Stochastic Search Methods for Nash Equilibrium Approximation in Simulation-Based Games

Yevgeniy Vorobeychik and Michael P. Wellman
University of Michigan
Computer Science & Engineering
Ann Arbor, MI 48109-2121 USA
{yvorobey,wellman}@umich.edu

ABSTRACT

We define the class of games called *simulation-based games*, in which the payoffs are available as an output of an oracle (simulator), rather than specified analytically or using a payoff matrix. We then describe a convergent algorithm based on a hierarchical application of simulated annealing for estimating Nash equilibria in simulation-based games with finite-dimensional strategy sets. Additionally, we present alternative algorithms for best response and Nash equilibrium estimation, with a particular focus on one-shot infinite games of incomplete information. Our experimental results demonstrate that all the approaches we introduce are efficacious, albeit some more so than others. We show, for example, that while iterative best response dynamics has relatively weak convergence guarantees, it outperforms our convergent method experimentally. Additionally, we provide considerable evidence that a method based on random search outperforms gradient descent in our setting.

Categories and Subject Descriptors

I.2.8 [**Problem Solving, Control Methods, and Search**]: Heuristic methods; I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

General Terms

Economics, Experimentation

Keywords

Empirical game, approximate equilibria, heuristic search

1. INTRODUCTION

The field of Game Theory has enjoyed considerable success as a framework for modeling strategic interactions between agents. A plethora of theoretical game models and analysis techniques have been developed over the years [3], and a number of numerical solvers exist, for example, GAMBIT [7] and GameTracer [1]. To the best of our knowledge, however, few solution or approximation tools exist for any general class of *infinite* games (that is, games with infinite sets of strategies). Reeves and Wellman solver [10] is one such tool; however, it can only be applied to two-player games with a restricted class of piecewise linear utility functions. The dearth of general-purpose solvers or approximation tools for infinite games considerably restricts the space of strategic models that can be studied: a model must either be analytic, fall into a highly restricted class covered by available solvers, or be amenable to coarse discretization that does not significantly sacrifice solution quality.

In this paper, we introduce several general-purpose Nash equilibrium approximation techniques for infinite games, focusing especially on one-shot infinite games of incomplete information. All of our techniques rely on a best response approximation subroutine, as shown in Figure 1. Our goal is to take as input a black-box specification of the players' payoff functions (and type distributions, if we are dealing with a game of incomplete information) and output an approximate Nash equilibrium. In a fairly general setting, we are able to demonstrate theoretical convergence of one of our methods to an actual Nash equilibrium. Additionally, we experimentally demonstrate efficacy of all methods we study. Our experimental evidence focuses on relatively simple auction settings, with most examples involving only one-dimensional types, and, consequently, allowing the use of low-dimensional strategy spaces. It is as yet unclear how our approaches will fare on considerably more complex games, as such games are not easily amenable to an experimental evaluation.

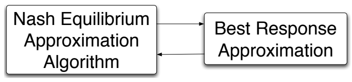

Figure 1: Approximation of Nash equilibria using a best response approximation subroutine.

2. NOTATION

A normal-form game is formally expressed by a tuple $[I, \{S_i\}, \{u_i(s)\}]$, where I refers to the set of players and $m = |I|$ is the number of players. S_i is the set of strategies available to player $i \in I$. The utility function, $u_i(s) : S_1 \times \cdots \times S_m \to \mathcal{R}$ defines the payoff of player i when players jointly play $s = (s_1, \ldots, s_m)$, where each player's strategy s_j is selected from his strategy set, S_j. It is often convenient to refer to the strategy (pure or mixed) of player i separately from that of the remaining players. To accommodate this, we use s_{-i} to denote the joint strategy of all players other than i.

Cite as: Stochastic Search Methods for Nash Equilibrium Approximation in Simulation-Based Games, Yevgeniy Vorobeychik and Michael P. Wellman, *Proc. of 7th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2008)*, Padgham, Parkes, Müller and Parsons (eds.), May, 12-16., 2008, Estoril, Portugal, pp. XXX-XXX.

Faced with a one-shot game, an agent would ideally play its best strategy given those played by the other agents. A configuration where all agents play strategies that are best responses to the others constitutes a *Nash equilibrium.*

DEFINITION 1. *A strategy profile* $s = (s_1, \ldots, s_m)$ *constitutes a* Nash equilibrium *of game* $[I, \{S_i\}, \{u_i(s)\}]$ *if for every* $i \in I$, $s'_i \in S_i$, $u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i})$.

An alternative way to look at the strategic landscape is to define a function which gives for each profile the maximum benefits any agent can gain from a unilateral deviation (*regret*): $\epsilon(s) = \max_{i \in I} \max_{a \in S_i} [u_i(a, s_{-i}) - u_i(s)] = \max_{i \in I} [u_i^*(s_{-i}) - u_i(s)]$.

We denote a *simulation-based game* by $[I, \{S_i\}, \mathcal{O}]$, where the oracle, $\mathcal{O}$, produces a (possibly noisy) sample from the joint payoff function of players, given a joint strategy profile. That is, $\mathcal{O}(s) = v$, where $v = (v_1, \ldots, v_m)$ and $E[v] = u(s)$. We call the normal form game $[I, \{S_i\}, \{u_i(s)\}]$ in this context the *underlying game.* As such, we will always evaluate $\epsilon(s)$ of a profile $s \in S$ with respect to the underlying game. Finally, we denote an estimate of a payoff for profile s, $u(s)$, based on k samples from $\mathcal{O}$ by $\hat{u}_k(s) = \frac{1}{k}\sum_{j=1}^{k} v(s)^j$, where each $v(s)^j$ is generated by invoking the oracle.

3. BEST RESPONSE APPROXIMATION

Best-response approximation is a subroutine in all the methods for equilibrium approximation we discuss below. Thus, we first describe this problem in some detail and present a globally convergent method for tackling it.

3.1 Continuous Stochastic Search for Black-Box Optimization

At the core of our algorithms for approximating best response lies a stochastic search subroutine which can find an approximate maximizer of a black-box objective function on continuous domains. The topic of one-stage black-box continuous optimization has been well-explored in the literature [11]. In this work, we utilize two algorithms: stochastic approximation and simulated annealing. The overall approach, of course, admits any satisfactory black-box optimization tool that can be effective in continuous settings. A part of our goal is to assess the relative difference between the performance of a local and a global search routine.

3.1.1 Stochastic Approximation

Stochastic approximation [11] is one of the early algorithms for continuous stochastic search. The idea of stochastic approximation is to implement gradient descent algorithms in the context of a noisy response function. As with all gradient descent algorithms, convergence is guaranteed only to a local optimum.[1] However, together with random restarts and other enhancements, stochastic approximation can perform reasonably well even in global optimization settings.

3.1.2 Simulated Annealing

Simulated annealing is a well-known black-box optimization routine [11] with provable global convergence [4]. Simulated annealing takes as input an oracle, f, that evaluates candidate solutions, a set X of feasible solutions, a candidate kernel $K(X_k, \cdot)$ which generates the next candidate solution given the current one, X_k, and the temperature schedule t_k that governs the Metropolis acceptance probability $p_k(f(x), f(y))$ at iteration k, which evaluates to $\exp[-\frac{f(y)-f(x)}{t_k}]$ when $f(y) < f(x)$ and 1 otherwise. It then follows a 3-step algorithm, iterating steps 2 and 3:

1. Start with $X_0 \in X$.
2. Generate Y_{k+1} using candidate kernel $K(X_k, \cdot)$.
3. Set $X_{k+1} = Y_{k+1}$ with Metropolis acceptance probability $p_k(f(X_k), f(Y_{k+1}))$, and $X_{k+1} = X_k$ otherwise.

3.2 Globally Convergent Algorithm for Best Response Approximation

We now present the application of simulated annealing search to the problem of best response in games in Algorithm 1, where t_k is a schedule of temperatures, and n_k is a schedule of the number of samples used to evaluate the candidate solutions at iteration k.

Algorithm 1 $\mathrm{BR}(\mathcal{O}, S_i, s_{-i}, K(\cdot,\cdot), t_k, n_k)$

1: Start with $a_0 \in S_i$
2: For $k > 0$, generate $b_{k+1} \in S_i$ using $K(s_k, \cdot)$
3: Generate $U_1 = \hat{u}_{n_k,i}(a_k, s_{-i})$ and $U_2 = \hat{u}_{n_k,i}(b_{k+1}, s_{-i})$ from $\mathcal{O}$
4: Set $a_{k+1} \leftarrow b_{k+1}$ w.p. $p_k(U_1, U_2)$ and $a_{k+1} \leftarrow a_k$ o.w.

For the analysis below, we need to formalize the notion of candidate Markov kernel, $K(\cdot)$, which describes a distribution over the next candidate given the current:

DEFINITION 2. *A function* $K : A \times \mathcal{B} \to [0, 1]$ *is a* candidate Markov kernel *if* $A \subset R^n$ *and* $\mathcal{B}$ *is a Borel* σ*-field over* A. *The first argument of* $K(\cdot, \cdot)$ *is the current candidate, and the second is a subset of candidates, for which* K *gives the probability measure.*

In order for simulated annealing to have any chance to converge, the kernel must satisfy several properties, in which case we refer to it as an *admissible kernel.*

DEFINITION 3. *A kernel* $K : A \times \mathcal{B} \to [0, 1]$ *is* admissible *if (a)* K *is absolutely continuous in second argument, (b)* $K(x, B) = \int_B f(x, y)dy$ *with* $\inf_{x,y \in A} f(x, y) > 0$, *and (c) For every open* $B \subset A$, $K(x, B)$ *is continuous in* x.

The following conditions map directly to the sufficient conditions for global convergence of simulated annealing observed by Ghate and Smith [4]:

1. *EXISTENCE* holds if S_i is closed and bounded and the payoff function $u_i(s_i, s_{-i})$ is continuous on S_i. This condition is so named because it implies that the best response exists by the Weierstrass theorem.
2. *ACCESSIBILITY* holds if for every maximal $a^* \in S_i$ and for any $\epsilon > 0$, the set $\{a \in S_i : \|a - a^*\| < \epsilon\}$ has positive Lebesgue measure
3. *DECREASING TEMPERATURES (DT)* holds if the sequence t_k of temperaturesconverges to 0

[1]For example, Kiefer and Wolfowitz [5] demonstrated convergence of this technique when gradient is estimated via a *finite difference* method—that is, based on the difference of function values at neighborhood points.

4. *CONVERGENCE OF RELATIVE ERRORS (CRE)* holds if the sequences $|\tilde{u}_{n_k,i}(a_k, s_{-i}) - u_i(a_k, s_{-i})|/t_k$ and $|\tilde{u}_{n_k,i}(b_{k+1}, s_{-i}) - u_i(b_{k+1}, s_{-i})|/t_k$, where b_{k+1} is the next candidate generated by the kernel, converge to 0 in probability

The first two conditions ensures that the global optimum actually exists and can be reached by random search with positive probability. The third and fourth conditions ensure that the iterates stabilize around optima, but do so slowly enough so that the noise does not lead the search to stabilize in suboptimal neighborhoods.

THEOREM 1 ([4]). *If the problem satisfies*EXISTENCE *and* ACCESSIBILITY, *and the algorithm paramters satisfy* DT, *and* CRE, *Algorithm 1 which uses an admissible candidate kernel converges in probability to* $u_i^*(s_{-i}) = \max_{a \in S_i} u_i(a, s_{-i})$.

Let $\hat{u}_{i,k}(s_{-i})$ to be the answer produced when Algorithm 1 is run for k iterations. This will be our estimate of $u_i^*(s_{-i})$, which, by Theorem 1, is consistent.

4. APPROXIMATING NASH EQUILIBRIA

Our goal in this paper is to take a simulation-based game as an input and return a profile constituting an approximate Nash equilibrium in the underyling game. Below, we present two general approaches to Nash equilibrium approximation: a well-known iterative best response approach, as well as our own algorithm based on simulated annealing. Both of these can use *any* best response approximation algorithm as a subroutine.

4.1 Equilibrium Approximation via Iterated Best Response

While the problem of best-response approximation is interesting in its own right, it may also be used iteratively to approximate a Nash equilibrium. For ease of exposition, we describe the procedure used for symmetric profiles (i.e., profiles in which all players play the same strategy):

1. Generate an initial symmetric profile s_0
2. Find approximate best response, $\hat{s}$, to current profile s_k
3. Set $s_{k+1} = \hat{s}$ and go back to step 2

When the procedure terminates after a finite number of steps K, we return the final iterate s_K as an approximate Nash equilibrium. Under very restrictive assumptions (e.g., in supermodular games with unique Nash equilibria [8] and in congestion games [9]) iterated best response is known to converge to a Nash equilibrium.

4.2 A Globally Convergent Algorithm for Equilibrium Approximation

In this section we are interested in developing a globally convergent algorithm for finding approximate Nash equilibria. The approach we take, visualized in Figure 2, is to minimize approximate regret, $\hat{\epsilon}(s)$, where approximations are produced by running Algorithm 1. For the task of minimizing regret we again use an adaptation of simulated annealing, but now need to establish the convergence conditions for this meta-problem.

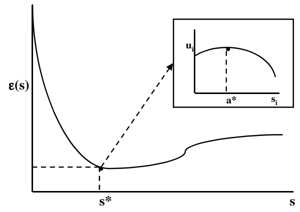

Figure 2: A diagramatic view of our algorithm based on approximate regret minimization.

First, let us define a candidate kernel for this problem as a combination of admissible kernels for each agent i:

$$K(x, B) = \int_B \prod_{i \in I} f_i(x, y_i) \prod_{i \in I} dy_i, \tag{1}$$

where $K^i(x, C) = \int_C f_i(x, y) dy$ with $f_i(\cdot)$ the Kernel density used by the simulated annealing routine for player i. We now confirm that the resulting kernel is admissible.

LEMMA 2. *The candidate kernel defined in Equation 1 is admissible.*

PROOF. Since each $f_i(x, y_i)$ is positive everywhere, so is the product. Furthermore, it is clear that if B is of measure-zero, then so is $K(x, B)$. Finally, $K(x, B)$ is continuous on x since each $f_i(x, y_i)$ is continuous and therefore so is the product. □

Thus, we need only define admissible kernels on each player's strategy set.

Given the candidate kernels for each player and the constructed candidate kernel for regret minimization, we present a meta-algorithm—Algorithm 2—to find approximate Nash equilibria.

Algorithm 2 EQEst$(\mathcal{O}, S, K(\cdot,\cdot, K^i(\cdot,\cdot), t_l, n_l, t_k^i, n_k^i)$

1: Start with $s_0 \in S$
2: Generate $q_{l+1} \in S$ using $K(s_l, \cdot)$
3: Generate $\hat{u}_{n_l,i}(s_l)$ and $\hat{u}_{n_l,i}(q_{l+1})$ from $\mathcal{O}$
4: Let $\hat{u}_{i,l}(s_{-i,l}) \leftarrow BR(\mathcal{O}, s_{-i,l}, S_i, K^i(\cdot,\cdot), t_k^i, n_k^i)$
5: Let $\hat{u}_{i,l}(q_{-i,l+1}) \leftarrow BR(\mathcal{O}, q_{-i,l}, S_i, K^i(\cdot,\cdot), t_k^i, n_k^i)$
6: Compute regrets $\hat{\epsilon}(s_l)$ and $\hat{\epsilon}(q_{l+1})$
7: Set $s_{l+1} \leftarrow q_{l+1}$ w.p. $p_l(s_l, q_{l+1})$ and $s_{l+1} \leftarrow s_l$ o.w.

We now present the sufficient conditions for convergence of Algorithm 2. First, we verify what we need for continuity of $\epsilon(s)$ in the following Lemma.

LEMMA 3. *If* $u_i(s)$ *are uniformly continuous on* S *for every* i, *then* $\epsilon(s)$ *is continuous on* S.

The proof of this and other results is in the Appendix of the extended version.

Based on Lemma 3, we need to modify the *EXISTENCE* criterion slightly as follows:

*EXISTENCE** holds if S_i is closed and bounded and the payoff function $u_i(s_i, s_{-i})$ is uniformly continuous on S_i for every player i.

Since we are concerned about every player now and, furthermore, need to avoid "undetectible" minima in $\epsilon(s)$, we also modify the *ACCESSIBILITY* condition:

*ACCESSIBILITY** holds if for any $\delta > 0$, for every profile s, for every player i, and for every maximal $a^* \in S_i$ the set $\{a \in S_i : \|a - a^*\| < \delta\}$ has positive Lebesgue measure; furthermore for every minimal $s^* \in S$ the set $\{s \in S : \|s - s^*\| < \delta\}$ has positive Lebesgue measure

We also need to augment the conditions on algorithm parameters to include both the conditions on the parameters for the problem of minimizing $\epsilon(s)$, as well as the conditions on parameters for finding each player's best response. For clarity, we will let l denote the iteration number of the meta-problem of minimizing $\epsilon(s)$ and k denote the iteration number of the best response subroutine.

DECREASING TEMPERATURES (DT*)* holds if for every agent i the sequence t_k^i of temperatures converges to 0, and the sequence t_l of temperatures converges to 0

CONVERGENCE OF RELATIVE PAYOFF ERRORS (CRPE) holds if for every agent i the sequences of ratios $|\hat{u}_{n_k^i,i}(a_k, s_{-i}) - u_i(a_k, s_{-i})|/t_k^i$ and the sequence of ratios $|\hat{u}_{n_k^i,i}(b_{k+1}, s_{-i}) - u_i(b_{k+1}, s_{-i})|/t_k^i$, where b_{k+1} is the next candidate generated by the kernel, converge to 0 in probability.

Now, define $\hat{\epsilon}_l(s) = \max_{i \in I}[\hat{u}_{i,n_l}(s_{-i}) - \hat{u}_{i,n_l}(s)]$.

LEMMA 4. *If* EXISTENCE*, ACCESSIBILITY*, DT*, *and* CRPE *hold, $\hat{\epsilon}_l(s)$ converges to $\epsilon(s)$ in probability for every $s \in S$.*

We need one more condition on the algorithm parameters before proving convergence:

CONVERGENCE OF RELATIVE EPSILON ERRORS (CREE) holds if the sequences of ratios $|\hat{\epsilon}_{l,i}(s_k, s_{-i,k}) - \epsilon_i(s_k, s_{-i,k})|/t_k$ and $|\hat{\epsilon}_{n_k,i}(r_{k+1}, s_{-i,k}) - \epsilon_i(r_{k+1}, s_{-i,k})|/t_k$, where r_{k+1} is the next candidate generated by the kernel, converge to 0 in probability.

THEOREM 5. *Under the conditions* EXISTENCE*, ACCESSIBILITY*, DT*, CRPE, *and* CREE, *Algorithm 2 converges to $\bar{\epsilon} = \min_{s \in S} \epsilon(s)$.*

PROOF. While Ghate and Smith [4] prove convergence for functions which are expectations of the noisy realizations, their proof goes through unchanged under the above sufficient conditions, as long as we ascertain that $\hat{\epsilon}_l \to \epsilon(s)$ for every $s \in S$. This we showed in Lemma 4. □

COROLLARY 6. *If there exists a Nash equilibrium on S, Algorithm 2 converges to a Nash equilibrium when the conditions* EXISTENCE*, ACCESSIBILITY*, DT*, CRPE, *and* CREE *obtain.*

5. INFINITE GAMES OF INCOMPLETE INFORMATION

Perhaps the most important application of the methods we have discussed is to infinite games of incomplete information. In what follows, we define one-shot games of incomplete information and adapt our methods to this domain. Additionally, we introduce another best response approximation method specifically designed for strategies that are functions of private information.

5.1 Definitions and Notation

We denote *one-shot games of incomplete information* by $[I, \{A_i\}, \{T_i\}, F(\cdot), \{u_i(r, t)\}]$, where I refers to the set of players and $m = |I|$ is the number of players. A_i is the set of actions available to player $i \in I$, and $A_1, \cdots, A_m$ is the joint action space. T_i is the set of types (private information) of player i, with $T = T_1 \times \cdots \times T_m$ representing the joint type space. A one-shot game of incomplete information is said to be *infinite* if both A and T are infinite. Since we presume that a player knows his type prior to taking an action, but does not know types of others, we allow him to condition his action on own type. Thus, we define a strategy of a player i to be a function $s_i : T_i \to A_i$, and use $s(t)$ to denote the vector $(s_1(t_1), \ldots, s_m(t_m))$. $F(\cdot)$ is the distribution over the joint type space. We define the payoff (utility) function of each player i by $u_i : A \times T \to \mathbb{R}$, where $u_i(a_i, a_{-i}, t_i, t_{-i})$ indicates the payoff to player i with type t_i for playing action $a_i \in A_i$ when the remaining players with joint types t_{-i} play r_{-i}. Given a strategy profile $s \in S$, the expected payoff of player i is $\tilde{u}_i(s) = E_t[u_i(s(t), t)]$.

Given a known strategy profile of players other than i, we define the best response of player i to s_{-i} to be the strategy s_i^* that maximizes expected utility $\tilde{u}_i(s_i, s_{-i})$. A configuration where all agents play best responses to each other in such a setting constitutes a *Bayes-Nash equilibrium.*

Since we defined the normal form games in terms of strategy sets which are subsets of $\mathbb{R}^n$, we cannot represent games of incomplete information perfectly in our restricted normal form. What we can do, however, is restrict the sets of strategies allowed for each player to a finite-dimensional function space on reals, and thereby parametrize each strategy using a vector $\theta_i \in \Theta_i \subset \mathbb{R}^n$. Let $\Theta = \Theta_1 \times \cdots \times \Theta_m$. Let us denote this restricted space $\mathcal{H}_i$ for each player i. Then, $h_{\theta_i,i}(t) \in \mathcal{H}_i$ is a particular type-conditional strategy of player i. We aggregate over all players to obtain $h_\theta(t) = (h_{\theta_1,1}, \ldots, h_{\theta_m,m})$. We then describe a restricted game of incomplete information by $[I, \{\mathcal{H}_i\}, \{T_i\}, F(\cdot), \{u_i(s)\}]$, where T_i is the set of player i's types and $F(\cdot)$ is the joint distribution over player types. We can map this game into the normal form as described previously by letting $S_i = \Theta_i$, the set of parametrizations, and for any $\theta \in \Theta$, $\tilde{u}_i(\theta) = E_F u_i(h_\theta(t))$. Thus, the transformed game is $[I, \{\Theta_i\}, \{\tilde{u}_i(\theta)\}]$. Now, Algorithm 2 is directly applicable and will guarantee convergence to a strategy profile with the smallest expected benefit for a unilateral deviation to any player.

5.2 Best Response Approximation

Given a best response subroutine, Algorithm 2 can be applied to the infinite games of incomplete information, although we guarantee convergence only when Algorithm 1 comprises this subroutine. Below, we describe two methods for approximating best response functions: the first is a direct adaptation of the techniques we described above; the second is based on regression. We note that both methods rely on an assumption that we can define a relatively low-dimensional hypothesis class for each player which contains good approximations of the actual best response. Later, we experimentally verify that this is indeed possible for a number of interesting and non-trivial games. More generally, an analyst may need to hand-craft low-dimensional restricted strategy sets in order to effectively apply our techniques.

Direct Method.

Our first method for approximating best response functions in infinite games is simply an application of Algorithm 1. Here, the oracle $\mathcal{O}$ performs two steps: first, generate a type $t \in T$ from the black-box type distribution;

and next, generate a payoff from the simulation-based payoff function for the strategy profile *evaluated* at t. As we have noted above, we can guarantee convergence to global best response function in the finite-dimensional hypothesis class $\mathcal{H}_i$; indeed convergence obtains even for an arbitrary black-box specification of the strategies of other players.

Regression to Pointwise Best Response.
Our second method takes an indirect route to approximating the best response, approximating best response *actions* for each of a subset of player types, and thereafter fitting a regression to these. The outline of this algorithm is as follows:

1. Draw L types, $\{t_1, \dots, t_L\}$, from the black-box type distribution
2. Use simulated annealing to approximate a pointwise best response for each t_j, $\hat{s}_j$
3. Fit a regression $\hat{s}(t)$ to the data set of points $\{t_j, \hat{s}_j\}$

The regression $\hat{s}(t)$ is the resulting approximation of the best response function.

6. EXPERIMENTAL EVALUATION OF BEST RESPONSE QUALITY

6.1 Experimental Setup

In this section, we explore the effectiveness of the two methods we introduced as best-response routines for infinite one-shot games of incomplete information. The best response methods were both allowed 5000 queries to the payoff function oracle for each iteration, and a total of 150 iterations. For both, the total running time was consistenty under 4 seconds. We compared our methods using both stochastic approximation (indicated by "_stochApprox_" in our plots) and the simulated annealing stochastic search subroutines. Besides comparing the methods to each other, we include as a reference the results of randomly selecting the slope parameter of the linear best response. We want to emphasize that our goal is *not* merely to beat the random method, but to use it as *calibration* for the approximation quality of the other two.

We test our methods on three infinite auction games. The first is the famed Vickrey, or second-price sealed-bid, auction [6]. The second is first-price sealed-bid auction [6]. The final game to which we apply our techniques is a *shared-good auction*, with payoff function specified in Equation 2.

$$u_i(a_i, a_{-i}, t_i, t_{-i}) = \begin{cases} \frac{1}{v}(t_i - \frac{m-1}{m} a_i + \frac{1}{m} \max_{j \neq i} a_j) \\ \qquad\qquad\qquad \text{if } a_i = \max_{j \neq i} a_j, \\ t_i - \frac{m-1}{m} a_i \qquad \text{if } a_i > \max_{j \neq i} a_j, \\ \frac{1}{m} \max_{j \neq i} a_j \qquad \text{otherwise.} \end{cases} \quad (2)$$

We experimented with two- and five-player games with uniform type distributions, noting that the best-response finder proposed by Reeves and Wellman [10] cannot be directly applied to games with more than two players. In our experiments with these auctions, we focus on the hypothesis class $\mathcal{H}$ of linear functions, with $\theta = (\alpha, \beta)$ such that $h_\theta(t_i) = \alpha t_i + \beta$, which includes an actual best response in many auction domains. In all best-response experiments, we sought a best response to a linear strategy of the form $s(t) = kt$, with k generated uniformly randomly in $[0, 1]$. The results were evaluated based on regret, $\epsilon(s)$, computed on the underlying game. We took the average value of $\epsilon(s)$ over 100–1000 trials and based our statistical tests on these samples. Statistically significant difference at the 99% confidence level between our two methods when both use simulated annealing was indicated by a "*". In our plots, we also include 99% confidence intervals for further statistical comparisons.

6.2 Two-Player One-Item Auctions

Our first experiment is largely a sanity check, as there is an *exact* best response finder for all three auction games we consider here [10].[2] Our results are shown in Figure 3. We

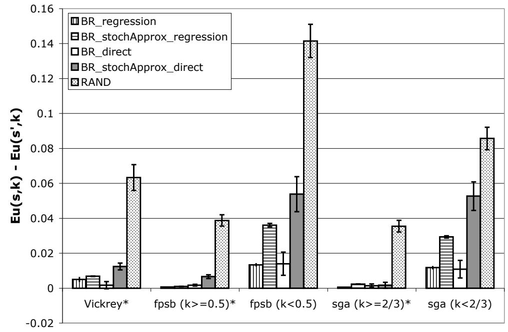

Figure 3: Comparison of best response methods in 2-player games with reserve price = 0.

group the initial results into two categories. The first category is comprised of settings in which there is a linear best response function. This is satisfied by Vickrey for any value of k, first-price sealed-bid auction (fpsb) with $k \geq 1/2$, and shared-good auction (sga) with $k \geq 2/3$.[3] In all of these settings, our best response approximations are orders of magnitude better than random. Indeed, *in every auction we study, the difference between all of our methods (using stochastic approximation or simulated annealing) and random is quite statistically significant (p-value $< 10^{-10}$).* Therefore, we omit the results for random from the subsequent figures. Additionally, in all but Vickrey, the regression-based method is better than direct, most likely because this method is particularly sample-efficient when the actual best response is linear and there are not many alternative best response options.

The settings in the second category yield non-linear best response functions. All the remaining comparisons in Figure 3 fall into this category. As expected, the performance of linear best response approximation is somewhat worse here, although in all cases far better than random. It is worth noting that in all of these the direct method performs statistically no worse, and in several cases much better than the regression-based method. The result is intuitive, since the direct method seeks the most profitable linear best response, whereas regression presumes that linearity is a good

[2] There is not a significant difference in running times between the exact best response finder and our approximators.
[3] Recall that k is the randomly generated slope of the line to which we are approximating a best response function.

fit for the *actual* best response, and may not do well when this assumption does not hold.

While it is good to see the effectiveness of our methods in settings which we can already solve, our goal is to apply them to problems in a class for which no general-purpose numerical solver exists. Our first such examples are two-player Vickrey and first-price sealed-bid auctions with reserve prices, denoted by vickrey_rp and fpsb_rp respectively (Figure 4).[4] In both of these, the direct method far outperforms the regression-based method. In nearly all cases simulated annealing yielded statistically significant improvement over stochastic approximation; indeed, at times it was better by more than a factor of magnitude. The reason, we believe, is that stochastic approximation expends much of its computing budget estimating gradients, while simulated annealing can guide and make use of all the function evaluations it generates along the search path. Additionally, stochastic approximation is a local search method (even with random restarts, which considerably enhance its performance), and our problems appear more suited to global search.

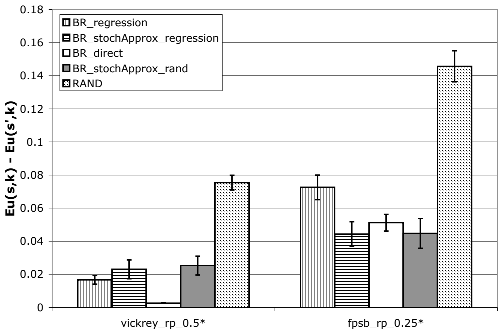

Figure 4: Comparison of best response methods in 2-player games with reserve price > 0. The reserve price for Vickrey is 0.5, while it is 0.25 for the first-price sealed-bid auction.

6.3 Five-Player One-Item Auctions

Since there is no general-purpose best response finder for five-player infinite games of incomplete information, the only viable comparison of our results is to each other. As we can see from Figure 5, in the five-player setting the direct method tends to produce substantially better approximate best response than the regression-based method. Additionally, in two of the three auctions in this setting, simulated annealing showed substantial advantage over stochastic approximation. Since these results echo those in smaller games, the reasons are likely to be the same.

6.4 Sampling and Iteration Efficiency

In this section, we compare the "regression" and "direct" methods in terms of efficiency in their use of both samples from the payoff function and iterations of the optimization algorithm. Our results below are roughly representative of the entire set of results involving several auction games with varying numbers of players.

First, we consider sampling efficiency. As Figure 6 suggests, the "direct" method seems no worse and at times substantially better than the "regression"-based method for various sample sizes we consider: when very few samples are taken, both methods seem to perform almost equally poorly, but as we increase the number of samples per iteration, "direct" method quickly surpasses "regression". Iteration efficiency results are presented in Figure 7. Interestingly, these results appear somewhat different from those for sampling efficiency: the "direct" method seems particularly affected by a dearth of iterations, while "regression" is quite robust. Note, however, that even in the case of sampling efficiency, regression is quite robust across different sample sizes; its flaw is that it fails to take sufficient advantage of additional sampling. We conjecture that the robustness of regression is partly because linear approximation of actual best response is somewhat reasonable in this setting, and, if so, regression smoothes out the noise much better when iterations are few.

[4] Of course, both of these are analytically tractable. Our study, however, is solely concerned with *numerical* methods for solving games.

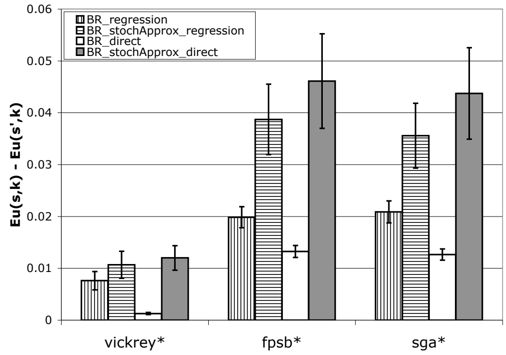

Figure 5: Comparison of best response methods in 5-player games with uniform type distribution.

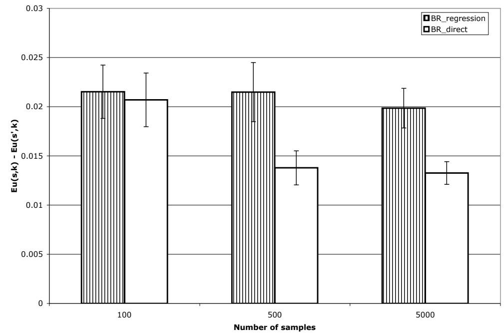

Figure 6: Comparison of sampling efficiency of best response methods in two-player first-price sealed-bid auctions.

7. EXPERIMENTAL EVALUATION OF EQUILIBRIUM QUALITY

7.1 Experimental Setup

We now turn to an application of best response techniques to Bayes-Nash equilibrium approximation in infinite one-shot games of incomplete information. One potential application is to extend the Automated Mechanism Design framework introduced by Vorobeychik et al. [12] beyond linear two-player games with piecewise-uniform type distri-

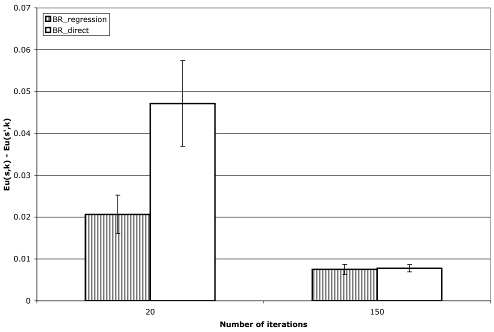

Figure 7: Comparison of iteration efficiency of best response methods in two-player first-price sealed-bid auctions.

butions, a possibility heretofore precluded due to lack of general-purpose numerical solvers.

Since the games we consider are all symmetric, we will focus on approximating *symmetric* equilibria, that is, equilibria in which all agents adopt the same strategy. Assuming that we can compute a best response to a particular strategy, we can use iterated best response dynamics to find equilibria, assuming, of course, that our dynamics converge. Here, we will avoid the issue of convergence by taking the last result of five best response iterations as the final approximation of the Bayes-Nash equilibrium. In all cases, we seed the iterative best response algorithm with truthful bidding, i.e., $a(t) = t$. All other elements of experimental setup are identical to the previous section.

As before, in every application the difference between both our methods and random is quite statistically significant and the actual experimental difference is always several orders of magnitude.

7.2 Two-Player One-Item Auctions

We first consider three two-player games for which we can numerically find the exact best response and two for which we cannot. We present the results in Figure 8. In all three

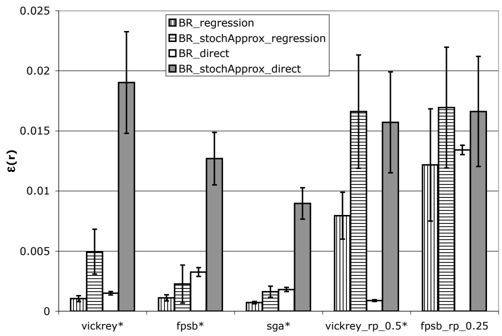

Figure 8: Comparison of equilibrium approximation quality in two-player games.

games that can be solved exactly (vickrey, fpsb, and sga), the regression-based method outperforms the direct method. In the case of Vickrey auction with reserve price of 0.5, this result is reversed, and the perfomance of the two methods on first-price sealed-bid auction with reserve price of 0.25 is not statistically different. As we have observed previously, simulated annealing tends to be considerably better than stochastic approximation.

7.3 Five-Player One-Item Auctions

Now we consider five-player games, for which no general-purpose numerical tool exists to compute a Bayes-Nash equilibrium or even a best response. Our results are presented in

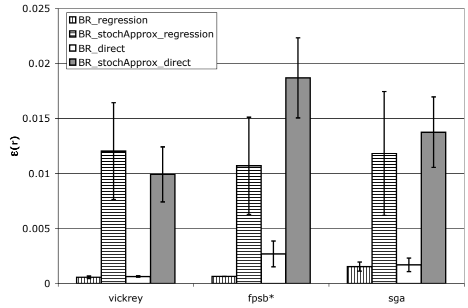

Figure 9: Comparison of equilibrium approximation quality in five-player games with uniform type distribution.

Figure 9. While the two methods are statistically indistinguishable from one another, the direct method was considerably better than the regression-based method in Vickrey auction experiments, and the opposite was true in the experiments involving the first-price sealed-bid auction. Interestingly, we can observe here considerable advantage from using simulated annealing, as compared to stochastic approximation, in all instances for both the regression-based and the direct methods.

7.4 Two-Player Two-Item First-Price Combinatorial Auction

In all our experiments above we faced auctions with one-dimensional player types. Here, we apply our methods to a first-price combinatorial auction—a considerably more complex domain—although we restrict the auction to two players and two items. We allocate the items between the two bidders according to the prescription of *winner determination problem* [2], which takes a particularly simple form in this case. We further restrict our problem to bidders with complementary valuations. Specifically, each bidder draws a value v^i for each item i from a uniform distribution on [0,1] and draws the value for the bundle of both items v^b from the uniform distribution on $[v^1 + v^2, 2]$. We let each player's value vector be denoted by $v = \{v^1, v^2, v^b\}$.

Since the game is symmetric, we seek a symmetric approximate equilibrium. Since the joint strategy space is an intractable function space, we restrict the hypothesis class to the functions of the form:

$$b^1(v) = k_1 v^1; b^2(v) = k_2 v^2; b^b(v) = b^1 + b^2 + k_3(v^b - b^1 - b^2).$$

Unlike the experiments above, verifying the approximation quality with respect to actual best responses is extremely difficult in this case. Thus, we instead measure the quality of our approximations against the best possible in our restricted strategy space. Finally, we use here the direct method with simulated annealing as the best response approximation tool. As we can see from Figure 10, the best response dynamics appears to converge quite quickly on the

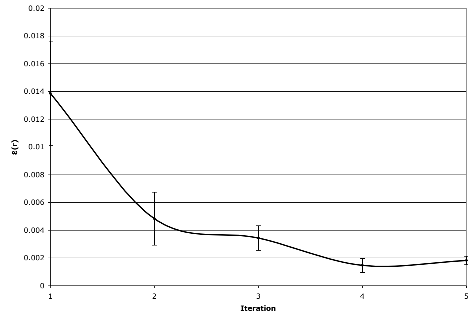

Figure 10: Iterations of approximate best response dynamics in the combinatorial auction.

restricted strategy space. Thus, at least according to this limited criterion, our direct method is quite effective in approximating equilibria even in this somewhat complicated case for which no general analytic solution is known to date.

8. COMPARISON OF EQUILIBRIUM APPROXIMATION METHODS

In this section we compare the approximation quality of best response dynamics and Algorithm 2—using simulated annealing in one case, and stochastic approximation in another.[5] In this setup, both best response dynamics and Algorithm 2 use the *direct* approximate best response method as a subroutine, and we only look at the five-player first-price sealed-bid auction (although we do not expect the results to be very different for the other auction domains above). As we can see from Figure 11, while not guaranteed to converge in general, best response dynamics seems more effective than Algorithm 2; thus, while convergence is guaranteed, it appears somewhat slow. Additionally, we observe that simulated annealing is significantly better than stochastic approximation even in the capacity of stochastic regret minimization.

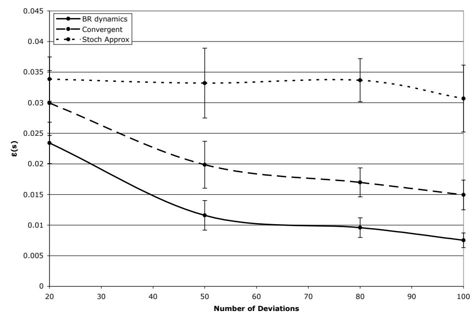

Figure 11: Comparison of equilibrium approximation quality of best response dynamics and the convergent method.

9. CONCLUSION

We study Nash equilibrium approximation techniques for games that are specified using simulations. Our algorithmic contributions include a set of methods, including a convergent algorithm, for best response and Nash equilibrium approximation. On the experimental side, we demonstrate that all methods that we introduce can effectively be used to approximate best response and Nash equilibria. However, there is considerable evidence in favor of using simulated annealing rather than a gradient descent-based algorithm as a black-box optimization workhorse. Of the two methods for approximating best response in games of incomplete information, we found that the method which directly optimized the parameters of the best response function outperformed a regression-based method on the more difficult problems, and was generally not very much inferior on others. Thus, faced with a new problem, the direct method seems preferable. There is a caveat, however: the regression-based method appeared more robust when the number of iterations is not very large. Our final result shows that, in spite of weak convergence guarantees, best response dynamics outperforms our globally convergent algorithm in the first-price sealed-bid auction setting.

While our results are generally very optimistic, our experimental work was restricted to relatively simple games. To be applicable to more difficult problems, particularly those with high-dimensional strategy sets, they will likely require the analyst to hand-craft restricted strategy sets given some knowledge of the problem structure.

[5]Naturally, when stochastic approximation is used, we lose the global convergence properties.

Acknowledgments

This work was supported in part by grant IIS-0414710 from the National Science Foundation.

10. REFERENCES

[1] B. Blum, C. R. Shelton, and D. Koller. A continuation method for Nash equilibria in structured games. *Journal of Artificial Intelligence Research*, 25:457–502, 2006.

[2] P. Cramton, Y. Shoham, and R. Steinberg, editors. *Combinatorial Auctions*. The MIT Press, 2006.

[3] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1991.

[4] A. Ghate and R. L. Smith. Adaptive search with stochastic acceptance probabilities for global optimization. draft, 2006.

[5] J. Kiefer and J. Wolfowitz. Stochastic estimation of the maximum of a regression function. *Annals of Mathematical Statistics*, 23(3):462–466, 1952.

[6] V. Krishna. *Auction Theory*. Academic Press, 1st edition, 2002.

[7] R. D. McKelvey, A. M. McLennan, and T. L. Turocy. Gambit: Software tools for game theory, version 0.2005.06.13, 2005.

[8] P. Milgrom and J. Roberts. Rationalizability, learning, and equilibrium in games with strategic complementarities. *Econometrica*, 58(6):1255–1277, 1990.

[9] D. Monderer and L. S. Shapley. Potential games. *Games and Economics Behavior*, 14:124–143, 1996.

[10] D. M. Reeves and M. P. Wellman. Computing best-response strategies in infinite games of incomplete information. In *Twentieth Conference on Uncertainty in Artificial Intelligence*, pages 470–478, 2004.

[11] J. C. Spall. *Introduction to Stochastic Search and Optimization*. John Wiley and Sons, Inc, 2003.

[12] Y. Vorobeychik and D. M. Reeves. Constrained automated mechanism design for infinite games of incomplete information. In *The Twenty Third Conference on Uncertainty in Artificial Intelligence*, 2007.