2021-Farina

Connecting Optimal Ex-Ante Collusion in Teams to Extensive-Form Correlation: Faster Algorithms and Positive Complexity

---

- **Date**: 2021
- **Link**
- **Authors:**
  - Farina, Gabriele
  - Celli, Andrea
  - Gatti, Nicola
  - Sandholm, Tuomas
- **Cites:**
- **Cited by:**
- **Keywords:** #extensive-form-game #multi-player-games #team-games #team-maxim-equilibrium #team-maxim-equilibrium-w-coordination-device #linear-program #extensive-form-correlation-plans #column-generation
- **Collections:**
- **Status:** #completed

---

0. Abstract

- **Problem:** find an optimal strategy for a team of players facing an opponent in an imperfect-information zero-sum extensive-form game.

- Coordination between team members is only allowed before the game.

> KEY IDEA - the best the team can do is sample a profile of potentially randomized strategies (one per player) from a joint (i.e. teammates' strategies can be correlated) probability distribution at the beginning of the game.

- In this paper:
  1. Compute optimal distributions by drawing a connection to extensive-form correlation.

  2. An algorithm that allows for the capping of the number of profiles employed in the solution.
     - Often a handful of well-chosen profiles is enough for optimal team utility.

  3. Efficient column-generation algorithm for finding an optimal distribution for the team.
     - 1000x faster than previous methods when tested across 3 benchmark games.
     - Solves previously unsolved games as well.

1 - Introduction

- Strategic multi-player games are not well understood in theory or practice.
  - Computing strong strategies is generally hard in the worst case.

- *Adversarial team games* - games in which a team of coordinating players face an opponent.
  - This paper focuses on the variant:
    - Two-player team facing a third player.
    - With no communication during the game.
    - Notable related papers: 1997-von-Stengel & 2018-Celli

  - Practical examples:
    - How should two players collude against a third player in poker?
    - Two defenders in Bridge playing against the declarer?

- Solving adversarial team games, two models:
  - Solutions are an optimal profile of (potentially mixed, i.e. randomized) strategies where each team member is assigned a strategy.

  - *Team-maxim equilibrium* (TME) - solution that yields maximum expected sum of utilities for the team's players against a rational (best responding) agent.

  - *Team-maxim equilibrium with coordination device* (TMECor) - solution that allows for team members to sample their strategies from a joint (correlated) distribution, called *ex-ante coordination*.
    - Two major advantages:
      1. Team's expected utility is greater than or equal to the EU provided by TME.
      2. Makes the optimal team strategy problem convex.
  - See 2018-Farina for more detail on TMECor.

- Paper's contributions:
  1. Proposes a problem formulation for finding a TMECor strategy via connection to extensive-form strategy polytopes, using the novel notion of *semi-randomized correlation plans*.

2. Introduces an algorithm for computing a TMECor strategy with a fixed number of semi-randomized correlation plan pairs.

3. Develops a column-generation algorithm for finding a TMECor strategy, using a best-response oracle for computing joint team best-response strategies.

2 - Preliminaries

- Let T_1 and T_2 represent the two players on the team, and O represent their opponent.
  Player c is the chance player, corresponding to the exogenous stochasticity in the environment.

- Extensive form game setup:
  - Each node v belongs to exactly one player.
  - Each edge leaving v represents a possible action at that node.
  - Leaf nodes represent the end state of the game.
    - Z is the set of leaf nodes.
    - Each $z \in Z$ has an associated tuple of payoffs, one for each player.
  - The product of the probabilities of all actions of c on the path from the root of the game to leaf z is denoted by $p_c(z)$.

- Private information:
  - *Information set* (infoset) - represents private information in the game.
    - Each player $i \in \{T_1, T_2, O\}$, has a collection $\mathcal{I}_i$ where $I \in \mathcal{I}_i$ is a set of nodes belonging to the player that are indistinguishable to the player, given what they've observed.

  - Naturally, states in the same information set I, must have the same set of actions A_I .

  - For two players i and j, two of their infosets $I_i \in \mathcal{I}_i$, $I_j \in \mathcal{I}_j$ are *connected*, $I_i \leftrightarrow I_j$, if there exists $v \in I_j$ and $w \in I_j$ such that the path from the root to v passes through w or vice versa.

- Sequences:
  - *Sequence* - infoset-action pair for a player.

  - *Sequences* Σ_i - set of all possible sequences for Player i - $\Sigma_i := \{(I, a) : I \in \mathcal{I}_i , a \in A_I\} \cup \{\varnothing\}$.

  - *Parent sequence* $\sigma(v)$ - the last sequence for Player i encountered on the path from the root to the node v.
    - Note: because agents are assumed to have perfect recall, all nodes in an infoset must share the same parent, $\sigma(I) \in \Sigma_i$.

  - *Relevant sequences* $\sigma_i \bowtie \sigma_j$ - a pair of sequences for two players $\sigma_i \in \Sigma_i$, $\sigma_j \in \Sigma_j$ where the corresponding infosets I_i , I_j are connected.
    - i.e. the action a_i by Player i at infoset I_i could potentially lead to Player j taking the action a_j at infoset I_j , or vice versa.

- Reduced-normal-form plans:
  - *Reduced-normal-form plan* π_i - Player i defines a choice of action for every information set $I \in \mathcal{I}_i$ that is still reachable based on the actions in π_i . The set of all plans for the player is denoted Π_i.
    - Note: I *think* these are deterministic plans.

  - Subsets of Π_i :
    - Let $\Pi_i(I)$ be the subset of Π_i that prescribe all actions for Player i on the path from the root to the information set I.
    - Let $\Pi_i(\sigma)$ be the subset of $\Pi_i(I)$ where Player i plays action a at I.
    - Let $\Pi_i(z)$ be the set of plans where Player i plays so as to reach z .

- Sequence-form strategy:
  - *Sequence-form strategy* - given a normal-form strategy $\mu \in \Delta(\Pi_i)$ for Player i where $\Delta(\Pi_i)$ is the probability simplex over the set of reduced-normal-form plans.

    Note: I *think* this defines a probability distribution over the deterministic plans, according to the likelihood of the player selecting the given plan.

    The sequence-form strategy induced by μ and indexed on the sequence σ is defined as

    $$y[\sigma] := \sum_{\pi \in \Pi_i(\sigma)} \mu(\pi)$$

    - $\mu(\pi)$ gives the probability we are following the plan π.
    - We sum over all plans that include σ - $\Pi_i(\sigma)$.
    - This yields the probability of the player performing the sequence σ - $y[\sigma]$.

  - The set of sequence-form strategies that can be induced as μ varies over $\Delta(\Pi_i)$ is denoted by $\mathcal{Y}_i$.
    - $\mathcal{Y}_i$ is a convex polytope.

- TMECor as a Bilinear Saddle-Point Problem:
  - *TMECor strategy* - a probability distribution μ_T over the set of randomized strategy profiles $\mathcal{Y}_{T1} \times \mathcal{Y}_{T2}$ that guarantees maximum expected utility for the team against the best responding opponent O.

  - Note: rather than defining the strategy as a distribution over profiles of randomized strategies ($\mathcal{Y}_{T1} \times \mathcal{Y}_{T2}$), we can write it equivalently as a distribution over the deterministic strategies $\Pi_{T1} \times \Pi_{T2}$.
    - $\Pi_{T1} \times \Pi_{T2}$ has the benefit of being finite.

  - Expected team utility:

    $$u_T(\mu_T, \mu_O) := \sum_{z \in Z} \hat{u}_T \left(\sum_{\pi_{T1} \in \Pi_{T1}(z),\, \pi_{T2} \in \Pi_{T2}(z)} \mu_T(\pi_{T1}, \pi_{T2}) \right) \left(\sum_{\pi \in \Pi_O(z)} \mu_O(\pi) \right)$$

- Where:
  - $\hat{u}_T(z) := \left(u_{T1}(z) + u_{T2}(z) \right) p_c(z)$ is the expected team utility at the leaf node z .
  - $\mu_T \in \Delta(\, \Pi_{T1} \times \Pi_{T2} \,)$, $\mu_O \in \Delta(\, \Pi_O \,)$ are the probabilities of the team playing by the given strategy profile and the adversary playing by the given plan.

- This computation amounts to multiplying the combined utility of the teammates at the leaf node times the probability of reaching the leaf node, then summing over all leaf nodes to get the total team utility.
  - The probability of reaching a leaf node is equal to the product of probabilities of following each edge in the tree on the path from the root to the leaf.

- *Team-maxmin equilibrium with coordination device* (TMECor) - in the zero-sum setting, it is the Nash equilibrium that is computed by finding a solution to the optimization problem:

$$\text{argmax}_{\mu_T \in \Delta(\Pi_{T1} \times \Pi_{T2})} \min_{\mu_O \in \Delta(\Pi_O)} \mu_T(\, \mu_T,\, \mu_O)$$

3 - A New Formulation of TMECor Based on Extensive-Form Correlation Plans

- Propose a different representation of μ_T where it is represented as a vector with only a polynomial number of components.
  - The proposed representation is inspired by *realization form* from 2018-Farina.

- Disadvantages of the proposed representation:
  - Number of components scales as the product of the number of sequences increases for the two players which is significantly greater than the number of leaves.

- Advantages of the proposed representation:
  1. Allows for faster best response computation, shown empirically.
  2. Allows for TMECor computation in polynomial time, for some classes of games such as Goofspiel.

3.1 - Extensive-Form Correlation Plans

- *Extensive-form correlation plans* - maps the correlated distribution of play μ_T to the vector ξ_T indexed over pairs of related sequences, $(\sigma_{T1},\, \sigma_{T2}) \in \Sigma_{T1} \bowtie \Sigma_{T2}$:

$$\xi_T[(\sigma_{T1},\, \sigma_{T2})] := \sum_{\substack{\pi_{T1} \in \Pi_{T1}(\sigma_{T1}) \\ \pi_{T2} \in \Pi_{T2}(\sigma_{T2})}} \mu_T[(\pi_{T1},\, \pi_{T2})]$$

  - $\xi_T[(\sigma_{T1},\, \sigma_{T2})]$ gives the probability of $T1$ playing the sequence σ_{T1} and $T2$ playing the sequence σ_{T2} .

- We only care about related sequences because, intuitively, there can be no coordination between unrelated sequences. Our likelihood of performing one sequence can not affect the likelihood of another sequence, if the first sequence can not lead to the second.

- This has the benefit of significantly reducing the size of ξ_T .

- This concept was first introduced by 2008-von-Stengel.

- *Polytope of correlation plans* Ξ_T - set of all valid ξ_T, is a convex polytope.

3.2 - Computing a TMECor using Correlation Plans

- Let - $\xi_T[\sigma_{T1}(z),\, \sigma_{T2}(z)] = \rho_T[z]$.

- Rewriting the TMECor problem using correlation plans:

$$\text{argmax}_{\;\xi_T \,\in\, \Xi_T} \min_{y_O \in \mathcal{Y}_O} \sum_{z \,\in\, Z} \hat{u}_T(z)\, \xi_T\left[\sigma_{T1}(z),\, \sigma_{T2}(z)\right]\, y\left[\sigma_O(z)\right]$$

- *Proposition 1* - By dualizing the inner linear minimization problem over y_O, we get an extensive-form correlation plan ξ_T is a TMECor if and only if it is a solution to the linear program:

  $\text{argmax}_{\xi_T} v_\varnothing$, subject to:

  1. $$v_I - \sum_{\substack{I' \in \mathcal{I}_O \\ \sigma_O(I')=(I,a)}} v_{I'} \;\le\; \sum_{\substack{z \in Z \\ \sigma_O(z)=(I,a)}} \hat{u}_T(z)\, \xi_T\left[\sigma_{T1}(z),\, \sigma_{T2}(z)\right] \quad \forall\, (I,a) \in \Sigma_O/\{\varnothing\}$$

  2. $$v_\varnothing - \sum_{\substack{I' \in \mathcal{I}_O \\ \sigma_O(I')=\varnothing}} v_{I'} \;\le\; \sum_{\substack{z \in Z \\ \sigma_O(z)=\varnothing}} \hat{u}_T(z)\, \xi_T\left[\sigma_{T1}(z),\, \sigma_{T2}(z)\right]$$

  3. $v_\varnothing$ free, v_I free $\forall I \in \mathcal{I}_O$.
  4. $\xi_T \in \Xi_T$.

- *Note to self: I need to learn more about linear programs.*
  I do not understand this proposition fully.

- Prop 1 shows that a TMECor can be found as the solution to a linear program with a polynomial number of variables.

- Prop 1 implies TMECor can be found in polynomial time whenever Ξ_T can be represented as the intersection of a set of polynomially many linear constraints (See Section 4).

4 - Semi-Randomized Correlation Plans and the Structure of Ξ_T

- The set of linear constraints that define Ξ_T is not known in general, and potentially exponentially large.

- Therefore, alternative characterizations of the set are needed in order to solve the LP in Prop 1.

4.1 - Containment in the von Stengel-Forges Polytope

- Fact about Ξ_T first proposed in 2008-von-Stengel.

- *Definition 1* - *von Stengel-Forges polytope* of the team $\mathcal{V}_T$ is the polytope of all vectors $\xi \in \mathbb{R}_{\geq 0}^{|\Sigma_{T1} \triangleright\triangleleft \Sigma_{T2}|}$ indexed over relevant sequence pairs that satisfy the following (polynomially-sized set) of linear constraints:
  1. $\xi[\varnothing, \varnothing] = 1$
  2. $\sum_{a_{T1} \in A_{I_{T1}}} \xi\Big[(I_{T1}, a_{T1}), \sigma_{T2}\Big] = \xi\Big[\sigma(I_{T1}), \sigma_{T2}\Big] \quad \forall\, I_{T1} \triangleright\triangleleft \sigma_{T2}$
  3. $\sum_{a_{T2} \in A_{I_{T2}}} \xi\Big[\sigma_{T1}, (I_{T2}, a_{T2})\Big] = \xi\Big[\sigma_{T1}, \sigma(I_{T2})\Big] \quad \forall\, \sigma_{T1} \triangleright\triangleleft I_{T2}$

- These are probability mass conservation constraints.
  - The sum over the probability of taking all actions at a node is equal to the probability of reaching that node.

- From these constraints we get Prop 2.

- *Proposition 2* - The set of extensive-form correlation plans is a subset of the von Stengel-Forges polytope, $\Xi_T \subseteq \mathcal{V}_T$.

4.2 - Triangle-Freeness and Polynomial-Time Computation of TMECor

- Fact about Ξ_T first proposed in 2020-Farina.

- *Definition 2* - the interaction between team members $T1, T2$ is *triangle-free* if, for any choice of distinct information sets that share a parent sequence: $I_1,\, I_2 \in \mathcal{I}_{T1}$ with $\sigma_{T1}(I_1) = \sigma_{T1}(I_2)$ and likewise $J_1,\, J_2 \in \mathcal{I}_{T2}$ with $\sigma_{T2}(J_1) = \sigma_{T2}(J_2)$, it is never the case that $(I_1 \leftrightarrow J_1) \wedge (I_2 \leftrightarrow J_2) \wedge (I_1 \leftrightarrow J_2)$.

- Team interactions are only triangle-free if all chance outcomes are public.

- The triangle-free condition can be checked in polynomial time.
  Iterate over all infoset quadruples.

- Consequence of triangle-free infoset structure among team members:
  - Then $\Xi_T = \mathcal{V}_T$.
  - The von Stengel-Forges constraints (Def 1) can be substituted in for constraint 4 in Prop 1, this yields a **TMECor computation in polynomial time!**

- See Table 1c for the proof that Goofspiel is triangle-free.

4.3 - Semi-Randomized Correlation Plans

- New result about the structure of Ξ_T .

- *Semi-randomized correlation plans* - subset of the von Stengel-Forges polytope where one of the players on the team plays a deterministic strategy, while the other player plays a randomized strategy.
  - Defined for each team member:

$$\Xi_{T1}^* = \Big\{\xi \in \mathcal{V}_T \;:\; \xi[\varnothing, \sigma_{T2}] \in \{0, 1\} \quad \forall \sigma_{T2} \in \Sigma_{T2}\Big\}$$

$$\Xi_{T2}^* = \Big\{\xi \in \mathcal{V}_T \;:\; \xi[\sigma_{T2}, \varnothing] \in \{0, 1\} \quad \forall \sigma_{T1} \in \Sigma_{T1}\Big\}$$

- Advantage - any point $\xi \in \Xi_i^*$ can be expressed using real and binary variables.
  From this we can write Prop 3.

- *Proposition 3* - Ξ_T is the convex hull of the set Ξ_{T1}^* , or equivalently Ξ_{T2}^* .

$$\Xi_T = \operatorname{co} \Xi_{T1}^* = \operatorname{co} \Xi_{T2}^* = \operatorname{co}\big(\Xi_{T1}^* \cup \Xi_{T2}^*\big)$$

  - *Note:* If I'm interpreting this correctly, Ξ_T is the set of probabilities over teammates both performing a given sequence pair. Ξ_{T1}^* is the set of probabilities where $T1$ takes deterministic actions. $\operatorname{co} \Xi_{T1}^*$ is a linear combination of deterministic actions, making the weights a probability distribution over actions. Therefore, $\Xi_T = \operatorname{co} \Xi_{T1}^*$. And by symmetry, the same is true for $T2$.

- This is reminiscent of *auxiliary game* construction in 2018-Farina with a different setting.

5 - Computing TMECor with a Small Support of Semi-Randomized Plans of Fixed Size

- Given Prop 3, we can replace Constraint 4 in Prop 1 with the constraint that ξ_T be a convex combination of elements from Ξ_{T1}^* and Ξ_{T2}^*.

  - Constraint definition:

$$\xi_T = \sum_{i=1}^{n} \lambda^{(i)} \xi_T^{(i)}$$

  where:

- $\xi_T^{(1)}, \ldots, \xi_T^{(n)} \in \Xi_{T1}^* \cup \Xi_{T2}^*$
- Convex combination coefficients: $\lambda^{(1)}, \ldots, \lambda^{(n)}$.
- n - number of semi-randomized correlation plans that can make-up the strategy.

- This yields the mixed integer LP:
$\underset{\xi_T^{(1)}, \ldots, \xi_T^{(n)}, \lambda^{(1)}, \ldots, \lambda^{(n)}}{\text{argmax}} \varnothing$ subject to:

  - Constraints 1, 2, 3 in Prop 1.

  4. $\xi_T = \sum_{i=1}^n \lambda^{(i)} \xi_T^{(i)}$

  5. $\xi_T^{(1)} \in \Xi_{T1}^*, \xi_T^{(2)} \in \Xi_{T2}^*, \ldots$
     - Note - we alternate which player's turn it is to play a deterministic strategy, this yields better results in practice.

  6. $\sum_{i=1}^n \lambda^{(i)} = 1, \quad \lambda^{(i)} \geq 0 \ \forall i \in \{1, \ldots, n\}$

- Choice of n :
  - The higher the n the better the solution.
  - For an optimal solution $n \geq |\Sigma_1 \bowtie \Sigma_2| + 1$ (by Caratheodory's theorem).
  - Empirically, the choice of n can be significantly less than optimal and still yield near-optimal results. Even one or two base randomized strategies, if well chosen, can produce optimal coordination.

6 - A Fast Column Generation Approach

- Use column generation as a scalable approach to solving the constraints in Prop 1.
  - Column generation was first introduced in 1958-Ford.

- Column generation algorithm:
  - *Seeding* phase - pick a set S of correlation plans $\xi_T^{(1)}, \ldots, \xi_T^{(m)} \in \Xi_T$.

  - Loop:
    - For each $i \in \{1, \ldots, |S|\}$, compute:

$$\beta^{(i)}(\sigma_O) := \sum_{\substack{z \in Z \\ \sigma_O(z) = \sigma_O}} \hat{u}_T(z)\, \xi_T^{(i)}\big[\sigma_{T1}(z), \sigma_{T2}(z)\big] \quad \forall\, \sigma_O \in \Sigma_O$$

      - This is the team expected utility summed over the leaf-nodes who's parent sequence is σ_O . Which is to say, this is the team's expected utility if the opponent were to play σ_O .

    - Then, solve the following *master* LP:

    $\underset{\lambda^{(1)}, \ldots, \lambda^{(|S|)}}{\text{argmax}} \varnothing$, subject to:

    1. $$v_I - \sum_{\substack{I' \in \mathcal{I}_O \\ \sigma_O(I') = \sigma_O}} v_{I'} - \sum_{i=1}^{|S|} \beta^{(i)}(\sigma_O) \lambda^{(i)} \leq 0 \quad \forall \ \sigma_O \in \Sigma_O / \{\varnothing\}$$

    2. $$v_\varnothing - \sum_{\substack{I' \in \mathcal{I}_O \\ \sigma_O(I') = \varnothing}} v_{I'} - \sum_{i=1}^{|S|} \beta^{(i)}(\varnothing) \lambda^{(i)} \leq 0$$

    3. $\sum_{i=1}^{|S|} \lambda^{(i)} = 1$
    4. $\lambda^{(i)} \geq 0 \quad \forall i \in \{1, \ldots, |S|\}$
    5. $v_\varnothing$ free, v_I free $\forall I \in \mathcal{I}_O$.

      - *Note:* this is a similar LP to the one in Section 5 however the correlation plans are given here.

    - *Pricing problem* - given the solution to the master LP, generate a new element $\xi_T^{|S|+1}$ to add to S so as to increase the team utility on the next iteration.

  - Repeat the loop until termination.

6.1 - The Pricing Problem

- We must find a correlation plan $\hat{\xi}_T \in \Xi_T$ such that it would lead to the maximum gradient of the objective, i.e. maximum *reduced cost*.

- This can be found by exploiting the theory of LP duality:
  - Let γ be the $|\Sigma_O|$-dimensional vector of dual variables corresponding to Constraints 1 and 2 of the master LP.

  - Then, the reduced cost of any candidate $\hat{\xi}_T$:

$$c(\hat{\xi}_T) := -\gamma + \sum_{z \in Z} \hat{u}_T(z)\, \hat{\xi}_T\big[\sigma_{T1}(z), \sigma_{T2}(z)\big]\, \gamma\big[\sigma_O(z)\big]$$

- We know from Section 4 that the constraints for Ξ_T are not known; therefore, it is impossible to solve for $\max_{\hat{\xi}_T \in \Xi_T} c(\hat{\xi}_T)$.

- However, we know that from Section 4.3, $\Xi_T = \text{co}\, \Xi_{T1}^*$.

- Ξ_{T1}^* is a well-defined mixed integer LP (MIP).

- We can then solve:

$$\max_{\hat{\xi}_T \in \Xi^*_{T1}} c(\hat{\xi}_T)$$

- When the objective value of the pricing problem is non-positive, then there is no variable we can add to S to improve its value. And, we terminate.

6.2 - Implementation Details

- Speeding up the algorithm:
  - **Seeding phase:**
    - The better the set of initial correlation plans are, the faster the loop converges to optimal.
    - Run self-play CFR+ players to generate pure normal-form strategies that can be used to compute correlation plans.
  - **Linear relaxation:**
    - Try to solve the linear relaxation of the pricing problem, $\operatorname{argmax}_{\hat{\xi}_T \in \mathcal{V}_T} c(\hat{\xi}_T)$ before solving the full MIP.
    - Oftentimes, this yields a semi-randomized correlation plan, circumventing the need to solve the harder problem.
  - **Solution pools:**
    - Commercial MIP solvers output the optimal solution along with a set of sub-optimal solutions that were found during computation.
    - These can be used without loss of optimality and could speed up convergence.
  - **Termination:**
    - Stop when the reduced costs generated by the pricing problem are within numeric tolerance, 10^{-6}.
  - **Dual values**:
    - We do not need to track dual values, modern solvers take care of this.

7 - Experimental Evaluation

- See the tables in the paper for results.