Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2006-021 March 20, 2006

Maximum Entropy Correlated Equilibria

Luis E. Ortiz, Robert E. Schapire, and Sham M. Kakade

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Maximum Entropy Correlated Equilibria

Luis E. Ortiz[1] ⋆, Robert E. Schapire[2], and Sham M. Kakade[3]

[1] Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
[2] Department of Computer Science
Princeton University
[3] Department of Computer and Information Science
University of Pennsylvania

Abstract. We study maximum entropy correlated equilibria in (multi-player) games and provide two gradient-based algorithms that are guaranteed to converge to such equilibria. Although we do not provide convergence rates for these algorithms, they do have strong connections to other algorithms (such as iterative scaling) which are effective heuristics for tasks such as statistical estimation.

1 Introduction

The computer science community has been vigorously attacking problems in game theory and economics, especially as it relates to the computation of the most important notions of equilibria. The general interest is threefold: to understand the barriers on the related computational questions, to push the boundaries of what can be efficiently computed, and to understand how equilibria result from learning dynamics. This paper presents work in this direction.

The problem of computing and learning equilibria is not only of interest to game theory and economics. The artificial intelligence community wants to analyze, understand, and more generally, reason about the possible behavior of a large population of agents interacting in an environment. Before this can be done, there must be study of the types of behaviors that can arise from such interactions. Large population game theory provides a suitable mathematical framework.

There is one caveat. Compact representations of multi-player games are needed to adequately model a large number of agents. Artificial intelligence has something to offer in this regard. Mostly based on the rich work in artificial intelligence on compact representations, new compact representations for multi-player games have been recently introduced to game theory which exploit the strategic structure of the players in the game [La Mura, 2000, Kearns et al., 2001, Koller and Milch, 2003]. In this paper, we are primarily concerned with a simple but powerful model called a *graphical game* [Kearns et al., 2001].

Here, we study the important equilibrium notion called *correlated equilibrium* [Aumann, 1974]. This deviates from Nash equilibrium, the classical and until recently the dominating notion of equilibrium in game theory, in several respects. While Nash equilibria assume that players play fully independently, correlated equilibria allow a limited amount of correlation between all the players'

⋆ Orig: Oct. 5, 2004; Last Modified: Feb. 2, 2005.

actions, while still keeping each player's response "greedy" or game-theoretic. This notion also turns out to be consistent with and optimal from a Bayesian perspective [Aumann, 1987]. From a learning standpoint, it has been found that when agents in an environment simultaneously learn to act using natural learning rules, convergence of their empirical play is guaranteed, not to Nash, but to (the set of) correlated equilibria (see, for instance, Foster and Vohra [1999] and Hart and Mas-Colell [2000]). Currently, there is little understanding of the (class of) equilibria they converge to — a point we return to below.

By definition, a correlated equilibrium is a joint probability distribution over the players' actions and as such can in fact be arbitrarily complex. This is so, even when the game has some strategic structure. Therefore, roughly speaking, the representation size can be large as a function of the representation size of the game.

From a representational standpoint, the interest is in "simpler" correlated equilibria, especially the ones that can be compactly represented. Motivated by the recent representational results of Kakade et al. [2003] on correlated equilibria in graphical games, we initially considered the following question: Are there natural learning rules directly leading to "compactly representable" and "reasonably structured" correlated equilibria that exploit the strategic structure of the graphical game? It turns out that this question is harder than it first appears. In fact, there has been considerable interest for a while now on trying to characterize, in *any* way, the behavior of natural learning rules leading to correlated equilibria in games. However, here we present preliminary results that might be helpful in trying to address this question.

A game can have many equilibria. This then leads to the obvious question: 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 choosen out of all distributions consistent with the data. The *Maximum Entropy Principle*, due to Jaynes, has proved to be a natural and useful guiding principle [Jaynes, 1957]. This principle has also found wide applicability beyond statistics. In many natural systems, maximum entropy has often helped to characterize equilibria (particularly, in thermodynamics). Similarly, maximum entropy may be useful when studying equilibria that are actually reached under simple learning dynamics. 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 the maximum entropy principle can also serve as a useful guiding principle to the question of equilibrium selection in game theory. [4]

The main contribution of this paper is the computational study of maximum entropy correlated equilibria. First, we show that maximum entropy (approximate) correlated equilibria can be computed using natural and simple gradient-

[4] We point out in passing that although product distributions are in general more entropic, it is not always the case that the most entropic correlated equilibrium is a Nash equilibrium, as can be shown using simple games as examples.

based algorithms. [5] Each round of the algorithms can be performed efficiently for graphical games with bounded tree-width (which include normal-form games). [6] The convergence rates of the algorithms, however, remain an open problem.[7] Regardless, the class of algorithms we present have been used quite effectively for a variety of problems on many domains. In particular, these results highlight an interesting connection between iterative scaling algorithms for statistical estimation in natural language processing (see Berger et al. [1996]) and our algorithms for computing the game-theoretic notion of maximum entropy correlated equilibria. Second, building on the work of Kakade et al. [2003] (which showed that maximum entropy correlated equilibria can be compactly represented), we give a specific (compact) expression of the maximum entropy equilibria in terms of the game payoffs that elucidates its probabilistic properties. Third, we can sample the equilibria efficiently for graphical games with bounded tree width; for general graphical games, the equilibria representation naturally lends itself to heuristics such as a Gibbs sampler. All of our results also easily extend to the case of maximum *relative* entropy correlated equilibria (with respect to a reference distribution).

We note in passing that most prior work on computing correlated equilibria has concentrated in the computation of equilibria which optimize some linear function of the probabilities defining the equilibria. There has not been much work on computing equilibria with other specific properties, as we do here.

Finally, we provide a game theoretic (and, in some sense, a learning) interpretation of our algorithm — as a way by which players, with the help of an external arbiter, distributively compute the maximum entropy correlated equilibrium for the game. It remains an open problem whether there exist more natural learning rules (such as regret based ones) which converge to such equilibria.

2 Preliminaries

Let $\mathcal{A}$ be the finite set of m actions for each player. An n-player, m-action game in *strategic, normal or matrix form* $\mathcal{G} = \{M_i : i = 1, \ldots, n\}$ is given by a set of *payoff functions* (or multi-dimensional matrices) $M_i : \mathcal{A}^n \to \mathbb{R}$, one for each player $i = 1, \ldots, n$. The payoff function M_i for player i maps each joint-action $\boldsymbol{a} \in \mathcal{A}^n$ for all players to a real value $M_i(\boldsymbol{a})$.

An equilibrium is generally considered the solution notion in games. Denote by $\boldsymbol{a}[i : a_i'] \equiv (a_1, \ldots, a_{i-1}, a_i', a_{i+1}, \ldots, a_n)$ the vector resulting from setting the i^{th} element of $\boldsymbol{a}$ to a_i' while keeping all its other elements the same. For any $\epsilon \geq 0$, an *ϵ-correlated equilibrium (ϵ-CE)* for an n-player, m-action game in

[5] Our results only apply to *approximate* equilibria due to a technical condition. It is open whether this could be made to apply to non-approximate equilibria.

[6] For more general graphical games, we can heuristically perform stochastic gradient ascent by using a Gibbs sampler to estimate the gradient at each round.

[7] We are currently exploring whether it is possible to extend the ideas used by Bartlett et al. [2004] for their exponentiated gradient algorithm to this context.

normal form $\mathcal{G} = \{M_i\}$ is a joint probability distribution P over the joint action space $\mathcal{A}^n$ such that for every player i, and every action pair $(j,k) \in \mathcal{A}^2, j \neq k$,

$$\sum_{\boldsymbol{a} \in \mathcal{A}^n : a_i = j} P(\boldsymbol{a}) M_i(\boldsymbol{a}) \geq \sum_{\boldsymbol{a} \in \mathcal{A}^n : a_i = j} P(\boldsymbol{a}) M_i(\boldsymbol{a}[i:k]) - \epsilon \qquad (1)$$

A *correlated equilibrium (CE)* is a 0-CE. [8]

A correlated equilibrium can be explained conceptually by introducing an "arbiter" or agent external to the players. The arbiter "holds" the correlated equilibrium and implements it; that is, having access to a randomization device, the arbiter randomly samples joint actions from the correlated equilibrium and uses them as individual suggestions to the agents of which action to take. Knowing only their own suggestion sent to them by the arbiter and the correlated equilibrium, each agent has no incentive to unilaterally deviate from the suggestion sent to them. For $\epsilon = 0$, condition 1 above can be interpreted as saying that, for every player i, and for every possible action a_i that i can possibly be told to play by the arbiter (i.e., $P(a_i) > 0$), if i is in fact told to play a_i, then the *conditional* expected payoff i can obtain by actually playing a_i is not worst than what it can obtain by ignoring the suggestion and playing any other action $a_i' \neq a_i$ instead. Hence, the player has no incentive to unilaterally deviate from playing the suggested action.

A Nash equilibrium (NE) is a CE such that P is a product distribution; that is $P(\boldsymbol{a}) = \prod_{i=1}^{n} P(a_i)$. The classical result of Nash is that for any game there exists a NE [Nash, 1951]. Since a NE is a CE, for any game there exists a CE.

Here we consider *maximum entropy ϵ-correlated equilibrium (ME ϵ-CE)*. This is the ϵ-CE P such that its entropy $H(P) \equiv \sum_{\boldsymbol{a}} P(\boldsymbol{a}) \ln(1/P(\boldsymbol{a}))$ is maximal.

The representation size of games in normal form is 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. We now formally define graphical games, which provide an alternative more succinct representation.

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 sometimes refer to the player i and its neighbors as the *neighborhood* $N_i \subset V$ 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 $k = |N_i|$ be the size of its neighborhood with $N_i = \{i_1, i_2, \ldots, i_k\} \subset V$ and denote by $\boldsymbol{a}[N_i] \equiv (a_{i_1}, a_{i_2}, \ldots, a_{i_k})$ 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 $\boldsymbol{a}[N_i] \in \mathcal{A}^{|N_i|}$ of players in its neighborhood in G to a real number $M_i'(\boldsymbol{a}[N_i])$. For each player i, the payoff function M_i of player i is such that $M_i(\boldsymbol{a}) = M_i'(\boldsymbol{a}[N_i])$. A graphical

[8] Note that the definition of approximate or ϵ-correlated equilibria we give above is the same as that used in the literature on learning in games; see, for instance, Hart and Mas-Colell [2000].

game $\mathcal{GG} = (G, \{M_i'\})$ is defined by a graph G and the set of local payoff functions $\{M_i'\}$.

If $k = \max_i |N_i|$ is the largest neighborhood size in the graph of the graphical game, the representation size of a graphical game is $O(nm^k)$; thus the representation size is 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 we have not lost the generality of the normal-form game since we can represent any game by using a fully connected graph; we just gained representationally for games with richer strategic structure. The equilibrium concepts for normal-form games also naturally apply to graphical games.

3 Related Work

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 maximum entropy CE and give an explicit expression for that equilibrium in terms of the players' payoff functions.

Of the quickly growing number of results for computing equilibria in games, including some for graphical games, two of them particularly regarding the computation of *exact* CE should be noted. First is the result of Kakade et al. [2003]. This states that one can compute a single CE in graphical games with bounded tree width in time polynomial on the representation size of the game. The result of Papadimitriou and Roughgarden [2004] also presents an alternative algorithm for computing CE in bounded-tree-width graphical games and strengthens the result by stating that, assuming $P \neq NP$, there is no polynomial time algorithm for computing a correlated equilibrium of an (arbitrary) graphical game that maximizes the expected sum of players payoff (i.e., the "social optimum" CE). For bounded-tree-width graphical games, both algorithms can guarantee a representation size for the resulting CE that is exponential in the tree-width of the graph of the graphical game. The number of parameters for that representation is therefore exponentially larger (as a function of the tree width) than that for our maximum entropy ϵ-CE. It is also important to note that the kind of equilibrium that they compute corresponds, roughly speaking, to those that maximize some linear function of the neighborhood marginals of the graphical game. The maximum entropy CE is likely to have different properties in general.

Concurrently with this paper, Papadimitriou [2004] has shown that one can compute a single correlated equilibrium in polynomial time for a large class of "compact" games, including graphical games. This is a surprising result, especially given the previous negative complexity result on computing the socially optimal CE. His goal is to find *any* correlated equilibria, while ours is to find specifically the most entropic CE. Similar to ours, his algorithm is based on duality, but it is used in a different way than ours. It is also interesting that the CE found by his algorithm can be represented by a mixture of a polynomial

number of product distributions as components, which is a different representation than that for the most entropic CE. In fact, as we will see below, the most entropic CE is a Gibbs distribution (or, equivalently, a Markov random field) which allows us to make qualitative conditional independence statements about the players' play at equilibria, a property that is not evident in the CE obtained by Papadimitriou's algorithm.

4 Results

We now present the main results of the paper and leave their proof for subsequent sections.

Theorem 1. *(ME ϵ-CE Representation) Given $\epsilon > 0$ and an n-player, m-action game $\mathcal{G}$, the maximum entropy ϵ-CE has the form*

$$P(\boldsymbol{a}) = \left(\prod_{i=1}^{n} \exp\left(\sum_{a_i' \neq a_i} \lambda_{i,a_i,a_i'} (M_i(\boldsymbol{a}) - M_i(\boldsymbol{a}[i:a_i'])) \right) \right) \Big/ Z(\lambda)$$

where $Z(\lambda) = \sum_{\boldsymbol{a}} \left(\prod_{i=1}^{n} \exp\left(\sum_{a_i' \in \mathcal{A}: a_i' \neq a_i} \lambda_{i,a_i,a_i'} (M_i(\boldsymbol{a}) - M_i(\boldsymbol{a}[i:a_i'])) \right) \right)$, for some values for the parameters $\{\lambda_{i,a_i,a_i'} \geq 0 : i = 1, \ldots, n, (a_i, a_i') \in \mathcal{A}^2, a_i \neq a_i'\}$,

The parameters λ have a natural interpretation. For example, the parameter $\lambda_{i,a_i,a_i'}$ roughly measures the tendency that player i has to prefer a_i' over a_i. For instance, if player i strictly prefers a_i over a_i', then $\lambda_{i,a_i,a_i'} = 0$; otherwise $\lambda_{i,a_i,a_i'} > 0$.

A *local Markov Network* is a probabilistic graphical model particularly suited for representing correlated equilibria in graphical games. It borrows the same ideas from other traditional probabilistic graphical models and exploits probabilistic structure to compactly represent joint probability distributions. We refer the reader to Kakade et al. [2003] for a formal definition.

Corollary 1. *(ME ϵ-CE Probabilistic Structure) Given $\epsilon > 0$ and an n-player, m-action graphical game $\mathcal{GG}$ with graph G, the maximum entropy ϵ-CE of $\mathcal{GG}$ is a local MN with respect to G.*

This corollary follows trivially from the proof of the Representation Theorem of Kakade et al. [2003]. It also follows trivially from Theorem 1 above and the strategic structure of the graphical game. This corollary is significant because it makes explicit the probabilistic structure of the equilibrium. It 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 equilibrium such as which players play independently *conditioned* on fixing the actions of a separate set of players.

Another consequence of the form of the most entropic ϵ-CE given above is that it permits efficient sampling for bounded-tree width graphical games and lends itself to natural sampling heuristics (e.g., Gibbs sampling) for more general graphs.

Theorem 2. *(ME ϵ-CE Computation) Given $\epsilon > 0$ and an n-player, m-action game $\mathcal{G}$, there exist gradient-based algorithms with guaranteed convergence to the maximum entropy ϵ-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.

5 Computing Maximum Entropy Correlated Equilibria

5.1 Optimization Setup

The problem of computing the most entropic ϵ-CE can be formally stated as follows:

$$\arg\min_{P} \sum_{\boldsymbol{a}} P(\boldsymbol{a}) \log P(\boldsymbol{a})$$

such that the ϵ-CE constraints (1) on P hold and P is a proper probability distribution: for all $\boldsymbol{a}$,

$$P(\boldsymbol{a}) \geq 0$$

$$\sum_{\boldsymbol{a}} P(\boldsymbol{a}) = 1.$$

To simplify notation, we recast the problem above. First, fix a pair of positive integers c and r. Let the vector $p \in \mathbb{R}^c$ and denote its jth element as p_j. Let the matrix $A \in \mathbb{R}^{r \times c}$ and denote its ith-row,jth-column entry as a_{ij}. For $p \in \mathbb{R}^c_+$, let $H(p) \equiv \sum_{j=1}^{c} p_j \log(1/p_j)$ be the *entropy* of p. Given $\epsilon > 0$, we want to find [9]

$$\arg\min_{p \in \mathbb{R}^c} -H(p) \tag{2}$$

subject to

$$\sum_{j=1}^{c} p_j = 1, \forall j = 1, \ldots, c, p_j \geq 0, \tag{3}$$

and

$$\forall i = 1, \ldots, r, \sum_{j=1}^{c} p_j a_{ij} \leq \epsilon. \tag{4}$$

In the context of our original problem, it is natural to refer to a p satisfying the constraints in (3) above as a *probability distribution (PD)* and a PD p additionally satisfying the constraints in (4) as an ϵ-CE.

In what follows, to simplify notation, assume that the index i has range $1, \ldots, r$, and the index j has range $1, \ldots, c$.

In our problem we can assume the following about the matrix A.

[9] Note that the optimization is well defined since $p \in \mathbb{R}^c_+$ in the constraint space, and therefore $H(p)$ is well defined there.

Assumption 1 *(Linear Transformation Invariance)* $\sum_i |a_{ij}| \leq 1$ *for all* j.

Assumption 2 *(CE Existence) There exists a PD* p *such that* $\sum_j p_j a_{ij} \leq 0$ *for all* i.

Assumption 3 *(No Dominated Actions) For every* i*, there exist a column* j^+ *and a column* j^- *such that* $a_{ij^+} > 0$ *and* $a_{ij^-} < 0$.

The first assumption results because a CE is unaffected by linear transformations of each player's payoffs. The second assumption corresponds to the CE conditions and a CE always exists for any game. The third assumption follows because otherwise there are dominated strategies for a player which can easily be removed from consideration (i.e., the action and related constraints can be ignored without changing the result). [10]

The above is the same problem as that recently studied by Dudík et al. [2004], except for the small technicality that for some i, j, $a_{ij} < 0$. The approach used follows closely that in Collins et al. [2002].

We can solve the optimization problem given in (2), (4) and (3) above by using duality. In what follows, we naturally refer to this problem as the *primal problem* and the variables p as the *primal variables*. Before continuing, we note in passing that for this problem there is no optimal duality gap.[11] Following a standard argument (see Appendix 7.1 for the details), we obtain that the Lagrange dual function reduces to $g(\lambda) = -\ln Z(\lambda) - \epsilon \sum_i \lambda_i$, where $Z(\lambda) = \sum_j \exp(-\sum_i \lambda_i a_{ij})$. The *dual problem* reduces to finding $\sup_{\lambda \geq 0} g(\lambda)$. The relationship between the dual variables λ and the primal variables p is as follows: for all j,

$$p_j(\lambda) = \exp\left(-\sum_i \lambda_i a_{ij}\right) \Big/ Z(\lambda).$$

5.2 Algorithms

We now present two gradient-based algorithms for computing $\sup_{\lambda \geq 0} g(\lambda)$.

We call the first algorithm the *logarithmic gradient* algorithm. First, we initialize λ^0 arbitrarily to some vector with nonnegative components (for instance, $\lambda_i^0 = 0$ for all i). Then, at every round $t \geq 0$, we set

$$\lambda_i^{t+1} \leftarrow \max\left(0, \lambda_i^t + \delta_i'(\lambda^t)\right) \tag{5}$$

where

$$\delta_i'(\lambda^t) \equiv \ln\left(\frac{\sqrt{4W_i^-(t)W_i^+(t) + \epsilon^2} - \epsilon}{2W_i^-(t)}\right),$$

[10] Note also that if for some i, $a_{ij} = 0$ for all j, then the i^{th} constraint can be ignored and removed from the system of inequality constraints since it will always hold.

[11] For this problem, the (weaker) Slater's condition holds if there is a ϵ-CE with full support $p > 0$. If $\tilde{p}$ satisfies Assumption 1 above and p is a properly constructed mixture of $\tilde{p}$ and the uniform distribution, then p satisfies the needed condition.

$$W_i^-(t) \equiv W_i^-(p^t) = \sum_{j:a_{ij}<0} p_j^t |a_{ij}|,$$

$$W_i^+(t) \equiv W_i^+(p^t) = \sum_{j:a_{ij}>0} p_j^t |a_{ij}|$$

and p^t is such that

$$p_j^t \equiv p_j(\lambda^t)$$

for all j.

The second algorithm is a dynamic-step-size *gradient-ascent* algorithm. We first initialize λ^0 arbitrarily to some vector with nonnegative components (as done above for the first algorithm), and then, at every round, we set

$$\lambda_i^{t+1} \leftarrow \lambda_i^t + \max\left(\delta_i'(\lambda^t), -1, -\lambda_i^t\right) \tag{6}$$

where

$$\delta_i'(\lambda^t) \equiv (1/2)\frac{1}{W_i^+(t) + W_i^-(t)} \nabla_i g(\lambda^t)$$

and

$$\nabla_i g(\lambda^t) = W_i^+(t) - W_i^-(t) - \epsilon = \sum_j p_j^t a_{ij} - \epsilon$$

is the i^{th} component of the gradient of g evaluated at λ^t.

5.3 Convergence

We first show that, for both algorithms presented above, $\{g(\lambda^t)\}$ converges monotonically to the optimal value of the reduced dual problem $\sup_{\lambda \geq 0} g(\lambda)$. For this, we use an auxiliary function approach similar to Dudík et al. [2004] and Collins et al. [2002]. This leads to the following theorem. We refer the reader to Appendix 7.2 for the proof.

Theorem 3. *Let $\{\lambda^t\}$ be the sequence of parameters generated by the gradient-based algorithms. Then the sequence $\{g(\lambda^t)\}$ converges monotonically to $\sup_{\lambda \geq 0} g(\lambda)$.*

We then show that the corresponding sequence of primal values in fact converges to the optimal value. Note that this does not follow immediately from the last theorem. Note also that the primal optimal is unique. [12] We now present the main theorem.

Theorem 4. *Let $\{\lambda^t\}$ be the sequence of parameters generated by the gradient-based algorithms. Then the sequence $\{p(\lambda^t)\}$ converges to the primal optimal.*

The proof is in Appendix 7.3.

Theorem 1 follows from the expression for the primal optimal. Theorem 2 follows immediately from the last theorem.

[12] To see this, let $\mathcal{CE}$ be the set of all CE. Suppose there exists $p, q \in \arg\max_{p \in \mathcal{CE}} H(p), p \neq q$. Since $\mathcal{CE}$ is a convex set, for $\alpha \in (0,1)$, $p' = \alpha p + (1-\alpha)q \in \mathcal{CE}$. Since H is strictly concave, $H(p') > H(p) = H(q)$, a contradiction.

6 A Distributed Game-Theoretic Interpretation

We can view the gradient-based algorithms presented above by which we can compute the most-entropic ϵ-CE as a type of distributed message-passing process. We introduce an additional "external player" or "arbiter" which, at each round t, takes suggestions for the probability distribution of play P_i^t over the joint actions space $\mathcal{A}^n$ from each player i. The arbiter then processes those suggestions by forming a single joint probability distribution P^t over the joint actions space $\mathcal{A}^n$, which it then sends to each player.

Let each player use the update rule above. At round $t = 0$, each player i sends to the arbiter the probability distribution P_i^0 based on its initial parameter values $\Lambda_i^0 = \{\lambda_{i,a_i,a_i'}^0 \geq 0 : (a_i, a_i') \in \mathcal{A}^2, a_i \neq a_i'\}$. The arbiter computes P^t as

$$P^t(\boldsymbol{a}) = (1/Z^t) \prod_{i=1}^{n} P_i^t(\boldsymbol{a})$$

where

$$Z^t = \sum_{\boldsymbol{a}} \prod_{i=1}^{n} P_i^t(\boldsymbol{a})$$

is the normalizing constant and $\sum_{\boldsymbol{a}} P_i^t(\boldsymbol{a}) = 1$ for all i. Then, the arbiter sends P^t to each player. Now each player, knowing the current value of its own parameters Λ_i^t, its own payoff matrices and P^t, updates the parameters values to obtain Λ_i^{t+1}, using one of the update rules presented above. Each player then computes the new suggestion P_i^{t+1}, which is such that for all $\boldsymbol{a}$,

$$P_i^{t+1}(\boldsymbol{a}) = \exp\left(\sum_{a_i' \neq a_i} \lambda_{i,a_i,a_i'}^{t+1} [M_i(\boldsymbol{a}) - M_i(\boldsymbol{a}[i : a_i'])]\right) \Big/ Z_i^{t+1}$$

where Z_i^{t+1} is the new (local) normalizing constant.

Note that in this message-passing process, for each player i, only player i knows its own payoff and parameter values, or in other words, any other player different than i nor the arbiter know player i's payoff or its parameter values. The only "communication" between the players is through the arbiter, or more specifically, through the global probability distribution P^t. The players do not see each other's suggestions to the arbiter.

Acknowledgments. We would like to thank Michael Collins for pointers to the literature on maximum entropy models as well as many useful suggestions.

References

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.

Peter Bartlett, Michael Collins, Ben Taskar, and David McAllester. Exponentiated gradient algorithms for large-margin structured classification. In *NIPS*, 2004.

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

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

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

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

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.

Daphne Koller and Brian Milch. Multi-agent influence diagrams for representing and solving games. *Games and Economic Behavior*, 45(1):181–221, 2003.

Pierfrancesco La Mura. Game networks. In *Proceedings of the 16th Annual Conference on Uncertainty in Artificial Intelligence (UAI-00)*, 2000.

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

Christos H. Papadimitriou. Computing correlated equilibria in multiplayer games. To appear in STOC, 2004.

Christos H. Papadimitriou and Tim Roughgarden. Computing equilibria in multi-player games. Preprint, April 2004.

7 Appendix

7.1 Lagrange Dual Function

We now present the standard argument leading to a reduced Lagrange dual function in our case. The Lagrangian for this problem is

$$L(p,\lambda,\alpha,\beta) = -H(p) + \sum_i \lambda_i \left(\sum_j p_j a_{ij} - \epsilon \right) - \sum_j \alpha_j p_j + \beta \left(\sum_j p_j - 1 \right).$$

We can obtain the Lagrange dual function as follows. We first take derivatives with respect to p_j for all j,

$$\partial L / \partial p_j = \log p_j + 1 + \sum_i \lambda_i a_{ij} - \alpha_j + \beta.$$

Setting the derivatives to 0 we get that, abusing notation,[13]

$$p^* \equiv p^*(\lambda,\alpha,\beta) = \arg\inf_p L(p,\lambda,\alpha,\beta)$$

[13] In what follows, we often abuse the notation to simplify the presentation.

is such that

$$p_j^* = \exp\left(-\beta - 1 + \alpha_j - \sum_i \lambda_i a_{ij}\right).$$

for all j. From this we get that the Lagrange dual function is

$$g(\lambda, \alpha, \beta) = \inf_p L(p, \lambda, \alpha, \beta) = -\sum_j p_j^* - \beta - \epsilon \sum_i \lambda_i.$$

We now optimize the Lagrange dual function. We first do so in terms of β. The derivative with respect to β is

$$\partial g / \partial \beta = \sum_j \exp\left(-\beta - 1 + \alpha_j - \sum_i \lambda_i a_{ij}\right) - 1.$$

Setting the derivative to 0 we get that

$$\beta^* \equiv \beta^*(\lambda, \alpha) = \arg\sup_\beta g(\lambda, \alpha, \beta)$$

is such that

$$\exp(\beta^* + 1) = \sum_j \exp\left(\alpha_j - \sum_i \lambda_i a_{ij}\right) \equiv Z(\lambda, \alpha).$$

We then have

$$g(\lambda, \alpha) \equiv g(\lambda, \alpha, \beta^*) = -\ln Z(\lambda, \alpha) - \epsilon \sum_i \lambda_i.$$

Now in maximizing over α we note that since $Z(\lambda, \alpha)$ increases with α_j for all j, we have that

$$\alpha^* \equiv \alpha^*(\lambda) = \arg\sup_{\alpha \geq 0} g(\lambda, \alpha)$$

is such that

$$\alpha_j^* = 0$$

for all j. Therefore, we have

$$Z(\lambda) \equiv Z(\lambda, \alpha^*) = \sum_j \exp\left(-\sum_i \lambda_i a_{ij}\right)$$

and

$$g(\lambda) \equiv g(\lambda, \alpha^*) = -\ln Z(\lambda) - \epsilon \sum_i \lambda_i.$$

7.2 Proof of Theorem 3

Here we prove the theorem for the case of the logarithmic gradient algorithm; the case for the dynamic-step-size gradient-ascent algorithm is almost identical (details omitted). The proof is given as an itemized list of individual statements followed, when needed, by the main arguments involved in their proof. Some of the statements are actually used later in the proof of Theorem 4.

- Let $s_{ij} = 1$ if $a_{ij} \geq 0$, and -1 otherwise. To simplify notation, let $p_j \equiv p_j(\lambda)$, $W_i^+ \equiv W_i^+(p)$ and $W_i^- \equiv W_i^-(p)$. For $\lambda \geq 0$, $\delta \geq -\lambda$,

$$\begin{aligned} g(\lambda) - g(\lambda+\delta) &= \ln \sum_j p_j \exp\left(-\sum_i \delta_i a_{ij}\right) + \epsilon \sum_i \delta_i \\ &= \ln \sum_j p_j \exp\left(-\sum_i \delta_i |a_{ij}| s_{ij}\right) + \epsilon \sum_i \delta_i \\ &\leq \ln\left(1 + \sum_j p_j \sum_i |a_{ij}|(\exp(-\delta_i s_{ij}) - 1)\right) + \epsilon \sum_i \delta_i \\ &\leq \sum_j p_j \sum_i |a_{ij}|(\exp(-\delta_i s_{ij}) - 1) + \epsilon \sum_i \delta_i \\ &= \sum_i (e^{-\delta_i} - 1) W_i^+ + (e^{\delta_i} - 1) W_i^- + \epsilon \delta_i \equiv A(\lambda, \delta). \end{aligned}$$

- Let $\delta^* \equiv \delta^*(\lambda) = \arg\inf_{\delta \geq -\lambda} A(\lambda, \delta)$. Using simple calculus, we find that δ^* is such that

$$\delta_i^* = \max(-\lambda_i, \delta_i')$$

where

$$\delta_i' \equiv \delta_i'(\lambda) = \ln\left(\frac{\sqrt{4W_i^+ W_i^- + \epsilon^2} - \epsilon}{2W_i^-}\right).$$

Note also that the *auxiliary function* $A(\lambda) \equiv A(\lambda, \delta^*) \leq A(\lambda, 0) = 0$ and continuous.

- $A(\lambda) = 0$ *if and only if* $\delta^*(\lambda) = 0$.
Pf. For all i, let $f_i(x) = (\exp(-x) - 1)W_i^+ + (\exp(x) - 1)W_i^- + \epsilon x$. Note that $\delta_i^* = \arg\inf_{\delta_i \geq -\lambda_i} f_i(\delta_i)$ and $f_i(\delta_i^*) \leq f_i(0) = 0$. If $A(\lambda) = 0$ then $\sum_i f_i(\delta_i^*) = 0$ which implies that for all i, $f_i(\delta_i^*) = 0$. Since f_i is a strictly convex function, $f_i(\delta_i^*) = 0$ if and only if $\delta_i^* = 0$, or in other words, if $\delta_i^* \neq 0$ then $f_i(\delta_i^*) < f_i(0) = 0$.
- $\delta^*(\lambda) = 0$ *if and only if* $p(\lambda)$ *and* λ *are primal and dual optimal.*
Pf. We just need to check that the Karush-Kuhn-Tucker (KKT) conditions are satisfied. Note that $\delta_i^* = 0$ implies that either (a) $\delta_i' = \delta_i^* = 0$, or (b) $-\lambda_i = \delta_i^* = 0 \geq \delta_i'$. If (a), then, using some algebra,

$$\partial L(p, \lambda)/\partial \lambda_i = \sum_j p_j a_{ij} - \epsilon = W_i^+ - W_i^- - \epsilon = 0.$$

Similarly, if (b), then $\lambda_i = 0$ and

$$\partial L(p,\lambda)/\partial\lambda_i = \sum_j p_j a_{ij} - \epsilon = W_i^+ - W_i^- - \epsilon \le 0.$$

- Let $g^t \equiv g(\lambda^t)$. The sequence $\{g^t\}$ is nondecreasing.
- *The reduced Lagrange dual function $g(\lambda)$ is non-positive and continuous.*
  Pf. Let $H(p;q) \equiv \sum_j q_j \ln(1/p_j)$. For $\tilde{p}$ a CE and for all q,

$$\begin{aligned} g(\lambda) &= -H(p;q) - \sum_i \lambda_i \left(\epsilon - \sum_j q_j a_{ij}\right) \qquad (7) \\ &\le -\sum_i \lambda_i \left(\epsilon - \sum_j \tilde{p}_j a_{ij}\right) \\ &\le 0. \end{aligned}$$

- The sequence $\{g^t\}$ converges. Let g^* be its limit point.
- *The sequence of Lagrange multipliers $\{\lambda^t\}$ lies in a compact set.*
  Pf. Note that

$$\sum_i \lambda_i^t \left(\epsilon - \sum_j q_j a_{ij}\right) \le -g(\lambda^t) \le -g(\lambda^0)$$

  and for $q = \tilde{p}$ a CE, $\epsilon - \sum_j \tilde{p}_j a_{ij} \ge \epsilon > 0$. Thus, $0 \le \lambda_i^t \le -g(\lambda^0)/\epsilon$ for all i and t.
  Note that the need for the λ^t to lie in a compact set is the only reason why we consider ϵ-CE instead of just a CE.
- The sequence of Lagrange multipliers $\{\lambda^t\}$ has a converging subsequence. Let $\{\lambda^{t_l}\}$ be such subsequence and $\hat{\lambda}$ its limit point.
- The subsequence $\{g^{t_l}\}$ converges to $g^* = \lim_l g(\lambda^{t_l}) = g(\hat{\lambda})$.
- $A(\hat{\lambda}) = 0 \implies \delta^*(\hat{\lambda}) = 0 \implies p(\hat{\lambda}), \hat{\lambda}$ *are primal and dual optimal.*
  Pf.

$$g(\lambda^{t_l}) - g(\lambda^{t_l+1}) \le g(\lambda^{t_l}) - g(\lambda^{t_l+1}) \le A(\lambda^{t_l}) \le 0$$

$$\lim_l g(\lambda^{t_l}) - g(\lambda^{t_l+1}) = g(\hat{\lambda}) - g(\hat{\lambda}) = 0 \le \lim_l A(\lambda^{t_l}) = A(\hat{\lambda}) \le 0.$$

- $g^* = \sup_\lambda g(\lambda)$.

7.3 Proof of Theorem 4

Here we use some notation and intermediate results from the proof for Theorem 3.

We now show that $\lim_t p^t$ exists and is equal to $\hat{p} \equiv p(\hat{\lambda})$.

Since $\hat{p}$ and $\hat{\lambda}$ are primal and dual optimal, by the KKT conditions, we have that

$$\nabla_i g(\hat{\lambda}) = \sum_j \hat{p}_j a_{ij} - \epsilon \leq 0$$

for all i, and

$$\sum_i \hat{\lambda}_i \nabla_i g(\hat{\lambda}) = 0.$$

From the alternative expression for g given in 7 above and the KKT conditions just stated, we have that [14]

$$g(\hat{\lambda}) = -H(\hat{p}) + \sum_i \hat{\lambda}_i \nabla_i g(\hat{\lambda}) = -H(\hat{p})$$

and

$$g(\lambda^t) = -H(p^t; \hat{p}) + \sum_i \lambda_i^t \nabla_i g(\hat{\lambda}) \leq -H(p^t; \hat{p}).$$

Hence,

$$g(\lambda^t) - g(\hat{\lambda}) \leq -H(p^t; \hat{p}) + H(\hat{p}) = -\mathrm{KL}(\hat{p} \parallel p^t)$$

where $\mathrm{KL}(p \parallel q)$ is the Kullback-Liebler (KL) divergence between distributions p and q (i.e., $\mathrm{KL}(p \parallel q) \equiv \sum_j p_j \ln(p_j/q_j)$). Rearranging terms,

$$0 \leq \mathrm{KL}(\hat{p} \parallel p^t) \leq g(\hat{\lambda}) - g(\lambda^t).$$

Thus,

$$\lim_t \mathrm{KL}(\hat{p} \parallel p^t) = 0. \tag{8}$$

For any two probability distributions p and q, a well known inequality

$$\mathrm{KL}(p \parallel q) \geq (1/2)(\parallel p - q \parallel_1)^2 \tag{9}$$

establishes the relationship between the KL divergence and the L_1 norm (*ie*, for any vectors $x, y \in \mathbb{R}^n$, $\parallel x - y \parallel_1 \equiv \sum_{k=1}^n |x_k - y_k|$).

From equation 8 above, we have that for any $\iota > 0$, there exists a $T > 0$, such that for all $t \geq T$, $\mathrm{KL}(\hat{p} \parallel p^t) < \iota$. From this we get that, for all j,

$$|\hat{p}_j - p_j^t| \leq \sum_{j'} |\hat{p}_{j'} - p_{j'}^t| = \parallel \hat{p} - p^t \parallel_1 < \sqrt{2\iota}$$

where the last inequality (first from the right) holds as a result of inequality 9 above. This completes the proof that $\lim_t p^t = \hat{p}$.

[14] Another reason why $g(\hat{\lambda}) = -H(\hat{p})$ is that there is no duality gap.