*$ of outcomes.

The dynamics in Theorem 4.1 differ from those in Theorem 3.2 only in that the restriction to ϵ-moves is dropped. That is, in each iteration, the player i and deviation s_i' are chosen to maximize $C_i(\mathbf{s}) - C_i(s_i', \mathbf{s}_{-i})$.

Theorem 4.1 states that for all but a small number of outcomes in the sequence, the cost is essentially as low as if best-response dynamics had already converged to a PNE. These "bad outcomes" need not be successive — since an iteration of best-response dynamics can strictly increase the overall cost, a "good outcome" can be followed by a bad one.

Proof of Theorem 4.1: In the proof of Theorem 3.2, we showed that every iteration of the dynamics significantly decreased the potential function, thus limiting the number of iterations. Here, the plan is to show that whenever there is a *bad state* $\mathbf{s}$ — one that fails to obey the guarantee in (10) — the potential function decreases significantly. This will yield the desired bound on the number of bad states.

For an outcome $\mathbf{s}^t$, define $\delta_i(\mathbf{s}^t) = C_i(\mathbf{s}^t) - C_i(s_i^*, \mathbf{s}_{-i}^t)$ as the cost decrease i experiences by switching to s_i^*, and $\Delta(\mathbf{s}^t) = \sum_{i=1}^k \delta_i(\mathbf{s}^t)$. The value $\delta(\mathbf{s}^t)$ is nonpositive when $\mathbf{s}^t$ is a PNE, but in general it can be positive or negative. Using this notation and the smoothness assumption, we can derive

$$\mathrm{cost}(\mathbf{s}^t) \leq \sum_{i=1}^k C_i(\mathbf{s}^t) = \sum_{i=1}^k \left[C_i(s_i^*, \mathbf{s}_{-i}^t) + \delta_i(\mathbf{s}^t)\right] \leq \lambda \cdot \mathrm{cost}(\mathbf{s}^*) + \mu \cdot \mathrm{cost}(\mathbf{s}^t) + \sum_{i=1}^k \delta_i(\mathbf{s}^t),$$

and hence

$$\mathrm{cost}(\mathbf{s}^t) \leq \frac{\lambda}{1-\mu} \cdot \mathrm{cost}(\mathbf{s}^*) + \frac{1}{1-\mu} \sum_{i=1}^k \delta_i(\mathbf{s}^t).$$

This inequality is stating that the cost of an outcome $\mathbf{s}^t$ is large only when the amount $\sum_i \delta_i(\mathbf{s}^t)$ players have to gain by unilateral deviations is large. It reduces the theorem to proving that only $O(\frac{k}{\eta(1-\mu)} \log \frac{\Phi(\mathbf{s}^0)}{\Phi_{\min}})$ states $\mathbf{s}^t$ are bad. Along the lines of Theorem 3.2, we next prove that bad states lead to a large decrease in the potential function.

In a bad state $\mathbf{s}^t$, since Φ underestimates cost,

$$\Delta(\mathbf{s}^t) \geq \eta(1-\mu)\mathrm{cost}(\mathbf{s}^t) \geq \eta(1-\mu)\Phi(\mathbf{s}^t).$$

If a player i chooses a best response in the outcome $\mathbf{s}^t$, its cost decreases by at least $\delta_i(\mathbf{s}^t)$. Thus, in a bad state $\mathbf{s}^t$, the cost of the player chosen by maximum-gain best-response dynamics decreases by at least $\frac{\eta(1-\mu)}{k}\Phi(\mathbf{s}^t)$. Since Φ is a potential function, $\Phi(\mathbf{s}^{t+1}) \leq (1 - \frac{\eta(1-\mu)}{k})\Phi(\mathbf{s}^t)$ whenever $\mathbf{s}^t$ is a bad state. This, together with the fact that Φ can only decrease in each iteration, implies that after $\frac{k}{\eta(1-\mu)}$ bad states the potential function decreases by a constant factor. This implies the claimed bound of $O(\frac{k}{\eta(1-\mu)} \log \frac{\Phi(\mathbf{s}^0)}{\Phi_{min}})$ bad states in all. ■

References

[1] B. Awerbuch, Y. Azar, A. Epstein, V. S. Mirrkoni, and A. Skopalik. Fast convergence to nearly optimal solutions in potential games. In *Proceedings of the 9th ACM Conference on Electronic Commerce (EC)*, pages 264–273, 2008.

[2] S. Chien and A. Sinclair. Convergence to approximate Nash equilibria in congestion games. *Games and Economic Behavior*, 71(2):315–327, 2011.

[3] D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14(1):124–143, 1996.

[4] T. Roughgarden. Intrinsic robustness of the price of anarchy. In *41st ACM Symposium on Theory of Computing (STOC)*, pages 513–522, 2009.

[5] A. Skopalik and B. Vöcking. Inapproximability of pure Nash equilibria. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 355–364, 2008.