Maximum Entropy Correlated Equilibria

Luis E. Ortiz
ECE Deptartment
Univ. of Puerto Rico at Mayagüez
Mayagüez, PR 00681
leortiz@ece.uprm.edu

Robert E. Schapire
Dept. of Computer Science
Princeton University
Princeton, NJ 08540
schapire@cs.princeton.edu

Sham M. Kakade
Toyota Technological Institute
Chicago, IL 60637
sham@tti-c.org

Abstract

We study *maximum entropy correlated equilibria (Maxent CE)* in multi-player games. After motivating and deriving some interesting important properties of Maxent CE, we provide two gradient-based algorithms that are guaranteed to converge to it. The proposed algorithms have strong connections to algorithms for statistical estimation (e.g., iterative scaling), and permit a distributed learning-dynamics interpretation. We also briefly discuss possible connections of this work, and more generally of the Maximum Entropy Principle in statistics, to the work on learning in games and the problem of equilibrium selection.

1 INTRODUCTION

The Internet has been a significant source of new problems of great scientific interest. One example is understanding the largely decentralized mechanisms leading to the Internet's own formation. Another example is to study the outcomes of the interaction of many individual human and artificial agents resulting from the modern mechanisms that the Internet has facilitated during the last ten years (e.g., online auction). Such problems have played a major role in the increased interest in artificial intelligence to the study of multi-agent systems, as evidenced by the large body of recent work.

Game theory and mathematical economics have become the most popular framework in which to formulate and formally study multi-agent systems in artificial intelligence. This has led the computer science community to become very active in pursuing fundamental computational questions in game theory and economics. Computational game theory and economics has thus emerged as a very important area of study.

An equilibrium is perhaps the most important solution concept in game theory and in mathematical economics, where game theory is applied to model competitive economies [Arrow and Debreu, 1954]. An equilibrium is a point of strategic stance between the players in a game that characterizes and serves as a prediction to a possible outcome of the game.

The general problem of equilibrium computation is fundamental in computer science [Papadimitriou, 2001]. In the last decade, there have been *many* great advances in our understanding of the general prospects and limitations of computing equilibria in games. Just late last year, there was a breakthrough on what was considered a major open problem in the field: computing equilibria was shown to be hard, even for two-player games [Chen and Deng, 2005b,a, Daskalakis and Papadimitriou, 2005, Daskalakis et al., 2005, Goldberg and Papadimitriou, 2005]. However, there are still important open questions. The work presented here contributes in part to the problem of computing equilibria with *specific* properties.

One important notion of equilibrium in non-cooperative games of recent interest in computational game theory is that of correlated equilibria (CE), a concept due to Aumann [1974]. A CE generalizes the more widely-known equilibrium notion due to Nash [1951]. It also offers some interesting advantages over Nash equilibria (NE); among them, (a) it allows a weak form of "cooperation" that can lead to "better" and "fairer" outcomes while maintaining individual player rationality (i.e., each individual still plays a best-response strategy), (b) it is consistent with Bayesian rationality [Aumann, 1987], and (c) unlike NE, reasonably natural learning dynamics in repeated games converge to CE (see, for example, Foster and Vohra [1997], Foster and Vohra [1999], Hart and Mas-Colell [2000]). The textbook example of a CE is the traffic light at a road intersection: a correlating device for which the best response of the individual drivers approaching the intersection is to obey the signal, therefore leading to a better overall outcome not generally achievable by any NE (e.g., avoids accidents without forcing any unnecessary, discriminatory delays).

In this paper, we study *maximum entropy correlated equi-*

libria (Maxent CE). We argue that the Maximum Entropy Principle (Maxent), due to Jaynes [1957], can serve as a guiding principle for selecting among the many possible CE of a game. This is in part because, as we show, Maxent CE affords individual players additional guarantees over arbitrary CE and exhibits other representational and computationally attractive properties. While in many cases CE has been shown more computationally amenable than NE, computing CE with specific properties has been found to be hard in general.

We provide simple gradient-based algorithms for computing Maxent CE that closely resemble statistical inference algorithms used successfully in practice (see, for example, Berger et al. [1996] and Della Pietra et al. [1997]). The algorithms have a "learning" interpretation. In particular, they can be seen as natural update rules which individual players use during some form of pre-play negotiation. The update rules can also be implemented efficiently. We relate this work to that on learning dynamics and discuss other possible connections to the general literature on learning in games. More specifically, we comment on how Maxent may help us characterize, just as it does for many natural systems (e.g., in thermodynamics), the equilibrium behavior that is likely to arise from the dynamics of rational individuals, each following his or her own individually controlled learning mechanism.

2 PRELIMINARIES

In this section, we introduce some basic game-theoretic concepts, terminology and notation. (For a thorough introduction to game theory, see Fudenberg and Tirole [1991], for instance.)

Let n be the number of players in the game. For each player $i = 1, \ldots, n$, let $\mathcal{A}_i$ be its finite set of *actions* (also called *pure strategies*). Let $\mathcal{A} \equiv \times_{i=1}^{n} \mathcal{A}_i$ be the *joint-action space*; each element $a \equiv (a_1, \ldots, a_n) \in \mathcal{A}$ is a called a *joint action*, i.e., if a_i is the ith component of a, player i plays a_i in a. Let $\mathcal{A}_{-i} \equiv \times_{j=1, j\neq i}^{n} \mathcal{A}_j$ denote the joint action space of all the players *except* that of i. Similarly, given $a \in \mathcal{A}$, let $a_{-i} \in \mathcal{A}_{-i}$ denote the joint action, in a, of all the players *except* that of i. It is often convenient to denote the joint action a by (a_i, a_{-i}) to highlight the action of player i in a.

Let $M_i : \mathcal{A}^n \to \mathbb{R}$ be the *payoff function* of player i, i.e., if players play joint action $a \in \mathcal{A}$, then each player i individually receives a payoff value of $M_i(a)$. Let $\mathcal{M} \equiv \{M_1, \ldots, M_n\}$ be the set of payoff functions, one for each player.

An *n-player game in normal form* is defined by the tuple $\mathcal{G} \equiv (\mathcal{A}, \mathcal{M})$. [1]

[1]Such games are also often referred to as *strategic-form* or *matrix* games because in the finite-action case one can view each payoff function as an n-dimensional matrix indexed by the joint-actions.

Each player's objective is to maximize their *own (expected) payoff*. In the context of a game, a probability distribution $\mathbf{P}$ over $\mathcal{A}^n$ is called a *joint mixed strategy*, i.e., a randomized strategy where players play $a \in \mathcal{A}^n$ with probability $\mathbf{P}(a)$. Given a joint mixed strategy $\mathbf{P}$, let $\mathbf{P}(a_i)$ denote the individual *mixed strategy* of player i, i.e., the marginal probability that player i plays a_i in $\mathcal{A}_i$, and $\mathbf{P}(a_{-i}|a_i)$ the *conditional joint mixed strategy* of all the players except i given the action of player i, i.e., the conditional probability that, given that player i plays a_i, the other players play a_{-i}.

An *equilibrium* is generally considered the solution of a game. An equilibrium can be viewed as a point of strategic stance, where every player is "happy," i.e., no player has any incentive to unilaterally deviate from the way they play.

This paper concerns correlated equilibria. Let $G_i(a'_i, a_i, a_{-i}) \equiv M_i(a'_i, a_{-i}) - M_i(a_i, a_{-i})$ denote player i's *payoff gain* from playing a'_i instead of a_i when the other players play a_{-i}. A *correlated equilibrium (CE)* for a game $\mathcal{G}$ is a joint mixed strategy $\mathbf{P}$ such that for every joint action a drawn according to $\mathbf{P}$, each player i, knowing $\mathbf{P}$ and its own action a_i in a *only*, has no payoff gain, in expectation, from *unilateral* changing its play to another action $a'_i \in \mathcal{A}_i$ instead; formally, if for every player i, and every action pair $(a_i, a'_i) \in \mathcal{A}_i^2$, such that $a_i \neq a'_i$ and $\mathbf{P}(a_i) > 0$,

$$\sum_{a_{-i} \in \mathcal{A}_{-i}} \mathbf{P}(a_{-i}|a_i) G_i(a'_i, a_i, a_{-i}) \leq 0.$$

One way to establish the existence of CE for any game is through its connection to *Nash equilibria (NE)*. An NE is a CE $\mathbf{P}$ that is a product distribution, i.e.,$\mathbf{P}(a) = \prod_{i=1}^{n} \mathbf{P}(a_i)$ for all a, and therefore players play independently. The classical result of Nash [1951] is that for any game there exists a (Nash) equilibrium, thus CE always exist.

Given a game $\mathcal{G}$, it is computationally convenient to express its CE conditions as the following equivalent system of *linear* constraints on the joint-mixed-strategy values $\mathbf{P}(a)$: [2] for each player i, and for all $(a_i, a'_i) \in \mathcal{A}_i^2$ such that $a_i \neq a'_i$,

$$\sum_{a_{-i} \in \mathcal{A}_{-i}} \mathbf{P}(a_i, a_{-i}) G_i(a'_i, a_i, a_{-i}) \leq 0; \qquad (1)$$

for all $a \in \mathcal{A}$, $\mathbf{P}(a) \geq 0$, and $\sum_{a \in \mathcal{A}} \mathbf{P}(a) = 1$. We refer to the *set* of all $\mathbf{P}$ that satisfy this linear system as the correlated equilibria of $\mathcal{G}$ and denote it simply by $\mathcal{CE}$, i.e., $\mathbf{P} \in \mathcal{CE}$ if and only if $\mathbf{P}$ is a CE.

[2]In contrast, because an NE is product distribution, the conditions are not linear on the individual players mixed strategies $\mathbf{P}(a_i)$.

3 MAXENT CE AND ITS PROPERTIES

Given a joint mixed strategy $\mathbf{P}$, let $H(\mathbf{P}) \equiv \sum_{a \in \mathcal{A}} \mathbf{P}(a) \ln(1/\mathbf{P}(a))$ be its *(Shannon) entropy*. Formally, a Maxent CE is the joint mixed strategy $\mathbf{P}^* = \arg\max_{\mathbf{P} \in \mathcal{CE}} H(\mathbf{P})$.

We state the results presented in the remaining of this section without proofs; the proofs will become clear from the descriptions in the next section.

3.1 CONCEPTUAL PROPERTIES

Consider the following scenario. A rational player is willing to negotiate and agree to some form of "joint" strategy with the other players. At the same time, the player wants to try to conceal its own behavior, relative to the agreed-upon joint strategy, by making it difficult to predict. If every player has the same objective, how should they (agree to) play the game? Alternatively, can we suggest a joint strategy that satisfies all the players but complicates their prediction of each others' individual strategies? We propose Maxent CE as a formal and natural answer to these questions.

The concept of *conditional entropy* in information theory provides a measure of the predictability of a random process from another [Cover and Thomas, 2006]. As such, conditional entropy helps quantify the uncertainty inherent in the players' strategies. Let $\mathbf{P}(a_i|a_{-i})$ denote the *conditional mixed strategy* of player i given the actions of the other players, i.e., the probability, with respect to $\mathbf{P}$, that player i plays a_i given that the other players play a_{-i}. Let us view A_i and A_{-i} as random variables corresponding to the play of player i and the others, respectively, under $\mathbf{P}$. We can then denote by $H_{A_i|A_{-i}}(\mathbf{P}) = -\sum_{a_{-i} \in \mathcal{A}_{-i}} \mathbf{P}(a_{-i}) \sum_{a_i \in \mathcal{A}_i} \mathbf{P}(a_i|a_{-i}) \log \mathbf{P}(a_i|a_{-i})$ the entropy of the play of player i conditioned on knowledge about the play of the other players.

Consider any group of players that wishes to predict the strategy of another player based on their own play. Conceptually, we can think of the conditional entropy as a measure on how hard this is: the larger the conditional entropy, the harder the prediction. Let us allow each individual player i the opportunity to change from a suggested CE $\mathbf{P}$ to another CE $\mathbf{P}'$ but the player can only change its conditional mixed strategy $\mathbf{P}'(a_i|a_{-i})$, not the joint mixed strategy of the other players $\mathbf{P}(a_{-i})$; thus, $\mathbf{P}'(a)$ must equal $\mathbf{P}'(a_i|a_{-i})\mathbf{P}(a_{-i})$. We show that an important property of the Maxent CE $\mathbf{P}^*$ of a game is that no player can unilaterally change to or suggest another allowed CE and increase its conditional entropy of play. More formally, given a joint mixed strategy $\mathbf{P}'$ and a player i, let P'_{-i} denote the marginal of P' over A_{-i}, and let $\mathcal{CE}_i(\mathbf{P}'_{-i}) = \{\mathbf{P} \in \mathcal{CE} | \mathbf{P}(a_{-i}) = \mathbf{P}'_{-i}(a_{-i}) \text{ for all } a_{-i} \in \mathcal{A}_{-i}\}$.

Proposition 1 (Player Optimality of Maxent CE) *If $\mathbf{P}^*$ is the Maxent CE of a game $\mathcal{G}$, then for each player i, we have $\mathbf{P}^* = \arg\max_{\mathbf{P} \in \mathcal{CE}(\mathbf{P}^*_{-i})} H_{A_i|A_{-i}}(\mathbf{P})$.*

The proof follows easily from the definition of a Maxent CE and the chain rule for (joint) entropies.

3.2 SIMPLE AND COMPACT REPRESENTATION

Another interesting property of Maxent is that it is, in a sense, the most simple CE, and permits a very compact representation. Let $m = \max_i |\mathcal{A}_i|$ be the maximum number of actions of any player. Then, the number of *parameters* needed is $O(nm^2)$.

Theorem 1 (Maxent CE Representation) *Given an n-player game $\mathcal{G}$, its Maxent CE $\mathbf{P}^*$ has the following parametric form: for all a,*

$$\mathbf{P}^*(a) \propto \exp\left(-\sum_{i=1}^{n} \sum_{a'_i \in \mathcal{A}_i \setminus \{a_i\}} \lambda^*_{i,a_i,a'_i} G_i(a'_i, a_i, a_{-i})\right),$$

where $\lambda^*_{i,a_i,a'_i} \geq 0$, for all players $i = 1, \ldots, n$, and for all of its action pairs $(a_i, a'_i) \in \mathcal{A}_i^2$ such that $a_i \neq a'_i$, are the parameters.

This form is not surprising for those familiar with maxent optimization, as the parameters corresponds to Lagrange multipliers (i.e., the dual variables). In the game context, the parameters λ^* have a natural interpretation. For example, the parameter λ^*_{i,a_i,a'_i} roughly measures the tendency that player i has to prefer a_i over a'_i. For instance, $\lambda^*_{i,a_i,a'_i} > 0$ if and only if player i is *indifferent* between a_i and a'_i, i.e., the corresponding CE (inequality) condition holds with equality. Thus, if the player has a *strict preference* of a_i over a'_i, i.e., the corresponding CE condition holds with strict inequality, then λ_{i,a_i,a'_i} must be 0.

We note that due to a technical condition, the results presented in this paper should be appropriately qualified. The expression for the Maxent CE given above hints at this complication: although the Maxent always exists, because CE always exists, one can see the need to appropriately modify the representation when the Maxent CE does not have full support. The results do hold, for instance, if the game has *some* CE with full support. They hold more generally for the case of Maxent *approximate* CE, where we allow the possibility that each player can gain by deviating, but no more than some small amount $\epsilon > 0$, i.e., for all i, and for all $a_i \neq a'_i$, we allow $\sum_{a_{-i}} \mathbf{P}(a_i, a_{-i}) G_i(a'_i, a_i, a_{-i}) \leq \epsilon$. The modifications to the case of approximate CE are easy, and the details are presented in a companion technical report [Ortiz et al., 2006]. We ignore this issue throughout this paper in the interest of presentation.

3.3 PROBABILISTIC STRUCTURE

The Maxent CE can exploit available strategic structure in the game. As an example, we now discuss the probabilistic structure of Maxent CE in the particular context of *graphical games* [Kearns et al., 2001], a graphical model for game theory. The representation size of games in normal form is $O(nm^n)$, exponential in the number of players. Such sizes render the normal-form representation of games inadequate when modeling problems as games with many players, as it would be the case in many real-world problems, specially in the Internet. Just like probabilistic graphical models allow us to model large but structured probabilistic systems, graphical games allow us to deal with large-population games by providing compact representation for them.

Let $G = (V, E)$ be an undirected graph where the vertices or nodes $V = \{1, 2, \ldots, n\}$ correspond to the players in the game. The *neighbors* of a player i in G are those players that are connected to i by a single edge in E. Given a player i, we refer to the player i and its neighbors as the *neighborhood* $N_i = \{j|(i, j) \in E\} \cup \{i\}$ of player i (note that N_i includes i). The graph has a simple meaning: a player's payoff is *only* a function of the actions of players in its neighborhood. For every player i, let $k_i = |N_i|$ be the size of its neighborhood with $N_i = \{j_1, j_2, \ldots, j_{k_i}\} \subset V$ and denote by $a_{N_i} \equiv (a_{j_1}, a_{j_2}, \ldots, a_{j_{k_i}}) \in \times_{j \in N_i} \mathcal{A}_j \equiv \mathcal{A}_{N_i}$ the joint action of only players in the neighborhood of player i. Given a graph G, for each player i, its *local payoff function* $M'_i : \mathcal{A}_{N_i} \to \mathbb{R}$ maps the the joint-actions a_{N_i} of players in its *neighborhood* in G to a real number $M'_i(a_{N_i})$. For each player i, the payoff function M_i of player i is such that $M_i(a) = M'_i(a_{N_i})$. Let $\mathcal{M}' \equiv \{M'_1, \ldots, M'_n\}$ be the set of local payoff functions. A graphical game is defined by the 3-tuple $\mathcal{GG} = (\mathcal{A}, G, \mathcal{M}')$.

Recall that $m = \max_i |\mathcal{A}_i|$. Let $k = \max_i |N_i|$ be the largest neighborhood size in G. The representation size of a graphical game is $O(nm^k)$, exponential in the size of the largest neighborhood, *not* the number of players. In fact, if $k \ll n$, we obtain *exponential* savings in representation size. Note that the generality of the normal-form game is not lost because we can represent any game by using a fully connected graph; we just gain representationally for games with richer strategic structure.

Let G_N denote the *extended-neighborhood graph* of G, i.e., the graph that results from adding edges between every neighbor of a player; formally, if $G = (V, E)$, then $G_N = (V, E_N)$, where $E_N = E \cup \{(i, j)|i, j \in N_k \text{ for some } k\}$. The following result follows easily from Theorem 1 above and the strategic structure of the graphical game.

Corollary 1 (Probabilistic Structure of Maxent CE) *Given a graphical game $\mathcal{GG}$ with graph G, the Maxent CE of $\mathcal{GG}$ is a Markov random field with respect to G_N.*

There is exactly one potential in the MRF for each player, and each potential is only over the neighborhood of the player.

The representation result of Kakade et al. [2003] states that *given any CE*, one can represent it by another with the same size as the representation size of the graphical game, modulo expected payoff equivalence. The proof uses a maximum entropy argument based on matching the neighborhood marginals of the original CE. Here we concentrate explicitly on the Maxent CE and give an explicit expression for that equilibrium in terms of the players' payoff functions which uses only $O(nm^2)$ parameters, as opposed to $O(nm^k)$.

This corollary, which can also be derived from the Representation Theorem of Kakade et al. [2003], is significant because it elucidates the probabilistic structure of the CE. It also lets us exploit what is known from the literature on probabilistic graphical models. For instance, we can make qualitative statements about the structure of the CE (e.g., which players play independently *conditioned* on fixing the actions of a separate set of players) that depend only on the graph structure, and *not* the actual payoff values. This connection also establishes the efficient implementation of the Maxent CE (i.e., drawing samples efficiently from it) for bounded-tree-width graphical games, and allows approximate implementations for more general graphs through well-known and broadly used techniques such as Gibbs sampling.

3.4 COMPUTATION

The following result is the main computational contribution of this paper. The next section is devoted to its derivation.

Theorem 2 (Maxent CE Computation) *Given a game $\mathcal{G}$, there exist gradient-based algorithms with guaranteed convergence to the Maxent CE of $\mathcal{G}$.*

We also note that each iteration of the algorithms constitutes a natural step and can be computed in time polynomial in the representation size for normal-form games and, more generally, graphical games whose neighborhood graphs have bounded tree-width; as noted earlier, approximations are also possible for more general graphs.

Kakade et al. [2003] give an algorithm to compute a single exact CE in graphical games with bounded tree-width in time polynomial on the representation size of the game based on a simple linear program, [3] thus extending a well-known previous result for normal-form games [Gilboa and Zemel, 1989]. Papadimitriou and Roughgarden [2005] present an alternative algorithm for computing

[3] The result given there is for tree graphical games, but a simple application of standard arguments and ideas from the literature on probabilistic graphical models extends it to the case of bounded-tree-width graphs.

CE in bounded-tree-width graphical games and strengthens it by showing that it is hard to compute the socially optimum CE, i.e., the CE that maximizes the sum of the players' expected payoffs. For bounded-tree-width graphical games, both algorithms can guarantee a representation size for the resulting CE that is at most exponential in the tree-width of the graph of the graphical game. It is also important to note that the kind of equilibrium that previous algorithms compute corresponds, roughly speaking, to those that maximize some *linear* function of the neighborhood marginals of the graphical game.

In a surprising result, Papadimitriou [2005] gives a polynomial-time algorithm for computing a single exact CE for a large class of compactly-representable games, including graphical games. The goal there is to find *any* CE, and thus, not one with a specific property. The algorithm also uses duality, but in a different way than ours. It is also interesting that the CE found by that algorithm can be represented by a mixture of a polynomial number of product distributions as components, which is also different from the Maxent representation given above.

4 COMPUTING MAXENT CE

Computing the Maxent CE is a convex optimization problem, allowing the application of efficient algorithms already developed for such problems (see [Boyd and Vanderberghe, 2004]). Here, we solve this problem by using duality, and provide specific gradient-based algorithms that monotonically converge to the unique dual optimum. The algorithms also have a natural distributed learning-dynamics interpretation which we discuss in the next section.

Following a standard argument, we obtain that the Lagrange dual function reduces to $g(\lambda) = -\ln Z(\lambda)$ where $Z(\lambda) = \sum_a \exp(-\sum_i \sum_{a'_i \neq a_i} \lambda_{i,a_i,a'_i} G_i(a'_i, a_i, a_{-i}))$. The dual problem reduces to finding $\sup_{\lambda \geq 0} g(\lambda)$. The relationship between the dual variables λ and the primal variables $\mathbf{P} \equiv \mathbf{P}^\lambda$ is as usual: for all a,

$$\log \mathbf{P}(a) = -\sum_i \sum_{a'_i \neq a_i} \lambda_{i,a_i,a'_i} G_i(a'_i, a_i, a_{-i}) - \log Z(\lambda).$$

4.1 ALGORITHMS

We now present two gradient-based algorithms for computing $\sup_{\lambda \geq 0} g(\lambda)$. For convenience, we use the notation $[x]^+ \equiv \max(0, x)$. We also denote $\mathbf{P}^t \equiv \mathbf{P}^{\lambda^t}$,

$$R^+_{i,a_i,a'_i}(t) = \sum_{a_{-i}} \mathbf{P}^t(a_i, a_{-i})[G_i(a'_i, a_i, a_{-i})]^+,$$

$$R^-_{i,a_i,a'_i}(t) = \sum_{a_{-i}} \mathbf{P}^t(a_i, a_{-i})[-G_i(a'_i, a_i, a_{-i})]^+,$$

and

$$R_{i,a_i,a'_i}(t) \equiv \nabla_{i,a_i,a'_i} g(\lambda^t) = R^+_{i,a_i,a'_i}(t) - R^-_{i,a_i,a'_i}(t).$$

We call $R_{i,a_i,a'_i}(t)$ the *regret* that player i has for playing a_i instead of a'_i, with respect to $\mathbf{P}^t$.

The *logarithmic-gradient* algorithm is as follows: initialize $\lambda^0 \geq 0$ arbitrarily and at every round $t \geq 0$, set

$$\lambda^{t+1}_{i,a_i,a'_i} \leftarrow [\lambda^t_{i,a_i,a'_i} + \delta^t_{i,a_i,a'_i}]^+ \tag{2}$$

where

$$\delta^t_{i,a_i,a'_i} = \frac{1}{2c} \log \left(R^+_{i,a_i,a'_i}(t) / R^-_{i,a_i,a'_i}(t) \right) \tag{3}$$

and the constant c is such that $c \geq \sum_i \sum_{a'_i \neq a_i} |G_i(a'_i, a_i, a_{-i})|$ for all a_i, a_{-i}.

The dynamic-step-size *gradient-ascent* algorithm is the same except that it uses

$$\begin{aligned} \delta^t_{i,a_i,a'_i} &= \frac{1}{2c} \frac{1}{R^+_{i,a_i,a'_i}(t) + R^-_{i,a_i,a'_i}(t)} \nabla_{i,a_i,a'_i} g(\lambda^t) \\ &= \frac{1}{2c} \frac{R^+_{i,a_i,a'_i}(t) - R^-_{i,a_i,a'_i}(t)}{R^+_{i,a_i,a'_i}(t) + R^-_{i,a_i,a'_i}(t)} \\ &= \frac{1}{c} \left(\frac{R^+_{i,a_i,a'_i}(t)}{R^+_{i,a_i,a'_i}(t) + R^-_{i,a_i,a'_i}(t)} - \frac{1}{2} \right). \end{aligned} \tag{4}$$

4.2 CONVERGENCE

We show that if we use either algorithm at each round, then the sequence $\mathbf{P}^t$ converges to the Maxent CE. Due to space constraints, we omit the details of the proof and refer the reader to the companion technical report [Ortiz et al., 2006].

The core element in the proof is based on an auxiliary function, variational approach similar to Dudík et al. [2004] and Collins et al. [2002]. The idea is to lower bound the change in the dual value and then maximize that bound. The following is a sketch. Let $s_i(a'_i, a_i, a_{-i}) = 1$ if $G_i(a'_i, a_i, a_{-i}) \geq 0$, and -1 otherwise. First note that, by Jensen's inequality,

$$\begin{aligned} e^{\left(\sum_i \sum_{a'_i \neq a_i} \frac{|G_i(a'_i, a_i, a_{-i})|}{c} (-c\delta_{i,a_i,a'_i} s_i(a'_i, a_i, a_{-i})) \right)} \leq \\ \sum_i \sum_{a'_i \neq a_i} \frac{|G_i(a'_i, a_i, a_{-i})|}{c} e^{-c\delta_{i,a_i,a'_i} s_i(a'_i, a_i, a_{-i})} + \\ \left(1 - \sum_i \sum_{a'_i \neq a_i} \frac{|G_i(a'_i, a_i, a_{-i})|}{c} \right). \end{aligned}$$

Applying this inequality and linearity of expectation, we

get

$$\begin{aligned} & Z_{t+1}/Z_t \\ \leq & \sum_{i,a_i,a_i',a_i\neq a_i'} \left(e^{-c\delta_{i,a_i,a_i'}}-1\right)\frac{R^+_{i,a_i,a_i'}(t)}{c} + \\ & \left(e^{c\delta_{i,a_i,a_i'}}-1\right)\frac{R^+_{i,a_i,a_i'}(t)}{c}+1. \end{aligned}$$

Minimizing with respect to δ leads the logarithmic-gradient algorithm. The gradient-ascent algorithm follows similarly after applying the well-known inequality $1+x \leq e^x \leq 1+x+x^2$ for $|x|<1$ to the bound, before minimizing. We then show that in fact $\delta = 0$ if and only if $g(\lambda^t) = \max_{\lambda\geq 0} g(\lambda) = \max_{\mathbf{P}\in\mathcal{CE}} H(\mathbf{P})$.

5 A LEARNING-DYNAMICS VIEW

In this section, we connect the algorithms to simple learning dynamics [Singh et al., 2000]. Convergence of best-response-gradient dynamics to NE is not always possible [Hart and Mas-Colell, 2003]. In contrast, the analogous process presented here is guaranteed to converge to the Maxent CE.

Consider the following simple learning dynamics based on distributed gradient ascent. In this case, each player is known to play independently, and thus holds its own mixed strategy $\mathbf{P}_i$ over just $\mathcal{A}_i$. At every round t, each player i sends to all the other players its current strategy $\mathbf{P}_i^t$. Each player i then computes the expected payoff from playing a particular action when the others play according to their broadcasted strategy. Assuming the other players maintain their strategy, each player then improves over its strategy by changing it in the direction of the gradient of its expected payoff $\sum_a \mathbf{P}_i(a_i)(\prod_{j\neq i}\mathbf{P}_j^t(a_j))M_i(a_i,a_{-i})$ with respect to its own strategy $\mathbf{P}_i$. Formally, each player i gradient-ascent update rule is, for each a_i,

$$\mathbf{P}_i^{t+1}(a_i) \leftarrow \mathbf{P}_i^t(a_i) + \alpha^t \sum_{a_{-i}} \left(\prod_{j\neq i}\mathbf{P}_j^t(a_j)\right) M_i(a_i,a_{-i}),$$

where α^t is the step size, and ensuring that the updates lead to a proper probability distribution. There has been quite a bit of work that tries to understand the convergence properties of such update rules. However, it is known that in general, update rules of this kind are not guaranteed to converge to NE.

We can view the gradient-based algorithms presented in the previous section as a type of distributed learning via a message-passing process that is analogous to the gradient-ascent update rule just described. Now, however, because the players are negotiating a joint mixed strategy, each player suggests its own *joint* mixed strategy $\mathbf{P}_(a)$. For conceptual presentation only, let us introduce an additional "external player" or "arbiter;" just like in gradient dynamics, this is not really needed as each individual can broadcast its suggested joint mixed strategy and perform the functions of the arbiter. At each round t, the arbiter takes suggestions from each player i about its preferred joint mixed strategy $\mathbf{P}_i^t$. The arbiter then processes those suggestions by forming a single joint mixed $\mathbf{P}^t$ as follows,

$$\mathbf{P}^t(a) \propto \prod_{i=1}^n \mathbf{P}_i^t(a),$$

which it then sends back to each player for its consideration.

We can interpret $\mathbf{P}^t$ as follows. We can think of each player using their own joint mixed strategy $\mathbf{P}_i^t$ to draw joint action suggestions. Using this mechanism, the players actually perform a joint play only when *everybody* draws the *same* joint action. When play actually occurs, the probability that they all agree to play a is in fact $\mathbf{P}^t(a)$! The arbiter, in some sense, serves as a surrogate that accelerates joint-action agreements; thus, only one draw is needed!

Throughout the process, each player i maintains a parametric joint probability distribution which at any time t takes the form

$$\mathbf{P}_i^t(a) \propto \exp\left(-\sum_{a_i'\neq a_i} \lambda^t_{i,a_i,a_i'} G_i(a_i',a_i,a_{-i})\right),$$

where $\lambda^t_{i,a_i,a_i'}$ corresponds to the current value of its own parameters. Each player also updates the parameters using any of the update rules described in the previous section. We can see from the update rules that each player only needs to know its *own* payoff, current parameter values, and no other external knowledge other than the current global joint mixed strategy. When an arbiter is present, no knowledge of the other players' suggestions is needed, and without an arbiter, only the individual joint mixed strategies at the current time step are required, *not* the other players' parameters or payoff functions that define those strategies. Each player can thus maintain their own parameter and payoff values private.

The update rules have a very natural interpretation for each player. If $R_{i,a_i,a_i'}$ (i.e., the gradient of the dual) is positive, then player i has regret over playing a_i instead of a_i' with respect to the currently agreed (global) joint distribution of play. Therefore, assuming the other players maintain their strategy, the player should increase the value of $\lambda_{i,a_i,a_i'}$ so as to reduce the probability that the player wants to play a_i (and indirectly increase that of playing a_i'); a similar reasoning applies to the case that $R_{i,a_i,a_i'} < 0$. It is not hard to see that by using the learning (or pre-play negotiation) mechanism just described, the players are effectively computing the Maxent CE in a distributed way. Thus, the convergence result guarantees that this process in fact converges to the Maxent CE.

6 DISCUSSION

On the question of equilibrium selection. From a representational standpoint, "simpler" CE are better. Motivated by the representational results of Kakade et al. [2003] on CE in graphical games, we initially considered the following question: Are there natural learning rules directly leading to "compactly representable" and "reasonably structured" CE that exploit the strategic structure of the graphical game? It turns out that this question is harder than it first appears.

From a learning standpoint, it has been found that when agents in an environment simultaneously learn to act using natural learning rules, as in so called *no-regret* learning, convergence of their *empirical play* is guaranteed, not to Nash, but to the *set* of CE (see, for instance, Foster and Vohra [1999] and Hart and Mas-Colell [2000]). Currently, there is little understanding of the specific equilibria those types of learning rules might converge to. Thus, there has been considerable interest in trying to characterize in *any* way the behavior of natural learning rules leading to CE. While we believe our work is a step towards addressing this question, in our own preliminary experiments, the empirical play produced by *no-regret-type* learning rules does not seem to converge to a *specific* CE, let alone the Maxent CE. The question is, are there other kind of natural learning rules that do converge to the Maxent CE?

Because a game can have many equilibria, another related question is, which equilibria are more natural or likely to naturally arise from the players' dynamics? This "selection" question is not particular to game theory. In statistical estimation, for instance, one asks which probability distribution should be chosen out of all distributions consistent with the data. The Maximum Entropy Principle [Jaynes, 1957] has proved to be a natural and useful guiding principle with wide applicability beyond statistics. Maxent has often helped to characterize equilibria in many natural systems (particularly, in thermodynamics). We believe Maxent may be useful to study equilibria that are actually reached under simple learning dynamics and empirical play. From an information-theoretic point of view, the maximum entropy distribution is the "simplest" as a consequence of being the most "uninformative." Hence, it is not unreasonable to think that Maxent can also serve as a useful guiding principle to the question of equilibrium selection in game theory.

Maxent vs. Social Welfare. In game theory, an equilibrium is a *descriptive* concept of any *stable* outcome of the interaction of rational non-cooperative players in a game. As such, game theory by itself imposes no classification of equilibria as good or bad (as in the physical sciences, equilibria are only stable states). Understanding how stable outcomes can be achieved which are somehow "socially optimal" (a notion which in and of itself is thorny to define), particularly when players are not cooperative, is an important open problem.

It is nevertheless instructive to explore the connections between Maxent CE and other popular equilibrium-selection notions, like the socially optimum CE with respect to total welfare (i.e. the total sum of expected payoffs). We can compute the socially optimum CE using a linear program. In many simple, classical toy games (e.g., Battle of the Sexes, Shapley, etc.), the Maxent CE is a NE, and, as with NE, the Maxent CE can achieve lower payoff than the socially optimal one. As is well known, there can be both better and worse CE than NE with respect to sum total welfare in a game. Similarly, in general, we expect there to be both better and worse CE than the Maxent CE (with regards to sum total welfare). We should note that it is unclear whether the equilibrium which maximizes the sum total welfare is the most appropriate concept that characterizes complex systems.

Conclusions. We propose maximum entropy as a useful guiding principle to the question of equilibrium selection in game theory. We motivated and studied maximum entropy correlated equilibria and showed that it has some very useful representational and computational properties. In particular, we presented a logarithmic-gradient and a (dynamic-step-size) gradient-ascent algorithm and showed that they are guaranteed to converge to the Maxent CE for any game. Each round of the algorithms can be performed efficiently for a wide class of games, including normal-form and some graphical games (e.g., those with bounded tree-width); widely popular and effective approximation techniques such as Gibbs sampling can be used for general graphical games. The algorithms are similar to others used for statistical estimation which have been found effective in many practical applications. Finally, we provided a view of the algorithms in terms of learning or pre-play negotiation dynamics. We believe Maxent CE may help us understand the properties of the equilibria that the empirical play generated by some natural learning mechanisms might converge to.

Acknowledgements

We would like to thank Michael Collins for pointers to the literature on maximum entropy models as well as many useful suggestions. We also want to thank the anonymous reviewers for their useful comments and suggestions, specially those regarding the comparison between Maxent and socially optimum CE. Part of this work was done while the first and third author were at the CIS Dept. at Penn, and also while the first author was at MIT CSAIL. The current address for the first author is given.

References

Kenneth J. Arrow and Gerard Debreu. Existence of an equilibrium for a competitive economy. *Econometrica*, 22 (3):265–290, July 1954.

R.J. Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1, 1974.

R.J. Aumann. Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, 55, 1987.

A. Berger, S. Della Pietra, and V. Della Pietra. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1), March 1996.

Stephen Boyd and Lieven Vanderberghe. *Convex Optimization*. Cambridge University Press, 2004.

Xi Chen and Xiaotie Deng. 3-NASH is PPAD-complete. *Electronic Colloquium on Computational Complexity (ECCC)*, (134), 2005a.

Xi Chen and Xiaotie Deng. Settling the complexity of 2-player Nash-equilibrium. *Electronic Colloquium on Computational Complexity (ECCC)*, (140), 2005b.

Michael Collins, Robert E. Schapire, and Yoram Singer. Logistic regression, Adaboost and Bregman distances. *Machine Learning*, 48(1/2/3), 2002.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley & Sons, New York, second edition, 2006.

Konstantinos Daskalakis and Christos H. Papadimitriou. Three-player games are hard. *Electronic Colloquium on Computational Complexity (ECCC)*, (139), 2005.

Konstantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a Nash equilibrium. *Electronic Colloquium on Computational Complexity (ECCC)*, (115), 2005.

Stephen Della Pietra, Vincent Della Pietra, and John Lafferty. Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (4):1–13, April 1997.

Miroslav Dudík, Steven J. Phillips, and Robert E. Schapire. Performance guarantees for regularized maximum entropy density estimation. In *Proceedings of COLT*, 2004.

D. Foster and R. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 1997.

D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, pages 7 – 36, 1999.

Drew Fudenberg and Jean Tirole. *Game Theory*. The MIT Press, 1991.

I. Gilboa and E. Zemel. Nash and correlated equilibria: some complexity considerations. *Games and Economic Behavior*, 1:80–93, 1989.

Paul W. Goldberg and Christos H. Papadimitriou. Reducibility among equilibrium problems. *Electronic Colloquium on Computational Complexity (ECCC)*, (090), 2005.

Sergiu Hart and Andreu Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68(5):1127 – 1150, 2000.

Sergiu Hart and Andreu Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, 93(5):1830–1836, 2003.

E. T. Jaynes. Information theory and statistical mechanics I & II. *Phys. Rev.*, 1957.

Sham Kakade, Michael Kearns, John Langford, and Luis Ortiz. Correlated equilibria in graphical games. In *ACM Conference Proceedings on Electronic Commerce*, 2003.

M. Kearns, M. Littman, and S. Singh. Graphical models for game theory. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence*, pages 253–260, 2001.

J. F. Nash. Non-cooperative games. *Annals of Mathematics*, 54:286–295, 1951.

Luis E. Ortiz, Robert E. Schapire, and Sham M. Kakade. Maximum entropy correlated equilibrium. Technical Report TR-2006-21, CSAIL MIT, Cambridge, MA USA, March 2006.

Christos Papadimitriou. Algorithms, games, and the Internet. In *STOC '01: Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 749–753, 2001.

Christos H. Papadimitriou. Computing correlated equilibria in multi-player games. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 49–56, 2005.

Christos H. Papadimitriou and Tim Roughgarden. Computing equilibria in multi-player games. In *SODA '05: Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 82–91, 2005.

Satinder P. Singh, Michael J. Kearns, and Yishay Mansour. Nash convergence of gradient dynamics in general-sum games. In *UAI*, pages 541–548, 2000.