Computing an Extensive-Form Correlated Equilibrium in Polynomial Time

Wan Huang and Bernhard von Stengel

Department of Mathematics, London School of Economics,
London WC2A 2AE, United Kingdom
wan.huang@gmail.com, stengel@nash.lse.ac.uk

Abstract. We present a polynomial-time algorithm for finding one *extensive form correlated equilibrium* (EFCE) for multiplayer extensive games with perfect recall. This the first such algorithm for an equilibrium notion for games of this generality. The EFCE concept has been defined by von Stengel and Forges [1]. Our algorithm extends the constructive existence proof and polynomial-time algorithm for finding a correlated equilibrium in succinctly representable games by Papadimitriou and Roughgarden [2,3]. We describe the set of EFCE with a polynomial number of consistency and incentive constraints, and exponentially many variables. The algorithm employs linear programming duality, the ellipsoid algorithm, and Markov chain steady state computations. We also sketch a possible interpretation of the variables in the dual system.

1 Introduction

Extensive games with perfect recall are a fundamental model of noncooperative game theory. They are game trees where players may have imperfect information about the game state, modeled by information sets [4]. The standard rationality assumption of *perfect recall* is a condition on the information sets that asserts that a player never forgets what he knew or did earlier.

The game tree, with its information sets, moves, chance probabilities, and payoffs, is a *succinct* representation of a game. The *strategic form* of the game is in general exponentially larger because because a pure strategy of a player is a tuple of moves, one for each information set, so there are exponentially many strategies per player; in the terminology of [2,3], this means the game is *not* of "polynomial type". Already for zero-sum two-player games, finding an equilibrium is therefore an interesting computational problem. It is solved by the *sequence form* [5], which is a strategic description of the same size as the game tree, and allows to solve huge two-person zero-sum games, for example of poker (see [6], also for earlier references related to the sequence form).

Finding a Nash equilibrium of an extensive game with any number of players is as difficult as for a game in strategic form. For the latter, a (more general) *correlated equilibrium* (CE) [7] can be found in polynomial time. Papadimitriou and Roughgarden [2,3] give a polynomial-time algorithm for succinctly representable games. Applying the ellipsoid method [8] to a linear program derived from the existence proof for CE due to [9] and [10], the method generates a polynomial-sized LP. The solution to that LP gives a distribution on product distributions that describes the desired CE for the

C. Papadimitriou and S. Zhang (Eds.): WINE 2008, LNCS 5385, pp. 506–513, 2008.

succintly represented game. However, this method cannot be applied to extensive games because they are not of polynomial type.

We present the *first polynomial-time algorithm for finding an equilibrium for general extensive games* that have any number of players and perfect recall. We have to adapt the concept of correlated equilibrium, but preserve the spirit of the extensive game in the sense that any "uncorrelated" such equilibrium is a Nash equilibrium (this is not the case for the *agent form*, where [2,3] can be applied, as discussed in [1]). We consider the new concept of *extensive form correlated equilibrium* (EFCE) [1]. A strategic-form CE can be considered as a device that selects a profile of pure strategies from a joint probability distribution and sends each player privately his pure strategy in that profile as a recommendation of what to play. In an EFCE, such a profile is also selected by the device before the game starts. However, the recommendations in an EFCE are "delayed": rather than telling the entire strategy in advance, each recommended *move* is given only when the player reaches the information set where she can make that move.

For two-player perfect recall extensive game without chance moves, [1] give a polynomial-size system that describes the set of all EFCE. Hence, any solution to that system can be found in polynomial time. However, already in two-player games with chance moves, it is not possible to give a polynomial-sized description of the set of all EFCE, unless P = NP [11,12,1].

In this paper, we use a description of the set of EFCE for general extensive games with any number of players with perfect recall, and possible chance moves. We describe the set of EFCE by a polynomial number of constraints, but with exponentially many variables, which allows us to extend the method of [2,3].

In an EFCE, unlike in a CE, a player has only to decide whether a recommended *move* is optimal, which involves a small number of comparisons with the other moves at the respective information set. In addition, a player who considers deviating from a recommended move has to look at additional future moves at previously unreached own information sets. We represent these by suitable incentive constraints with variables and constraints that, essentially, mimick "dynamic programming" in a single-player decision tree. The resulting system is small but somewhat involved, because one has to study carefully the propagation of payoffs for possible deviations through the game tree via dual variables.

Our blueprint, the proof by [2,3], uses the constructive existence proof of CE of [9,10], and employs linear programming duality, the ellipsoid algorithm, Markov chain steady state computations, as well as application specific methods for computing expectations over product distributions. We run the ellipsoid algorithm on the, more complicated, dual system. In contrast to the computation of a CE for a succinct strategic-form game [2,3], the dual system contains additional "consistency" information in the form of certain equalities for the dual variables. At each step, the ellipsoid algorithm finds a violated convex combination of the constraints of the dual system. These correspond, dually, to product distributions on moves at information sets. These do *not* represent steady states of a Markov chain, unlike in [2,3]. However, by assuming that the required distribution exists for preceding information sets, it can be considered as a steady state. For this reason, our algorithm uses Markov chain computation at each information set from the root of the game tree down to the information sets closest to the leaves of the game tree.

We give our construction first for for perfect-recall extensive games without chance moves, and then consider chance as an extra player who gets no payoff and never deviates. For reasons of space, we have to omit all proofs, and, unfortunately, examples, which are available in a long version of this paper.

2 Incentive Constraints

We use the notation of [1], except that a player is typically denoted by p and N is the set of all players, often omitted. The set of all information sets of player p is H_p. For $h \in H_p$, the set of moves or choices at h is C_h. A pure strategy s_p of player p is an element of $\prod_{h \in H_p} C_h$, and a strategy profile s is an element of $\prod_p \prod_{h \in H_p} C_h$.

In an EFCE, a player receives a *move* recommendation when reaching an information set, unlike in a CE where a player gets a recommended strategy at the beginning of the game. The player then compares the expected payoffs of moves at that information set and chooses the move with maximum expected payoff.

We first consider games without chance moves. Given an extensive game with perfect recall, is there a "sequence form" to compute one EFCE? For two-player games without chance moves, the answer is yes, as described in detail in [1]. The system of consistency and incentive constraints of sequences defines the set of EFCE. This holds because the condition of perfect recall imposes strong restrictions on the players' information sets, so that the recommended move at each information set can be generated uniquely. However, for games of more than two players, or with chance moves, the consistency constraints of the sequences on the marginal probabilities of moves that are correlated across information sets are only necessary conditions. For this reason, our system does not use the sequence form.

Therefore, instead of introducing an auxiliary variable $u(\sigma)$ to denote the expected payoff contribution of a sequence σ as in [1], we use $u(c)$ to denote the expected payoff contribution of the move c when the player follows his recommendations at all information sets he reaches. Before giving the expression of $u(c)$, the relations between moves and strategies need to be clarified.

Definition 1. *An information set $h \in H_p$* precedes *another information set $k \in H_q$ if and only if $p = q$ and there are nodes $u \in h$ and $v \in k$ such that u is on the path from the root to v. Furthermore, $h \in H_p$* immediately precedes *$k \in H_p$ when h precedes k and there is no information set $l \in H_p$ that succeeds h and precedes k.*

Unlike in [1], the relation "precedes" is only between two information sets of the same player.

Definition 2. *A move $c \in C_h$* agrees *with a strategy profile s if and only if the information set h is reached and the move c is chosen when players play according to s. A move $c \in C_h$ of player p* terminates *a strategy profile s_{-p} if and only if h is reached and no further information set of player p is reached if p plays the moves leading to h at all preceding information sets and plays c at h, and the other players play s_{-p}. An information set $h \in H_p$ is* reachable *by a strategy profile $s \in S$ if and only if the player p reaches h at a certain stage if all players choose the strategies in s.*

Let $a^p(s)$ be the payoff to player p if all players choose the strategies in s, and let $z(s)$ be the probability according to which the correlation device selects the strategy profile s. For any $c \in C_p$, the variable $u(c)$ is given by

$$u(c) = \sum_{s \in S\,:\,c \text{ agrees with } s} a^p(s)z(s). \tag{1}$$

Thus for any move c so that no further information set of player p is reached afterwards, $u(c)$ is the expected payoff to player p if he plays the recommended move c. The following lemma shows that $u(c)$ is the expected payoff contribution also for a move c that leads to further information sets of player p.

Lemma 3. *Given a move $c \in C_h$ and the set $\{k \in H_p : h \text{ precedes } k\}$ is not empty, we have*

$$u(c) = \sum_{\substack{s \in S\,:\,c \text{ agrees with } s,\\ c \text{ terminates } s_{-p}}} a^p(s)z(s) + \sum_{\substack{l \in H_p\,:\,h \text{ immediately}\\ \text{precedes } l \text{ via } c}} \sum_{c' \in C_l} u(c').$$

The expected payoff $u(c)$ when the player chooses the recommended move c is compared with the possible payoff when the player deviates from his recommendation. Given a move $c \in h$ and an information set k such that $k = h$ or k succeeds h, we use $v(k,c)$ to denote the optimal expected payoff at k given the player is recommended move c at h. It is the maximum of the payoffs for the possible moves $d \in C_k$, which may either directly give a payoff $a^p(s_d^k)$ when d terminates s_{-p} (where s_d^k is the strategy profile that specifies moves leading to k at information sets preceding k, and d at k, and the same moves as s at all other information sets), or are obtained from subsequent optimal payoffs at later information sets. This is expressed by the following inequalities:

$$v(k,c) \geq \sum_{\substack{s \in S\,:\,c \text{ agrees with } s,\\ d \text{ terminates } s_{-p}}} a^p(s_d^k)z(s) + \sum_{\substack{l \in H_p\,:\,k \text{ immediately}\\ \text{precedes } l \text{ via } d}} v(l,c), \quad d \in C_k. \tag{2}$$

These incentive constraints constraints are completed by

$$u(c) = v(h,c), \tag{3}$$

for any move $c \in C_h$. That is, given a recommended move c, the player does not gain by deviating from move c.

Theorem 4. *In a perfect-recall extensive game, a probability distribution z that fulfills for all players the incentive constraints (1),(2) and (3) defines an EFCE. The number of constraints that describe the set of EFCE is polynomial in the size of the game tree.*

3 Existence Proof

In the system describing the set of EFCE, (3) states that the expected payoff contribution of the recommended move must be optimal, as expressed by (1) and (2). One can obviously substitute $v(h,c)$ with $u(c)$ when $c \in C_h$ in (2). We rewrite these simplified constraints as matrix inequalities and consider the linear program

$$\text{maximize} \sum_{s\in S} z(s), \qquad \text{subject to} \quad Az + Bv \geq 0, \quad z \geq 0. \tag{4}$$

So the entries of A are either 0 or linear terms of $a(s)$ for certain s and the entries of B, for (2) and (3), are either 0, 1 or -1. The LP (4) is either trivial with the objective function being 0 or unbounded. When it is unbounded the normalized solution is an EFCE. Therefore by duality, to prove the existence of EFCE, it suffices to show that the dual of (4)

$$A^\top y \leq -1, \quad B^\top y = 0, \quad y \geq 0 \tag{5}$$

is always infeasible. We need the following lemma, analogous to [9,10,2].

Lemma 5. *If $y \geq 0$, $B^\top y = 0$, then there is a product distribution z so that $z^\top A^\top y = 0$.*

Here, $z^\top A^\top y$ is a convex combination of left sides of the constraints $A^\top y \leq -1$ in (5), and hence for every feasible $y \geq 0$, $B^\top y = 0$, it should evaluate to something negative. Thus this lemma shows that (5) is infeasible.

The proof of Lemma 5 uses the following lemma which has a long but straightforward proof. There is one dual variable $y^k_{c,d}$ for each information set k and move $d \in k$ and $c \in h$ where $h = k$ or h precedes k. To prove that given $y \geq 0$ and $B^\top y = 0$, there is a convex combination of components of $A^\top y$ equal to 0, we first show how a component of $A^\top y$ can be expressed in terms of the payoff a^p and the dual variable y.

Lemma 6. *Given a strategy profile $s \in S$ and $y \geq 0$ such that $B^\top y = 0$,*

$$(A^\top y)_s = \sum_p \sum_{k\in H_p} \sum_{\substack{h \in H_p \,:\, h=k \text{ or} \\ h \text{ precedes } k \text{ and } h \\ \text{is reachable by } s}} \sum_{d\in C_k} y^k_{c^h_s,d}[a^p(s^k) - a^p(s^k_d)] \tag{6}$$

where s^k is the strategy profile in S that specifies moves leading to k at information sets preceding k and the same moves as s at all other information sets (k may not be reached according to the moves s specifies at information sets of other players), and c^h_s is the move that s specifies at h.

The following lemma provides the main step to prove Lemma 5.

Lemma 7. *For any y such that $y \geq 0$ and $B^\top y = 0$, there is a product distribution $z = \prod_{p\in N} \prod_{k\in H_p} z^k$ such that for any information set k, the probability distribution z^k on the moves d at k satisfies*

$$z^k(d)[\alpha^k(d) + \alpha^k(\emptyset)] = \sum_{c\in C_k} z^k(c)\beta^k_1(c,d) + \beta^k_2(d) \tag{7}$$

where for any $c \in C_k$,

$$\alpha^k(c) = [\prod_{l \in H^k} z^l(c^l_k)] \sum_{d \in C_k} y^k_{c,d}\,, \qquad \alpha^k(\emptyset) = \sum_{\substack{h \in H_p : h \\ \text{precedes } k}} \prod_{l \in H^h} z^l(c^l_h) \sum_{c \in C_h} z^h(c) \sum_{d \in C_k} y^k_{c,d}\,,$$

$$\beta^k_1(c,d) = [\prod_{l \in H^k} z^l(c^l_k)] y^k_{c,d}\,, \qquad \beta^k_2(d) = \sum_{\substack{h \in H_p : h \\ \text{precedes } k}} \prod_{l \in H^h} z^l(c^l_k) \sum_{c \in C_h} z^h(c) y^k_{c,d}\,.$$

Here H^k is the set of information sets l (of the same player as k) that precede k, and c^l_k is the unique move at l that leads to k.

Nau and McCardle [10] discussed "joint coherence" in noncooperative games, and thus gave a possible interpretation of the variables involved in both the primal and the dual system. Myerson [13] used this interpretation to obtain further properties of CE. For EFCE, we consider a certain *move transition matrix* T^k in order to prove Lemma 7, for each information set k. Any such move transition matrix for information set k can be interpreted as a random deviation plan for the player who will make a move at that information set. Each number $y^k_{c,d}$ in (6), where $d \in C_k$, represents the trend that player would deviate to the move d when c is recommended at this information set or some earlier stage of the game (and the player ignores any recommendation after getting c). More precisely, the trend that the player ignores a recommendation at some earlier stage and chooses move d at information set k is $\beta^k_2(d)$, and the trend that the player chooses d at information set k is $\sum_{c \in C_k} z^k(c)\beta^k_1(c,d) + \beta^k_2(d)$. On the other hand, the trend that the player would be getting recommended move d (he or she may ignore it) is $z^k(d)\alpha^k(c^k_s) + z^k(d)\alpha^k(\emptyset)$. One can then show that the deviation plan does not change the distribution on the player's actions at the information set k.

The preceding lemmas require several pages of proofs in full detail. Lemma 5 then implies the existence of EFCE.

Theorem 8. *Every game of extensive form without chance moves has an EFCE.*

4 Algorithm for Games without and with Chance Moves

To find an EFCE in polynomial time, we follow [2,3] and apply the ellipsoid algorithm to the dual (5) of the system (4) that characterizes the set of EFCE. The LP (4) has polynomially many constraints and exponentially many variables. Thus for the dual (5) the opposite holds, which makes it suitable for the ellipsoid algorithm.

In each iteration of the ellipsoid algorithm, an extra step is needed to maintain the candidate solution y_i to satisfy the consistency constraints $B^\top y_i = 0$. At the initial iteration, for the system $B^\top y = 0$, $y = 0$, let $\bar{y}$ be the free variables in y, and $y = \bar{B}\bar{y}$. Thus the system (5) is equivalent to

$$A^\top \bar{B}\bar{y} \leq -1, \qquad \bar{y} \geq 0. \tag{8}$$

We apply the ellipsoid algorithm to the system (8).

Let $\bar{y}_0 = 0$ be the candidate for the initial iteration. Thus every constraint $A^\top \bar{B}\bar{y} \leq -1$ is violated. Any product distribution z satisfies $z^\top A^\top y = 0$. Choose any product distribution z_0, and a violated inequality $z_0^\top A^\top y \leq -1$. At each iteration, the candidate $\bar{y}_{i-1}$ is replaced by $\bar{y}_i$. Let $y_i = \bar{B}\bar{y}_i$. By Lemma 5, a product distribution z_i such that $z_i^\top A^\top y = 0$ can be found. Thus the inequality $(z_i^\top A^\top)\bar{B}\bar{y} \leq -1$ is violated. We proceed to the next step.

Since we know that (5) is infeasible, the algorithm will end up with recognizing the system as infeasible after polynomially many iterations. Thus when the algorithm halts, we have polynomially many candidate solutions y_i and for each y_i a corresponding product distribution z_i.

We now claim that a convex combination, denoted $Z^\top \xi$, of these product distributions can be found in polynomial time, such that the system $AZ^\top \xi + Bv \geq 0$, $\xi \geq 0$ is unbounded. When the ellipsoid algorithm is applied to (8), in each iteration the inequality $(z_i^\top A^\top \bar{B})\bar{y} \leq -1$ is violated by $\bar{y}_i$. Let Z be the matrix where each row i is the product distribution z_i found by the ellipsoid algorithm. We consider the system of linear inequalities

$$[ZA^\top \bar{B}]\bar{y} \leq -1, \qquad \bar{y} \geq 0\,. \tag{9}$$

Clearly, the number of variables of (9) is equal to that of (8), and is polynomial in the size of the game tree. Thus the ellipsoid algorithm is appropriate to (9) too. Apply it to (9). Let the initial candidate solution be $\bar{y}_0 = 0$. In each iteration i, the ith constraint of (9) $(z_i^\top A^\top \bar{B})\bar{y} \leq -1$ is violated by the ith candidate solution y_i. Thus the algorithm will determine that (9) is infeasible too. That is,

$$[ZA^\top]y \leq -1, \qquad y = \bar{B}\bar{y}, \qquad \bar{y} \geq 0$$

or equivalently

$$[ZA^\top]y \leq -1, \qquad B^\top y = 0, \qquad y \geq 0$$

is infeasible. The dual problem

$$\text{maximize} \sum_i (\xi_A)_i \qquad \text{subject to} \quad [AZ^\top]\xi_A + B\xi_B \geq 0, \quad \xi_A \geq 0 \tag{10}$$

is unbounded. Here (ξ_A, ξ_B) is a partition of the variable vector ξ.

For any feasible solution ξ of (10), ξ_A after normalization is a probability distribution on the set of strategy profiles. The product $Z^\top \xi_A$ is a convex combination of the rows of $Z^\top$, which are the product distributions that are computed at all the iterations of the ellipsoid algorithm. Thus the nonnegative constraints $\xi_A \geq 0$ are satisfied if and only if $Z^\top \xi_A \geq 0$. Let $z = Z^\top \xi_A$, and $v = \xi_B$. The system (10) becomes

$$\text{maximize} \sum_s z(s) \qquad \text{subject to} \quad Az + Bv \geq 0, \quad z \geq 0$$

which is the system that characterizes an EFCE. Therefore, $(z, v) = (Z^\top \xi_A, \xi_B)$ is a nontrivial solution to (5) when ξ is a nontrivial solution to (10). Furthermore, $z = Z^\top \xi_A$ is the desired EFCE.

So far, all the arguments and inductions are based on the assumption that there are no chance moves in the game. With chance moves, the system (4) is no longer appropriate, because a move may "agree" with more than one strategy profile. However, the impact of the chance moves on the reachability of an information set can be expressed by considering chance as an extra player 0, without any incentive constraints. The chance moves become part of a strategy profile, but their probabilities in the construction of product distributions will be constants rather than variables, with minor modifications of the algorithm for games without chance moves. We obtain the following result.

Theorem 9. *Every multi-player, perfect-recall extensive game, which may have chance moves, has an EFCE, which can be computed in polynomial time.*

The EFCE concept is crucial to limit the number of incentive constraints. It is an *open question* if one can find one (strategic-form) CE for extensive games, even with only two players, in polynomial time as well. Because of the exponential number of strategies for each player, it is not even clear if such a CE has a polynomial-sized description and certification of the equilibrium property (analogous to the NP property for a decision problem).

References

1. von Stengel, B., Forges, F.: Extensive form correlated equilibrium: Definition and computational complexity. Math. Oper. Res. (in press, 2008)
2. Papadimitriou, C.H.: Computing correlated equilibria in multi-player games. In: Proc. 37rd Annual ACM Symposium on Theory of Computing (STOC), pp. 49–56 (2005)
3. Papadimitriou, C.H., Roughgarden, T.: Computing correlated equilibria in multi-player games. J. ACM 55, 14 (2008)
4. Kuhn, H.W.: Extensive games and the problem of information. In: Contributions to the theory of games. Annals of Mathematics Studies 28, vol. 2, pp. 193–216. Princeton Univ. Press, Princeton (1953)
5. von Stengel, B.: Efficient computation of behavior strategies. Games Econom. Behav. 14, 220–246 (1996)
6. Gilpin, A., Sandholm, T.: Lossless abstraction of imperfect information games. J. ACM 54, 25 (2007)
7. Aumann, R.J.: Subjectivity and correlation in randomized strategies. J. Math. Econom. 1, 67–96 (1974)
8. Khachiyan, L.G.: A polynomial algorithm in linear programming. Soviet Mathematics 20, 191–194 (1979); (transl. of Dokl. Akad. Nauk SSSR 244, 1093–1096)
9. Hart, S., Schmeidler, D.: Existence of correlated equilibria. Math. Oper. Res. 14, 18–25 (1989)
10. Nau, R.F., McCardle, K.F.: Coherent behavior in noncooperative games. J. Econom. Theory 50, 424–444 (1990)
11. Chu, F., Halpern, J.: On the NP-completeness of finding an optimal strategy in games with common payoffs. Internat. J. Game Theory 30, 99–106 (2001)
12. von Stengel, B.: Computational complexity of correlated equilibria for extensive games. Research Report LSE-CDAM-2001-03, London School of Economics (2001)
13. Myerson, R.B.: Dual reduction and elementary games. Games Econom. Behav. 21, 183–202 (1997)