Learning Extensive-Form Perfect Equilibria in Two-Player Zero-Sum Sequential Games

Martino Bernasconi
Politecnico di Milano

Alberto Marchesi
Politecnico di Milano

Francesco Trovò
Politecnico di Milano

Abstract

Designing efficient algorithms for computing *refinements* of the *Nash equilibrium* (NE) in two-player zero-sum sequential games is of paramount importance, since the NE may prescribe sub-optimal actions off the equilibrium path. The *extensive-form perfect equilibrium* (EFPE) amends such a weakness by accounting for the possibility that players may make mistakes. This is crucial in the real world, which involves humans with bounded rationality, and it is also key in boosting superhuman agents for games like Poker. Nevertheless, there are only few algorithms for computing NE refinements, which either lack convergence guarantees to exact equilibria or do *not* scale to large games. We provide the first *efficient iterative* algorithm that *provably converges* to an EFPE in two-player zero-sum sequential games. Our algorithm works by tracking a sequence of equilibria of regularized-perturbed games, by using a procedure that is specifically tailored to converge last iterate to such equilibria. The procedure can be implemented efficiently by visiting the game tree, making our method computationally appealing. We also empirically evaluate our algorithm, showing that its strategies are much more robust to players' mistakes than those of state-of-the-art algorithms.

1 INTRODUCTION

Computing the *Nash equilibria* (NEs) (Nash, 1951) of *two-player zero-sum* sequential (*i.e.*, extensive-form) games with *imperfect information* has been one of the flagship computational challenges of artificial intelligence since several years. The latest advances in the field of equilibrium computation have lead to the development of *superhuman* agents that are capable of beating top human professionals in several games, such as, *e.g.*, Go (Silver et al., 2016), heads-up no-limit Texas hold'em Poker (Brown and Sandholm, 2018, 2019), and Diplomacy (Bakhtin et al., 2022).

The NE is a solution concept that prescribes each player to play optimally under the assumption that the opponents are perfectly rational and do the same. However, this assumption is oftentimes unreasonable, especially when computed equilibria are deployed in the real world, where artificial agents usually face human opponents that naturally have bounded rationality. Indeed, it is well known that the NE has several weaknesses when played against opponents who may make mistakes since they are *not* perfectly rational. In particular, an NE may prescribe to perform sub-optimal actions at decision points (a.k.a. *information sets*) that are never reached assuming that the players play equilibrium strategies.

Over the last decades, game theorists introduced several *refinements* of the NE notion, in order to amend its weaknesses off the equilibrium path. Among them, the most studied and recognized one is the *extensive-form perfect equilibrium* (EFPE) originally introduced by Selten (1975). The EFPE is based on the idea of *trembling-hand perfection*, whose rationale is to let the players reasoning about the possibility that both themselves and their opponents may "tremble" in future, by playing sub-optimal actions with small probability at information sets that will be reached later in the game. An EFPE is defined as a limit point of a sequence of equilibria that is obtained by letting the "magnitude of trembles" (*i.e.*, the probabilities of playing sub-optimal actions at future information sets) going to zero. This makes the strategies prescribed by an EFPE robust against possible players' mistakes.

NE refinements have received some attention from the artificial intelligence community only recently. This is surprising, since designing artificial agents that

Proceedings of the 27$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

are capable of exploiting opponents' mistakes is of paramount importance when bringing equilibrium concepts into practice. Indeed, the adoption of NE refinements would foster the operationalization of equilibrium computation techniques into novel application scenarios where "humans are in the loop", such as, *e.g.*, military settings and businesses. Moreover, NE refinements could boost the performance of state-of-the-art superhuman agents playing recreational games like Poker, by equipping them with the ability of capitalizing over opponents' mistakes.

Despite there are several appealing reasons for switching the attention from the computation of NEs to that of its refinements, only few works addressed the latter problem in two-player zero-sum sequential games with imperfect information. Moreover, most of these works, such as, *e.g.*, (Farina and Gatti, 2017; Farina et al., 2018a; Farina and Sandholm, 2021), provide equilibrium computation algorithms that rely on the solution of *linear programs* (LPs). As a result, up to now, these methods have failed at scaling up to huge games like Texas hold'em Poker endgames. This was predictable, as the major breakthroughs in NE computation in two-player zero-sum games (such as, *e.g.*, (Brown and Sandholm, 2018, 2019)) were achieved by means of iterative techniques that work by repeatedly visiting the game tree; see, *e.g.*, the *counterfactual regret minimization* (CFR) algorithm (Zinkevich et al., 2007). To the best of our knowledge, the only iterative methods for computing NE refinements are those presented in (Farina et al., 2017; Kroer et al., 2017). However, these do *not* provide any convergence guarantee to exact refinements, but only to approximate equilibria of a perturbed game (see the following Section 2).

In this paper, we provide the first *iterative* algorithm that provably converges to (exact) NE refinements, focusing on the case of EFPEs. Our algorithm works by tracking a *sequence of equilibria* of suitably-defined, regularized-perturbed games, different from the sequence in the definition of EFPE. Indeed, these games modify the utility function of the original game by adding *regularization* and *perturbation* components, while the definition of EFPE only considers the latter. The regularization component is needed to ensure equilibrium uniqueness, while the perturbation component guarantees that the equilibrium strategies place at least a (small) probability on every action, resembling the idea of "trembles". By carefully controlling how the regularization and the perturbation components vanish, the resulting sequence of equilibria admits an EFPE of the original game as a limit point. This is achieved by letting the regularization component vanishing faster than the perturbation one.

The core component of our algorithm is a *last-iterate* procedure provably converging to an (approximate) equilibrium of a given regularized-perturbed game. Such a procedure extends the *optimistic online mirror descent* (OOMD) algorithm by Rakhlin and Sridharan (2013). Intuitively, it uses an OOMD-style update rule in order to converge last iterate, while it deals with the non-smooth terms appearing in the utility of regularized-perturbed games without resorting to linearization techniques. The crucial feature of our procedure is that its updates can be performed by visiting the game tree recursively, thus avoiding the solution of "one-piece" optimization problems. This considerably enhances the scalability of our algorithm compared to those solving LPs over the players' strategy spaces.

We conclude by experimentally evaluating the performances of our algorithm on a standard testbed of Poker-inspired game instances. Our analysis shows that our algorithm outperforms of orders of magnitude several state-of-the-art equilibrium-computation algorithms—including the CFR algorithm (Zinkevich et al., 2007) and the iterative methods for computing (approximate) NE refinements introduced in (Farina et al., 2017; Kroer et al., 2017)—in terms of average players' regrets over all the information sets of the game, which is the standard practical metric used for evaluating convergence to NE refinements.

2 RELATED WORKS

The first works addressing the problem of finding NE refinements in two-player zero-sum sequential games are (Miltersen and Sørensen, 2010; Farina and Gatti, 2017; Farina et al., 2018a). All the algorithms in these works rely on the solution of "perturbed" LPs, which are modified versions of the LP for finding an NE in two-player zero-sum games. Thus, such algorithms fail at scaling up to real-world-size game instances, where it is well-established that an (approximate) NE can be computed by means of iterative methods that exploit the tree structure of the game.

The only algorithms that compute EFPEs by iteratively visiting the game tree are those in (Farina et al., 2017) and (Kroer et al., 2017). The former is a modification of the CFR algorithm that takes into account opponents' mistakes so as to minimize regret at all the information sets, including those never reached at the equilibrium. The latter applies a similar idea to first-order methods. Such iterative methods are able to converge to an approximate equilibrium of a perturbed game. However, they are *not* concerned with the computation of a limit point of a sequence of equilibria, since they work by approximating an equilibrium while keeping the "magnitude of trembles" fixed. Thus, they do *not* provide any theoretical convergence

guarantee to an exact EFPE, which would require taking a limit point as "trembles" vanish.

It is also worth citing some works studying NE refinements in n-player general-sum sequential games (Miltersen and Sørensen, 2010; Hansen et al., 2010; Farina and Gatti, 2017; Gatti et al., 2020) and refinements of other solution concepts (Farina et al., 2018b; Marchesi et al., 2019; Marchesi and Gatti, 2021).

3 PRELIMINARIES

In this section, we provide all the definitions and concepts needed in the paper.

3.1 Extensive-Form Games

An *extensive-form game* (EFG) is usually described by means of a game tree. We denote by $\mathcal{H}$ the set of nodes of the tree and by $\mathcal{Z} \subset \mathcal{H}$ the set of terminal nodes, which are the leaves of the tree. Each non-terminal node $h \in \mathcal{H} \setminus \mathcal{Z}$ is either a decision node in which a given player acts or a chance node where a random event occurs. For any decision node h, we let $A(h)$ be the set of actions available to the player acting at h. On the other hand, each terminal node $z \in Z$ is associated with a payoff $u(z) \in [-1, 1]$ for the first player, while $-u(z)$ is the payoff of the second player since the game is zero sum.

In an EFG, imperfect information is described by means of *information sets* (infosets). A player's infoset I is a collection of decision nodes of that player which are indistinguishable for them; formally, it must be the case that $A(h) = A(h')$ for any pair $h, h' \in I$. We let $A(I)$ be the set of actions available at all nodes of infoset I, with $n_I \coloneqq |A(I)|$ being its cardinality. Moreover, we denote by $\mathcal{I}_1$ and $\mathcal{I}_2$ the sets all the infosets of player 1 and 2, respectively. As customary in the literature, we restrict the attention to games with *perfect recall*, where a player never forgets information once acquired. This is equivalent to assuming that both $\mathcal{I}_1$ and $\mathcal{I}_2$ are partially ordered according to a precedence relation denoted by $\preceq$. Namely, here $I \preceq J$ means that infoset I *precedes* infoset J. Formally, $I \preceq J$ if and only if there exists a path in the game tree that connects a node belonging to infoset I to a node in infoset J. Furthermore, give any infoset $I \in \mathcal{I}_i$ of player i and action $a \in A(I)$, we let $C_{I,a}$ be the set of all player i's infosets that *immediately follow* I *through action* a, according to the relation $\preceq$.

3.2 Sequence-Form Representation

Any node $h \in \mathcal{H}$ defines a *sequence* $\sigma_i(h)$ of player i's actions encountered on the path from the root of the game tree to h. In EFGs with perfect recall, an infoset $I \in \mathcal{I}_i$ uniquely determines a sequence of player i's actions, since $\sigma_i(h) = \sigma_i(h')$ for any $h, h' \in I$ by definition. We denote such a sequence by $\sigma_i(I)$. Since $\sigma_i(I)$ extended with any action $a \in A(I)$ is a valid sequence of player i's actions, we can identify player i's sequences with infoset-action pairs. Thus, we let $\Sigma_i \coloneqq \{(I, a) \mid I \in \mathcal{I}_i, a \in A(I)\} \cup \{\varnothing\}$ be the set of player i's sequences, where $\varnothing$ is the empty sequence defined by paths in which player i never plays.

By leveraging the sequence form, the *mixed strategies* of a player can be encoded in terms of realization probabilities of sequences (Von Stengel, 1996). Formally, a strategy of the first player is a vector $\boldsymbol{x} \in [0,1]^{|\Sigma_1|}$ such that, for each $\sigma \in \Sigma_1$, $\boldsymbol{x}[\sigma]$ is the probability of playing sequence σ. To be well defined, $\boldsymbol{x}$ must satisfy the following linear constraints:

$$\boldsymbol{x}[\varnothing] = 1 \text{ and } \boldsymbol{x}[\sigma_i(I)] = \textstyle\sum_{a \in A(I)} \boldsymbol{x}[\sigma_i(I)a] \quad \forall I \in \mathcal{I}_1,$$

which can be expressed as $\boldsymbol{F}_1 \boldsymbol{x} = \boldsymbol{f}_1$ using matrix notation. Similarly, we let $\boldsymbol{y} \in [0,1]^{|\Sigma_2|}$ be a second player's strategy, which must satisfy the constraints $\boldsymbol{F}_2 \boldsymbol{y} = \boldsymbol{f}_2$. We denote by $\mathcal{X}$ and $\mathcal{Y}$ the polytopes of valid strategies for player 1 and 2, respectively.

Thanks to the sequence-form strategy representation, we can define the first player's expected utility given two strategies $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$ as the bilinear term $\boldsymbol{x}^\top \boldsymbol{U} \boldsymbol{y}$, where $\boldsymbol{U} \in [-1,1]^{|\Sigma_1| \times |\Sigma_2|}$ is a utility matrix whose entry corresponding to $\sigma_1 \in \Sigma_1$ and $\sigma_2 \in \Sigma_2$ is defined as follows:

$$\boldsymbol{U}[\sigma_1, \sigma_2] \coloneqq \sum_{z \in \mathcal{Z} : \sigma_1(z) = \sigma_1 \wedge \sigma_2(z) = \sigma_2} p(z) u(z),$$

with $p(z) \in [0, 1]$ being the product of chance probabilities on the path from the root to $z \in \mathcal{Z}$. The second player's expected utility is $-\boldsymbol{x}^\top \boldsymbol{U} \boldsymbol{y}$.

3.3 Extensive-Form Perfect Equilibria

The EFPE refines the NE by considering the possibility that players may make mistakes and play off-equilibrium actions with "small and vanishing" probability. Formally, an EFPE is defined as a limit point as $\epsilon \to 0$ of a sequence of NEs of "perturbed games" parametrized by $\epsilon > 0$, where each action $a \in A(I)$ at each infoset I must be played with probability at least ϵ (Selten, 1975). In terms of sequence form, this is equivalent to ask, for the first player, that $\boldsymbol{x}[\sigma_i(I)a] \geq \epsilon \boldsymbol{x}[\sigma_i(I)]$ for every $I \in \mathcal{I}_1, a \in A(I)$. Such linear constraints can be expressed as a polytope $\boldsymbol{M}_1(\epsilon)\boldsymbol{x} \geq \boldsymbol{m}_1(\epsilon)$ (see (Farina and Gatti, 2017)), so that we can define the set of valid first player's strategies for a perturbed game parametrized by $\epsilon > 0$ as:

$\mathcal{X}^\epsilon := \{\boldsymbol{x} \in \mathcal{X} \mid \boldsymbol{F}_1\boldsymbol{x} = \boldsymbol{f}_1, \boldsymbol{M}_1(\epsilon)\boldsymbol{x} \geq \boldsymbol{m}_1(\epsilon)\}$. Similarly, $\mathcal{Y}^\epsilon := \{\boldsymbol{y} \in \mathcal{Y} \mid \boldsymbol{F}_2\boldsymbol{y} = \boldsymbol{f}_2, \boldsymbol{M}_2(\epsilon)\boldsymbol{y} \geq \boldsymbol{m}_2(\epsilon)\}$ is the set of valid strategies for the second player. Then, an EFPE is defined as follows.[1]

Definition 3.1 (EFPE)**.** Given any $\epsilon > 0$, an *ϵ-EFPE* is defined as any pair of strategies $(\boldsymbol{x}^\star, \boldsymbol{y}^\star) \in \mathcal{X}^\epsilon \times \mathcal{Y}^\epsilon$ such that, for all $\boldsymbol{x} \in \mathcal{X}^\epsilon$ and $\boldsymbol{y} \in \mathcal{Y}^\epsilon$, it holds:

$$\boldsymbol{x}^\top \boldsymbol{U}\boldsymbol{y}^\star \leq \boldsymbol{x}^{\star,\top}\boldsymbol{U}\boldsymbol{y}^\star \leq \boldsymbol{x}^{\star,\top}\boldsymbol{U}\boldsymbol{y}.$$

Finally, an *EFPE* is a limit point of ϵ-EFPEs as $\epsilon \to 0$.

In this work, we also introduce a relaxed notion of EFPE, which we call δ-approximate ϵ-EFPE and it is defined as δ-approximate NEs of ϵ-perturbed games.

Definition 3.2. Given any $\epsilon > 0$ and $\delta > 0$, a *δ-approximate ϵ-EFPE* is a pair of strategies $(\boldsymbol{x}^\star, \boldsymbol{y}^\star) \in \mathcal{X}^\epsilon \times \mathcal{Y}^\epsilon$ such that, for all $\boldsymbol{x} \in \mathcal{X}^\epsilon$ and $\boldsymbol{y} \in \mathcal{Y}^\epsilon$, it holds:

$$\boldsymbol{x}^\top \boldsymbol{U}\boldsymbol{y}^\star - \delta \leq \boldsymbol{x}^{\star,\top}\boldsymbol{U}\boldsymbol{y}^\star \leq \boldsymbol{x}^{\star,\top}\boldsymbol{U}\boldsymbol{y} + \delta.$$

Notice that, when $\delta = 0$, then a limit point as $\epsilon \to 0$ of δ-approximate ϵ-EFPEs is an EFPE.

Finally, we measure how well a pair of players' strategies approximates an NE in terms of *Nash gap*. For every $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in \mathcal{X} \times \mathcal{Y}$, this is defined as $Gap(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) := \max_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{x}^\top \boldsymbol{U}\tilde{\boldsymbol{y}} - \min_{\boldsymbol{y} \in \mathcal{Y}} \tilde{\boldsymbol{x}}^\top \boldsymbol{U}\boldsymbol{y}$.

4 SEQUENCES OF EQUILIBRIA LEADING TO EFPE

In this section, we provide the core results that allow us to design our algorithm. First, we introduce a modified version of an EFG, which alters the utility function of the game by adding suitable *regularization* and *perturbation* components, where the former guarantees equilibrium uniqueness, while the latter ensures that players' strategies at the equilibrium are valid for the "perturbed games" in the definition of EFPE. Then, we show that such a modified game allows us to identify a *sequence of equilibria* that admits an EFPE as a limit point, by carefully tuning the parameters controlling the regularization and the perturbation components in the modified game. In the following Section 5, we provide an algorithm that is able to track such a sequence, and converges to an EFPE.

4.1 Regularized and Perturbed Games

Next, we define regularized-perturbed games and show some of their properties useful in proving our core results, namely equilibrium uniqueness and the connection between their equilibria and the δ-approximate ϵ-EFPEs of the original game (Theorem 4.1).

[1]For ease of notation, $\epsilon \to 0$ is the limit from the right.

Given any EFG and two parameters $\lambda, \epsilon > 0$, we define the *regularized-perturbed game* $\mathcal{G}(\lambda, \epsilon)$ as a two-player zero-sum game in which the players' strategy sets are those of the original game (namely $\mathcal{X}$ and $\mathcal{Y}$) and the first player's utility for any pair of strategies $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$ is given by:

$$f_{\lambda,\epsilon}(\boldsymbol{x}, \boldsymbol{y}) := \boldsymbol{x}^\top \boldsymbol{U}\boldsymbol{y} - \frac{1}{\lambda} d_1^\epsilon(\boldsymbol{x}) + \frac{1}{\lambda} d_2^\epsilon(\boldsymbol{y}),$$

where $d_1^\epsilon : \mathcal{X} \to \mathbb{R}$ and $d_2^\epsilon : \mathcal{Y} \to \mathbb{R}$ are strongly convex distance-generating functions that are defined recursively over the sequence-form strategy sets $\mathcal{X}$ and $\mathcal{Y}$, respectively, as follows:

$$d_1^\epsilon(\boldsymbol{x}) := \sum_{I \in \mathcal{I}_1} \alpha_I \, \boldsymbol{x}[\sigma_1(I)] d_{\Delta_I}^\epsilon \left(\frac{\boldsymbol{x}[I]}{\boldsymbol{x}[\sigma_1(I)]} \right),$$

$$d_2^\epsilon(\boldsymbol{y}) := \sum_{I \in \mathcal{I}_2} \alpha_I \, \boldsymbol{y}[\sigma_2(I)] d_{\Delta_I}^\epsilon \left(\frac{\boldsymbol{y}[I]}{\boldsymbol{y}[\sigma_2(I)]} \right).$$

In the definitions of $d_1^\epsilon(\boldsymbol{x})$ and $d_2^\epsilon(\boldsymbol{y})$, we introduced the following useful additional notation:

- for every infoset $I \in \mathcal{I}_1$ (resp. $I \in \mathcal{I}_2$), the vector $\boldsymbol{x}[I] \in [0,1]^{n_I}$ (resp. $\boldsymbol{y}[I] \in [0,1]^{n_I}$) is the subvector of $\boldsymbol{x}$ (resp. $\boldsymbol{y}$) made by all the components $\boldsymbol{x}[\sigma_1(I)a]$ (resp. $\boldsymbol{y}[\sigma_2(I)a]$) for $a \in A(I)$;
- for every $I \in \mathcal{I}_1 \cup \mathcal{I}_2$, the function $d_{\Delta_I}^\epsilon : \Delta^{n_I} \to \mathbb{R}$ is such that $d_{\Delta_I}^\epsilon(\boldsymbol{w}) := (\boldsymbol{w} - \mathbf{1}\epsilon)^\top \log(\boldsymbol{w} - \mathbf{1}\epsilon)$;[2]
- for all $I \in \mathcal{I}_1 \cup \mathcal{I}_2$, the weights $\alpha_I \in \mathbb{R}_+$ are recursively defined as $\alpha_I := 2 + 2\max_{a \in A(I)} \sum_{J \in \mathcal{C}_{I,a}} \alpha_J$, so as to guarantee that d_1^ϵ and d_2^ϵ are 1-strongly convex functions with respect to the ℓ_2-norm (see Lemma A.2 in Appendix A).

In this paper, we will also use the *Bregman divergences* $D_i^\epsilon(\cdot|\cdot)$ associated with the distance-generating functions d_i^ϵ. Formally, for every pair $\boldsymbol{x}, \tilde{\boldsymbol{x}} \in \mathcal{X}$, we define $D_1^\epsilon(\boldsymbol{x}|\tilde{\boldsymbol{x}}) := d_1^\epsilon(\boldsymbol{x}) - d_i^\epsilon(\tilde{\boldsymbol{x}}) - \nabla d_i^\epsilon(\tilde{\boldsymbol{x}})^\top(\boldsymbol{x} - \tilde{\boldsymbol{x}})$, while $D_2^\epsilon(\boldsymbol{y}|\tilde{\boldsymbol{y}})$ is defined analogously.

Notice that the functions d_1^ϵ and d_2^ϵ are special cases of *dilated* distance-generating functions, which are defined by directly exploiting the tree form of sequence-form strategy sets $\mathcal{X}$ and $\mathcal{Y}$ (Hoda et al., 2010; Lee et al., 2021). In particular, d_1^ϵ and d_2^ϵ employ the $d_{\Delta_I}^\epsilon$ as base distance-generating functions for the simplexes defining the strategy spaces at every infoset I. These functions modify the classical negative entropy by offsetting by $-\epsilon$ the strategy given as input. Intuitively,

[2]In this work, Δ_I denotes the $(n_I - 1)$-dimensional simplex defined over the set $A(I)$ of actions at infoset $I \in \mathcal{I}_1 \cup \mathcal{I}_2$. We also use $\log(\boldsymbol{w})$ as a shorthand for the vector whose k-th component is $\log(\boldsymbol{w}[k])$. Moreover, for the function $d_{\Delta_I}^\epsilon$ to be well defined for every $I \in \mathcal{I}_1 \cup \mathcal{I}_2$, we assume w.l.o.g. that $\epsilon \leq \min_{I \in \mathcal{I}_1 \cup \mathcal{I}_2} 1/2n_I$.

this ensures that the equilibria of the game belong to $\mathcal{X}^\epsilon \times \mathcal{Y}^\epsilon$, and, thus, the parameter ϵ can be used to tune the perturbation component.

We also notice that the term $1/\lambda$ that multiplies the functions d_1^ϵ and d_2^ϵ is the one controlling the regularization component, and it is crucial to recover the strong convexity-concavity of the utility function $f_{\lambda,\epsilon}$, which, in its turn, guarantees equilibrium uniqueness. By letting $\boldsymbol{z}^\star_{\lambda,\epsilon} \coloneqq (\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}^\star_{\lambda,\epsilon}) \in \mathcal{X}^\epsilon \times \mathcal{Y}^\epsilon$ be an NE of the game $\mathcal{G}(\lambda,\epsilon)$, it is easy to check that it must satisfy the following conditions:

$$\boldsymbol{x}^\star_{\lambda,\epsilon} \in \arg\max_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{x}^\top \boldsymbol{U} \boldsymbol{y}^\star_{\lambda,\epsilon} - \frac{1}{\lambda} d_1^\epsilon(\boldsymbol{x}) \right\},$$

$$\boldsymbol{y}^\star_{\lambda,\epsilon} \in \arg\min_{\boldsymbol{y}\in\mathcal{Y}} \left\{ \boldsymbol{x}^{\star,\top}_{\lambda,\epsilon} \boldsymbol{U} \boldsymbol{y} + \frac{1}{\lambda} d_2^\epsilon(\boldsymbol{y}) \right\}.$$

Then, by the fact that the objectives of the problems above are strongly convex, one can conclude that a pair of solutions $(\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}^\star_{\lambda,\epsilon})$ is unique and changes smoothly when the parameters are suitably tuned.

We remark that equilibrium uniqueness in regularized-perturbed games plays a crucial in the construction underpinning our algorithm, as we will show in Section 5. Moreover, let us also notice that our approach to ensure equilibrium uniqueness is inspired by ideas taken from the *quantal equilibrium* (McKelvey and Palfrey, 1995). Indeed, it is easy to check that, setting $\epsilon = 0$ in $f_{\lambda,\epsilon}$ defines quantal equilibria.

Next, we prove that the unique NE of a game $\mathcal{G}(\lambda,\epsilon)$ is indeed a δ-approximate ϵ-EFPE of the original EFG, where the approximation δ linearly depends on the regularization component $1/\lambda$.[3]

Theorem 4.1. *Given an EFG and two parameters $\lambda, \epsilon > 0$, the unique NE $\boldsymbol{z}^\star_{\lambda,\epsilon}$ of $\mathcal{G}(\lambda,\epsilon)$ constitutes a $O\left(\frac{1}{\lambda}\right)$-approximate ϵ-EFPE of the EFG.*

4.2 How to Select a Sequence Leading to an EFPE

From Theorem 4.1, one could naïvely think that, by tracking a sequence of NEs of $\mathcal{G}(\lambda,\epsilon)$ as $\lambda \to +\infty$ and $\epsilon \to 0$, it is possible to recover an EFPE of the EFG. In the following, we show that this is *not* always the case, since, in order to identify the sequence of equilibria leading to an EFPE, one needs to carefully control how the parameters λ and ϵ converge to $+\infty$ and zero, respectively.

First, let us remark that, as an immediate corollary of Theorem 4.1, we have that taking the limit as $\lambda \to +\infty$ before letting $\epsilon \to 0$ allows to recover an EFPE, since the theorem shows that $\lim_{\lambda\to+\infty} \boldsymbol{z}^\star_{\lambda,\epsilon}$ is an NE of a perturbed game. However, doing so will result in losing all the benefits of the regularization component, which are needed in order to be able to design an efficient algorithm converging to an EFPE. Formally:

Corollary 4.2. *Given any EFG, let us define $\boldsymbol{z}^\star \coloneqq \lim_{\epsilon\to0} \lim_{\lambda\to\infty} \boldsymbol{z}^\star_{\lambda,\epsilon}$. Then, $\boldsymbol{z}^\star$ is an EFPE of the EFG.*

However, using *any* sequence $\{(\lambda_k,\epsilon_k)\}_{k\in\mathbb{N}}$ such that $\lambda_k \to +\infty$ and $\epsilon_k \to 0$ jointly as $k \to +\infty$ does *not* lead to EFPEs, as shown by the following proposition:

Proposition 4.1. *There exists an EFG for which $\lim_{\lambda\to\infty} \lim_{\epsilon\to0} \boldsymbol{z}^\star_{\lambda,\epsilon} \neq \lim_{\epsilon\to0} \lim_{\lambda\to\infty} \boldsymbol{z}^\star_{\lambda,\epsilon}$ and, additionally, $\lim_{\lambda\to\infty} \lim_{\epsilon\to0} \boldsymbol{z}^\star_{\lambda,\epsilon}$ is not an EFPE.*

By results on iterated limits (Steinlage, 1971), we have that the double limit $\lim_{(\lambda,\epsilon)\to(+\infty,0)} \boldsymbol{z}^\star_{\lambda,\epsilon}$ may *not* exist. This immediately implies that one cannot consider any arbitrary $\{(\lambda_k,\epsilon_k)\}_{k\in\mathbb{N}}$ in order to define a sequence of equilibria leading to an EFPE. Indeed, the only guarantee is that the sequence leads to *an* NE of the original EFG. As a result, the sequence $\{(\lambda_k,\epsilon_k)\}_{k\in\mathbb{N}}$ must be built in a specific way.

Next, we show that we need sequences $\{(\lambda_k,\epsilon_k)\}_{k\in\mathbb{N}}$ that are defined so that the sequence made by the λ_k converges faster than that of the ϵ_k. Formally:

Theorem 4.3. *Given any EFG, there exists a function $\mu : (0,1] \to \mathbb{R}_+$ such that, given any sequence $\{(\lambda_k,\epsilon_k)\}_{k\in\mathbb{N}}$ defined so that $\lambda_k \to +\infty$, $\epsilon_k \to 0$, and $1/\lambda_k < \mu(\epsilon_k)$ as $k \to +\infty$, the limit $\lim_{k\to+\infty} \boldsymbol{z}^\star_{\lambda_k,\epsilon_k}$ identifies an EFPE of the given EFG.*

The proof of Theorem 4.3 is based on the core idea that, by choosing $\lambda_k > 1/\mu(\epsilon_k)$ for a suitably-defined function $\mu : (0,1] \to \mathbb{R}_+$, the resulting sequence $(1/\lambda_k, \epsilon_k)$ is "close" to the one defined by the iterated limit in Corollary 4.2, which is guaranteed to be an EFPE. Indeed, this amounts to showing that, for k sufficiently large, the sequence $(1/\lambda_k, \epsilon_k)$ is close to $(0, \epsilon_k)$, where the latter sequence converges to an EFPE thanks to Corollary 4.2.

4.3 The Need of Regularization

Next, we argue why adding a regularization component is of paramount importance in our setting. Indeed, one could argue that it would be more natural to track the sequence of ϵ-EFPEs that appear in the definition of EFPE (Definition 3.1). However, such a methodology would *not* work. Indeed, such a natural sequence, tough always admitting a limit point, can have discontinuities, bifurcations, multiple branches, and other irregularities, which are drawback inherited by the convoluted structure of NE in linear games. In particular, non-uniqueness alone would doom any

[3] All the proofs are in the Appendixes A and B.

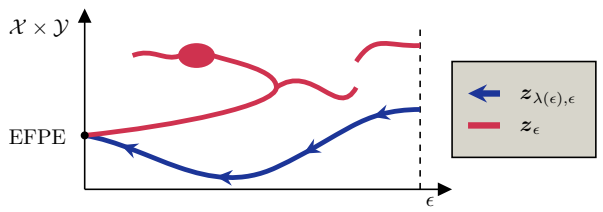

Figure 1: An example of the "smooth" sequence of equilibria $\boldsymbol{z}_{\lambda(\epsilon),\epsilon}$ of regularized-perturbed games (in *blue*) and the sequence of ϵ-EFPEs $\boldsymbol{z}_\epsilon$ in the definition of EFPE (in *red*).

method that relies on last-iterate convergence, since uniqueness is a key assumption in such methods.

Conversely, our sequence of equilibria defined in terms of regularized-perturbed games is "smooth"—as it is $\mathcal{C}^\infty$-differentiable—and it identifies a unique branch, while sharing the same limit point with the sequence in the definition of EFPE. This makes our sequence much easier to track. Figure 1 shows an example of the sequence used in the definition of EFPE and the one of our regularized-perturbed games.

5 EFFICIENT ITERATIVE ALGORITHM FOR EFPE

We are now ready to introduce our *last-iterate* algorithm that converges to an (exact) EFPE (Algorithm 1). The algorithm works by tracking a sequence of equilibria $\boldsymbol{z}_k^\star := \boldsymbol{z}_{\lambda_k,\epsilon_k}^\star$ of regularized-perturbed games $\mathcal{G}_k := \mathcal{G}(\lambda_k, \epsilon_k)$ by letting $k \to +\infty$, where the latter games are defined by means of a sequence $\{(\lambda_k, \epsilon_k)\}_{k\in\mathbb{N}}$ as in the statement of Theorem 4.3.

Algorithm 1 works in phases. For every $k \in \mathbb{N}$, the k-th phase of the algorithm is devoted to finding a suitable approximation of the equilibrium $\boldsymbol{z}_k^\star$ of the regularized-perturbed game $\mathcal{G}_k$. This is achieved by performing $T_k := \beta^k$ (for a given $\beta > 1$) iterations of a last-iterate sub-procedure (Algorithm 2) that converges linearly to an approximate equilibrium of $\mathcal{G}_k$ (see Theorem 5.1). In any given phase k, finding an approximation of $\boldsymbol{z}_k^\star$ is sufficient, since the sequence made by the equilibria $\boldsymbol{z}_k^\star$ has the only purpose of identifying EFPEs as its limit points. Algorithm 2 is an extension of the OOMD algorithm, when instantiated for sequence-form strategy sets. Similarly to the OOMD algorithm, Algorithm 2 performs suitably-defined intermediate updates, labeled with $t + \frac{1}{2}$.

Notice that, in our setting, it is *not* possible to use the standard update rule of OOMD, by directly feeding the algorithm with the gradients of f_{λ_k,ϵ_k} of regularized-perturbed games $\mathcal{G}_k$. This is because such functions are non-smooth due to the $d_i^{\epsilon_k}$ having unbounded gradient norm near the boundaries of $\mathcal{X}^{\epsilon_k} \times \mathcal{Y}^{\epsilon_k}$. Indeed, $d_i^{\epsilon_k}$ are Legendre functions (Cesa-Bianchi and Lugosi, 2006). In order to circumvent this issue, we use an idea introduced by the *composite objective mirror descent* algorithm (Duchi et al., 2010), which consists in avoiding the linearization of the non-smooth part in the utility function to deal with the non-Lipschitzness of the gradients. The convergence guarantees of Algorithm 2 are the following:

Theorem 5.1. *Define* $D^{\epsilon_k}(\boldsymbol{z}|\tilde{\boldsymbol{z}}) := D_1^{\epsilon_k}(\boldsymbol{x}|\tilde{\boldsymbol{x}}) + D_2^{\epsilon_k}(\boldsymbol{y}|\tilde{\boldsymbol{y}})$, *then for all* $\eta \le 1/\sqrt{2}\|\boldsymbol{U}\|_2$, $k \in \mathbb{N}$ *and round* $t \in \{1, \dots, T_k\}$, *Algorithm 2 guarantees:*

$$\|\boldsymbol{z}_k^\star - \boldsymbol{z}_t^{(k)}\|_2 \le 2\left[2D^{\epsilon_k}\left(\boldsymbol{z}_k^\star|\boldsymbol{z}_0^{(k)}\right)\right]^{1/2}\left(\frac{\lambda_k}{\lambda_k+\eta}\right)^{t/2}.$$

We also remark that the proof of the theorem above can be generalized to achieve last-iterate convergence to equilibria of EFGs for *any* dilated regularization functions d_i. Indeed, our results do *not* rely on the the specific d_i used here that are specifically tailored for EFPEs. Let us remark that the problem of designing learning algorithms with last-iterate converge guarantees to equilibria in games has recently received considerable attention (see, *e.g.*, (Cen et al., 2021, 2022; Liu et al., 2022; Sokota et al., 2022; Abe et al., 2023))

5.1 Convergence Analysis

By exploiting the analysis of Algorithm 2 in Theorem 5.1, the following theorem formally proves the convergence guarantees of Algorithm 1 to (exact) EFPEs.

Theorem 5.2. *Given any sequence* $\{(\lambda_k, \epsilon_k)\}_{k\in\mathbb{N}}$ *defined as in Theorem 4.3 and satisfying* $\eta \le \lambda_k^2 \le \eta\beta^k/2$ *for every* $k \in \mathbb{N}$, *Algorithm 1 grantees that* $\lim_{k\to\infty} \boldsymbol{z}_{T_k}^{(k)}$ *is an EFPE.*

Moreover, we can exploit the strongly convex-concave structure of the utility functions $f_{\lambda,\epsilon}$ to prove Nash gap guarantees for regularized-perturbed games $\mathcal{G}_{\lambda,\epsilon}$, which can then be combined with Theorem 4.1 to obtain guarantees on the exploitability of $\boldsymbol{z}_{T_k}^{(k)}$ in the original EFG. Formally:

Theorem 5.3. *Let* $\tilde{\boldsymbol{z}} := (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in \mathcal{X} \times \mathcal{Y}$ *be such that it holds* $\|\boldsymbol{z}_{\lambda,\epsilon}^\star - \tilde{\boldsymbol{z}}\|_2 \le \nu$ *for some* $\nu > 0$, *then:*

$$Gap(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \le \nu|\Sigma|^{3/2} + 8C/\lambda,$$

where $|\Sigma| := \max\{|\Sigma_1|, |\Sigma_2|\}$ *and* $C > 0$ *is a constant that depends polynomially in* $\max_{i\in\{1,2\}} |\mathcal{I}_i|$, $\max_{I\in\mathcal{I}_1\cup\mathcal{I}_2} \alpha_I$, *and* $\max_{I\in\mathcal{I}_1\cup\mathcal{I}_2} \log n_I$.

Algorithm 1 Compute EFPE

function COMPUTE-EFPE($\{(\lambda_k, \epsilon_k)\}, \beta, \eta$)
  $\{(\lambda_k, \epsilon_k)\}_k$: *As in Theorem 4.3.*
  $\beta > 1$: *Phase growing rate.*
  $\eta > 0$: *Learning rate.*
  $\boldsymbol{z}^{(0)} \in \mathcal{X} \times \mathcal{Y}$; $k \leftarrow 0$; $T \leftarrow 0$
  while *not* exceeding time limit **do**
    $k \leftarrow k + 1$
    Instantiate game $\mathcal{G}_k \coloneqq \mathcal{G}(\lambda_k, \epsilon_k)$
    $T_k \leftarrow \beta^k$; $T \leftarrow T + T_k$
    $\boldsymbol{z}^{(k)}_{T_k} \leftarrow \text{SOLVE}(\mathcal{G}_k, T_k, \boldsymbol{z}^{(k-1)}_{T_{k-1}}, \eta)$
  return $\boldsymbol{z}^{(k)}_{T_k}$

Algorithm 2 Solve game $\mathcal{G}_k$

function SOLVE($\mathcal{G}_k, T_k, \boldsymbol{z}^{(k)}_{-1/2}, \eta$)
  for $t = 0, \dots, T_k - 1$ **do**
    $\boldsymbol{x}^{(k)}_{t+\frac{1}{2}} \leftarrow \arg\max_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{x}^\top \boldsymbol{U} \boldsymbol{y}^{(k)}_t - \frac{d_1^{\epsilon_k}(\boldsymbol{x})}{\lambda_k} - \frac{1}{\eta} D_1^{\epsilon_k}\left(\boldsymbol{x} | \boldsymbol{x}^{(k)}_{t-\frac{1}{2}}\right) \right\}$
    $\boldsymbol{x}^{(k)}_{t+1} \leftarrow \arg\max_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{x}^\top \boldsymbol{U} \boldsymbol{y}^{(k)}_t - \frac{d^{\epsilon_k}(\boldsymbol{x})}{\lambda_k} - \frac{1}{\eta} D_1^{\epsilon_k}\left(\boldsymbol{x} | \boldsymbol{x}^{(k)}_{t+\frac{1}{2}}\right) \right\}$
    $\boldsymbol{y}^{(k)}_{t+\frac{1}{2}} \leftarrow \arg\min_{\boldsymbol{y} \in \mathcal{Y}} \left\{ \boldsymbol{x}^{(k),\top}_t \boldsymbol{U} \boldsymbol{y} + \frac{d_2^{\epsilon_k}(\boldsymbol{y})}{\lambda_k} + \frac{1}{\eta} D_2^{\epsilon_k}\left(\boldsymbol{y} | \boldsymbol{y}^{(k)}_{t-\frac{1}{2}}\right) \right\}$
    $\boldsymbol{y}^{(k)}_{t+1} \leftarrow \arg\min_{\boldsymbol{y} \in \mathcal{Y}} \left\{ \boldsymbol{x}^{(k),\top}_t \boldsymbol{U} \boldsymbol{y} + \frac{d_2^{\epsilon_k}(\boldsymbol{y})}{\lambda_k} + \frac{1}{\eta} D_2^{\epsilon_k}\left(\boldsymbol{y} | \boldsymbol{y}^{(k)}_{t+\frac{1}{2}}\right) \right\}$
  return $\boldsymbol{z}^{(k)}_{T_k} \coloneqq \left(\boldsymbol{x}^{(k)}_{T_k}, \boldsymbol{y}^{(k)}_{T_k}\right)$

As a direct corollary, we can prove the following guarantees in terms of Nash gap for Algorithm 1.

Corollary 5.4. *Given any sequence $\{(\lambda_k, \epsilon_k)\}_{k \in \mathbb{N}}$ as in Theorem 5.2, at the end of every phase $k \in \mathbb{N}$ of Algorithm 1, the following holds:*

$$Gap\left(\boldsymbol{x}^{(k)}_{T_k}, \boldsymbol{y}^{(k)}_{T_k}\right) \leq \frac{4}{\lambda_k} |\Sigma|^{\frac{3}{2}} \left[2 D^{\epsilon_k}(\boldsymbol{z}^\star_k | \boldsymbol{z}^{(k)}_0)\right]^{\frac{1}{2}} + \frac{8C}{\lambda_k}.$$

The result above shows that, after $T = \sum_{j=1}^{k} \beta^j = O(\beta^k)$ rounds, one gets a Nash gap of the order of $O(1/\lambda_k)$. Thus, by choosing $\lambda_k = O(\beta^{k/2})$ (which is the largest order of λ_k allowed by Corollary 5.4), we have a bound of $O(1/\sqrt{T})$ on the Nash gap.

Notice that Theorem 4.3 shows asymptotic convergence in terms of ℓ_2-distance to an EFPE, while Corollary 5.4 shows convergence in terms of Nash gap. Determining whether it is possible to show finite convergence rates also for the ℓ_2-distance to the set of EFPEs requires additional research. The difficulty of such an inquiry is that, even in (non-perturbed) normal-form games, our problem reduces to finding convergence rates to limit-quantal equilibria, which is still open.

5.2 Efficient Decomposition

The pseudo-code formulation of Algorithm 2 does *not* scale efficiently, since at each iteration it requires to solve convex problems over the whole set $\mathcal{X}$ and $\mathcal{Y}$.

Next, we show that each update of Algorithm 2 can be implemented *efficiently* by directly working on the game tree with closed form solutions. In particular, we show how to implement each update in Algorithm 2 so that: (i) it is performed recursively with a bottom-up visit of one player's infosets; and (ii) at every visited infoset, it only requires a *local* closed-form update.

To achieve this, we exploit a technique of (Hoda et al., 2010; Farina et al., 2021). By focusing on the first player, we have that if the following two facts hold:

(i) the overall update can be expressed as a *proximal gradient update*; formally, as the computation of the conjugate gradient $\nabla d_1^{\epsilon,*}(\tilde{\boldsymbol{g}}) \coloneqq \arg\max_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{x}^\top \tilde{\boldsymbol{g}} - d_1^\epsilon(\boldsymbol{x}) \right\}$ of d_1^ϵ for some vector $\tilde{\boldsymbol{g}} \in \mathbb{R}^{|\Sigma_1|}$, and

(ii) for every information set $I \in \mathcal{I}_1$ and vector $\tilde{\boldsymbol{g}} \in \mathbb{R}^{n_I}$, the *local conjugate gradient* $\nabla d_{\Delta_I}^{\epsilon,*}(\tilde{\boldsymbol{g}}) \coloneqq \arg\max_{\boldsymbol{w} \in \Delta_I} \left\{ \boldsymbol{w}^\top \tilde{\boldsymbol{g}} - d_{\Delta_I}^\epsilon(\boldsymbol{w}) \right\}$ and *local gradient* $\nabla d_{\Delta_I}^\epsilon(\tilde{\boldsymbol{g}})$ have a closed-form expression,

then the overall update can be computed efficiently in terms by closed form expressions.

However, Algorithm 2 employs update rules that are *a priori* different from those studied in (Hoda et al., 2010) and its follow ups. Thankfully, the following theorem shows that the approach described above can still be employed on the update rules of Algorithm 2.

Theorem 5.5. *The following hold:*

(i) *For any pair of vectors $\boldsymbol{g}, \tilde{\boldsymbol{x}} \in \mathbb{R}^{|\Sigma_1|}$ and $\epsilon > 0$, the updates of Algorithm 2 of the form $\arg\max_{\boldsymbol{x} \in \mathcal{X}} \left\{ \boldsymbol{x}^\top \boldsymbol{g} - \frac{d_1^\epsilon(\boldsymbol{x})}{\lambda} - \frac{1}{\eta} D_1^\epsilon(\boldsymbol{x} | \tilde{\boldsymbol{x}}) \right\}$ can be formulated as the computation of the conjugate gradient $\nabla d_1^{\epsilon,*}(\tilde{\boldsymbol{g}})$ for a suitably-defined, efficiently-computable vector $\tilde{\boldsymbol{g}} \in \mathbb{R}^{|\Sigma_1|}$.*

(ii) *For every $\epsilon > 0$, $I \in \mathcal{I}_1$, and $\tilde{\boldsymbol{g}} \in \mathbb{R}^{n_I}$, $\nabla d_{\Delta_I}^{\epsilon,*}(\tilde{\boldsymbol{g}})[a] = (1 - \epsilon\, n_I) \frac{e^{\tilde{\boldsymbol{g}}[a]}}{\|e^{\tilde{\boldsymbol{g}}}\|_1} + \epsilon$, and the local gradient can be computed as $\nabla d_{\Delta_I}^\epsilon(\tilde{\boldsymbol{g}})[a] = 1 + \log(\tilde{\boldsymbol{g}}[a] - \epsilon)$ for all $a \in A(I)$.*

6 EXPERIMENTAL EVALUATION

We conclude by evaluating our algorithm on a standard testbed of EFGs. We consider two simplified

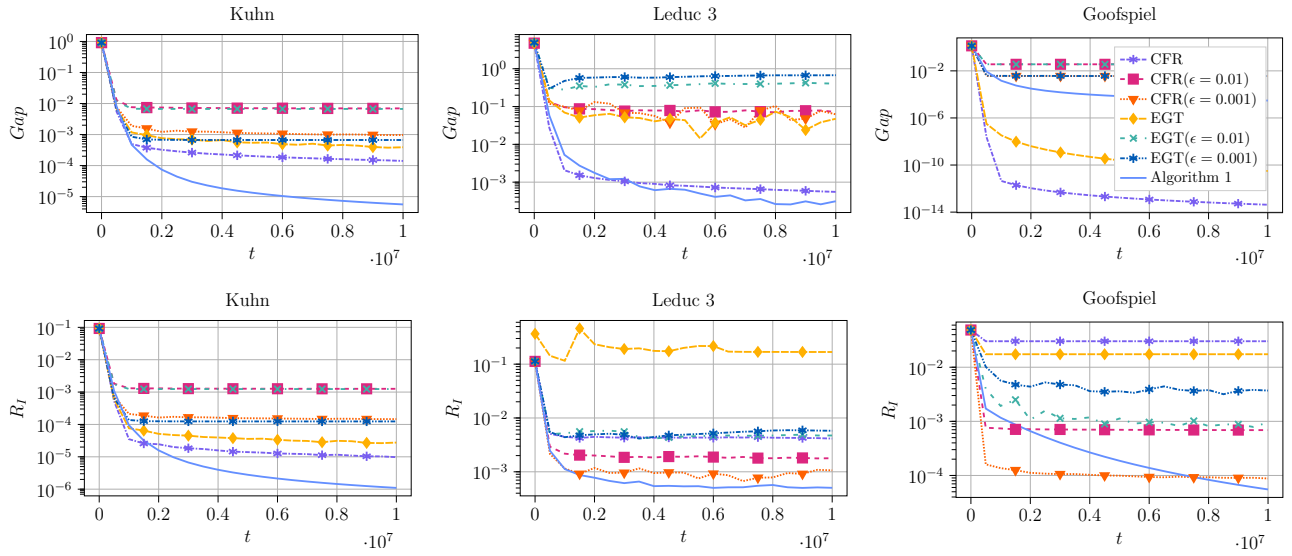

Figure 2: Results of the experimental evaluation. Algorithm 1 is compared with the baselines in terms of Nash gap (*First row*) and average infoset regret (*Second row*). CFR(ϵ) and EGT(ϵ) denote the methods by Farina et al. (2017) and Kroer et al. (2017), respectively, instantiated with perturbation ϵ.

versions of Poker, called *Kuhn* (Kuhn, 1950) and *Leduc* (Southey et al., 2005), and a two-player game called *Goofspiel* Ross (1971).[4]

We evaluate our algorithm in terms of two metrics. The first one is the *Nash gap*, which is the standard metric employed to evaluate NE approximation. The second metric is the *average infoset regret*, which we call R_I for short. Given a pair of players' strategies, this is defined as the average of the regrets that such strategies incur at all the infosets of the game. Specifically, the regret at an infoset is computed by assuming that such infoset is reached with probability one and by applying the *Bayes* rule to get a distribution encoding the probability of reaching each node in the infoset. Then, the regret is defined as the increase in utility by playing a best response to the opponent's strategy from that infoset on.

We compare Algorithm 1 with two baseline algorithms that are designed for computing NEs, namely the classical CFR algorithm (Zinkevich et al., 2007) and the *excessive gap technique* (EGT) algorithm for sequence-form strategy sets, as introduced by Farina et al. (2019). Moreover, we also compare the performances of Algorithm 1 with the extensions of such algorithms that have been designed for NE refinements. These are the iterative methods proposed by Farina et al. (2017) and Kroer et al. (2017), which take as input a fixed perturbation parameter $\epsilon > 0$, and work with the strategy space $\mathcal{X}^\epsilon \times \mathcal{Y}^\epsilon$. In particular, we test such algorithms for the perturbation values $\epsilon \in \{0.01, 0.001\}$. Moreover, we run Algorithm 1 with the following values of the parameters: $\beta = 1.001$, $\eta = 2$, $\epsilon_k = (1 - 1e^{-4})^k$, and $\lambda_k = 1/\epsilon_k^2$.

Figure 2 shows the main results of our experimental evaluation (additional ones are in Appendix C).[5] As it is clear from the plots, Algorithm 1 outperforms the others in terms of average infoset regret R_I. This was expected, since our algorithm is specifically tailored for converging to an EFPE, while the others only guarantee convergence to either an NE or an approximate EFPE. Notice that Algorithm 1 is also competitive with the other algorithms in terms of Nash gap. This shows that the additional guarantees provided by Algorithm 1 do *not* hinder the speed of decay of the Nash gap metric. Indeed, our algorithm surprisingly achieves the best performance in terms of Nash gap in the *Khun* and *Leduc* games.

Our results show the importance of tuning regularization and perturbation terms jointly in order to converge to exact EFPEs. Indeed, every method that finds approximate equilibria of the perturbed game provably fails in getting either zero Nash gap or zero average infoset regret. This shows that the regularization introduced into the game *not* only enables convergence to EFPEs, but it also has the potential of leading to the design of algorithms which are superior to current state-of-the-art equilibrium-computation algorithms.

[4] See Appendix C for a description of the games.

[5] For Algorithm 1, EGT, and EGT(ϵ) we considered the last iterates, while for CFR and CFR(ϵ) we considered the time average.

Acknowledgements

his paper is supported by the Italian MIUR PRIN 2022 Project "Targeted Learning Dynamics: Computing Efficient and Fair Equilibria through No-Regret Algorithms", by the FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence), and by the EU Horizon project ELIAS (European Lighthouse of AI for Sustainability, No. 101120237).

References

Kenshi Abe, Kaito Ariu, Mitsuki Sakamoto, Kentaro Toyoshima, and Atsushi Iwasaki. Last-iterate convergence with full and noisy feedback in two-player zero-sum games. In *International Conference on Artificial Intelligence and Statistics*, pages 7999–8028. PMLR, 2023.

Anton Bakhtin, Noam Brown, Emily Dinan, Gabriele Farina, Colin Flaherty, Daniel Fried, Andrew Goff, Jonathan Gray, Hengyuan Hu, Athul Paul Jacob, Mojtaba Komeili, Karthik Konath, Minae Kwon, Adam Lerer, Mike Lewis, Alexander H. Miller, Sasha Mitts, Adithya Renduchintala, Stephen Roller, Dirk Rowe, Weiyan Shi, Joe Spisak, Alexander Wei, David Wu, Hugh Zhang, and Markus Zijlstra. Human-level play in the game of diplomacy by combining language models with strategic reasoning. *Science*, 378(6624):1067–1074, 2022.

Noam Brown and Tuomas Sandholm. Superhuman ai for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374):418–424, 2018. ISSN 0036-8075.

Noam Brown and Tuomas Sandholm. Superhuman ai for multiplayer poker. *Science*, 365(6456):885–890, 2019.

Shicong Cen, Yuting Wei, and Yuejie Chi. Fast policy extragradient methods for competitive games with entropy regularization. *Advances in Neural Information Processing Systems*, 34:27952–27964, 2021.

Shicong Cen, Yuejie Chi, Simon S Du, and Lin Xiao. Faster last-iterate convergence of policy optimization in zero-sum markov games. *arXiv preprint arXiv:2210.01050*, 2022.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

John C Duchi, Shai Shalev-Shwartz, Yoram Singer, and Ambuj Tewari. Composite objective mirror descent. In *COLT*, volume 10, pages 14–26. Citeseer, 2010.

Gabriele Farina and Nicola Gatti. Extensive-form perfect equilibrium computation in two-player games. In *AAAI Conference on Artificial Intelligence*, pages 502–508, 2017.

Gabriele Farina and Tuomas Sandholm. Equilibrium refinement for the age of machines: The one-sided quasi-perfect equilibrium. *Advances in Neural Information Processing Systems*, 34:8845–8856, 2021.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Regret minimization in behaviorally-constrained zero-sum games. In *International Conference on Machine Learning*, pages 1107–1116, 2017.

Gabriele Farina, Nicola Gatti, and Tuomas Sandholm. Practical exact algorithm for trembling-hand equilibrium refinements in games. In *Advances in Neural Information Processing Systems*, pages 5039–5049, 2018a.

Gabriele Farina, Alberto Marchesi, Christian Kroer, Nicola Gatti, and Tuomas Sandholm. Trembling-hand perfection in extensive-form games with commitment. In *International Joint Conferences on Artificial Intelligence Organization*, pages 233–239, 2018b.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Optimistic regret minimization for extensive-form games via dilated distance-generating functions. *Advances in neural information processing systems*, 32, 2019.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Better regularization for sequential decision spaces: Fast convergence rates for nash, correlated, and team equilibria. *arXiv preprint arXiv:2105.12954*, 2021.

Nicola Gatti, Mario Gilli, and Alberto Marchesi. A characterization of quasi-perfect equilibria. *Games and Economic Behavior*, 122:240–255, 2020.

Kristoffer Arnsfelt Hansen, Peter Bro Miltersen, and Troels Bjerre Sørensen. The computational complexity of trembling hand perfection and other equilibrium refinements. In *International Symposium on Algorithmic Game Theory*, pages 198–209. Springer, 2010.

Samid Hoda, Andrew Gilpin, Javier Pena, and Tuomas Sandholm. Smoothing techniques for computing nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2):494–512, 2010.

Christian Kroer, Gabriele Farina, and Tuomas Sandholm. Smoothing method for approximate extensive-form perfect equilibrium. In *International Joint Conference on Artificial Intelligence*, pages 295–301, 2017.

HW Kuhn. Simplified two-person poker. *Contributions to the Theory of Games*, pages 97–103, 1950.

Chung-Wei Lee, Christian Kroer, and Haipeng Luo. Last-iterate convergence in extensive-form games. *Advances in Neural Information Processing Systems*, 34:14293–14305, 2021.

Mingyang Liu, Asuman Ozdaglar, Tiancheng Yu, and Kaiqing Zhang. The power of regularization in solving extensive-form games. *arXiv preprint arXiv:2206.09495*, 2022.

Alberto Marchesi and Nicola Gatti. Trembling-hand perfection and correlation in sequential games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5566–5574, 2021.

Alberto Marchesi, Gabriele Farina, Christian Kroer, Nicola Gatti, and Tuomas Sandholm. Quasi-perfect stackelberg equilibrium. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 2117–2124, 2019.

Richard D McKelvey and Thomas R Palfrey. Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1):6–38, 1995.

Peter Bro Miltersen and Troels Bjerre Sørensen. Computing a quasi-perfect equilibrium of a two-player game. *Economic Theory*, 42(1):175–192, 2010.

John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

Sasha Rakhlin and Karthik Sridharan. Optimization, learning, and games with predictable sequences. *Advances in Neural Information Processing Systems*, 26, 2013.

Sheldon M Ross. Goofspiel: The game of pure strategy. *Journal of Applied Probability*, 8(3):621–625, 1971.

Reinhard Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *International journal of game theory*, 4(1):25–55, 1975.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

Samuel Sokota, Ryan D'Orazio, J Zico Kolter, Nicolas Loizou, Marc Lanctot, Ioannis Mitliagkas, Noam Brown, and Christian Kroer. A unified approach to reinforcement learning, quantal response equilibria, and two-player zero-sum games. *arXiv preprint arXiv:2206.05825*, 2022.

Finnegan Southey, Michael Bowling, Bryce Larson, Carmelo Piccione, Neil Burch, Darse Billings, and Chris Rayner. Bayes' bluff: opponent modelling in poker. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, pages 550–558, 2005.

RC Steinlage. Nearly uniform convergence and interchange of limits. *Publications de l'Institut Mathématique*, 26(12):115–129, 1971.

Eric Van Damme. *Stability and perfection of Nash equilibria*, volume 339. Springer, 1991.

Bernhard Von Stengel. Efficient computation of behavior strategies. *Games and Economic Behavior*, 14(2):220–246, 1996.

Chen-Yu Wei, Chung-Wei Lee, Mengxiao Zhang, and Haipeng Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *International Conference on Learning Representations*, 2020.

Martin Zinkevich, Michael Johanson, Michael Bowling, and Carmelo Piccione. Regret minimization in games with incomplete information. *Advances in neural information processing systems*, 20, 2007.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model.
   Yes
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm.
   Yes
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries.
   No

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results.
   Yes
   (b) Complete proofs of all theoretical results.
   Yes
   (c) Clear explanations of any assumptions.
   Yes

3. For all figures and tables that present empirical results, check if you include:

(a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL).
No

(b) All the training details (e.g., data splits, hyperparameters, how they were chosen).
Yes

(c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times).
Yes

(d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider).
Yes

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:

(a) Citations of the creator If your work uses existing assets.
Not Applicable

(b) The license information of the assets, if applicable.
Not Applicable

(c) New assets either in the supplemental material or as a URL, if applicable.
Not Applicable

(d) Information about consent from data providers/curators.
Not Applicable

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content.
Not Applicable

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

(a) The full text of instructions given to participants and screenshots.
Not Applicable

(b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable.
Not Applicable

(c) The estimated hourly wage paid to participants and the total amount spent on participant compensation.
Not Applicable

Appendix Index

The appendix is structured as follows:

- Appendix A presents the proofs omitted from Section 4.
- Appendix B provides the proofs and additional details omitted from Section 5 in the construction of Algorithm 1.
- Appendix C provides additional experiments details and results.

A Proofs omitted from Section 4

Lemma A.1. *For every $I \in \mathcal{I}_1 \cup \mathcal{I}_2$ and $\epsilon > 0$, the function $d^{\epsilon}_{\Delta_I}$ is 1-strongly convex with respect to the Euclidean norm.*

Proof. For every $\boldsymbol{w} \in \Delta_I$, we have that:

$$\frac{\partial^2 d^{\epsilon}_{\Delta_I}(\boldsymbol{w})}{\partial \boldsymbol{w}[i] \partial \boldsymbol{w}[j]} = \begin{cases} \frac{1}{\boldsymbol{w}[i] - \epsilon} & \text{if} \quad i = j \\ 0 & \text{otherwise} \end{cases}.$$

Thus, $\nabla^2 d^{\epsilon}_{\Delta_I}(\boldsymbol{w}) \geq \boldsymbol{I}_{n_I}$, where $\boldsymbol{I}_n$ denotes the n-dimensional identity matrix. □

Lemma A.2. *For any $\epsilon > 0$, the functions d^{ϵ}_1 and d^{ϵ}_2 are 1-strongly convex with respect to the Euclidean norm.*

Proof. This follows from Lemma A.1, the expression of the weights α_I in $d^{\epsilon}_{\Delta_I}$ in the definitions of d^{ϵ}_1 and d^{ϵ}_2, and by using Farina et al. (2019, Corollary 1). □

Lemma A.3. *We have that $|d^{\epsilon}_1(\boldsymbol{x})| \leq C$ and $|d^{\epsilon}_2(\boldsymbol{y})| \leq C$ for all $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X} \times \mathcal{Y}$ and any $\epsilon \leq \min_{I \in \mathcal{I}_1 \cup \mathcal{I}_2} \frac{1}{2n_I}$. Moreover $C \leq \|\boldsymbol{\alpha}\|_\infty \cdot \max_{i \in \{1,2\}} |I_i| \max_{I \in \mathcal{I}_1 \cup \mathcal{I}_2} \log(2n_I)$, where $\boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{I}_1| + |\mathcal{I}_2|}$ is a vector that contains all the components α_I with $I \in \mathcal{I}_1 \cup \mathcal{I}_2$.*

Proof. Consider the following inequalities:

$$\begin{aligned} \max_{\boldsymbol{x} \in \mathcal{X}} |d^{\epsilon}_1(\boldsymbol{x})| &\leq \sum_{I \in \mathcal{I}_1} \alpha_I \max_{\boldsymbol{w} \in \Delta_I} |d^{\epsilon}_{\Delta_I}(\boldsymbol{w})| \\ &\leq \sum_{I \in \mathcal{I}_1} \alpha_I \log\left(\frac{n_I}{1 - \epsilon \cdot n_I}\right) \\ &\leq \sum_{I \in \mathcal{I}_1} \alpha_I \log(2n_I) := C_{\mathcal{X}}, \end{aligned}$$

where we used that the maximum of $d^{\epsilon}_{\Delta_I}(\boldsymbol{w})$ is attained in the center $\boldsymbol{w}[a] = 1/n_I$, for all $a \in A(I)$, and the last inequality follows from $\epsilon \leq \min_{I \in \mathcal{I}_1 \cup \mathcal{I}_2} \frac{1}{2n_I}$. We can define $C_{\mathcal{Y}}$ analogously for the second player and take $C := \max\{C_{\mathcal{X}}, C_{\mathcal{Y}}\}$.

Then define the vector $\boldsymbol{\alpha}$ as the vector that contains the components α_I for the first player *and* second player. Then, by Holder's inequality, it holds $C \leq \|\boldsymbol{\alpha}\|_\infty \cdot \max_{i \in \{1,2\}} |I_i| \max_{I \in \mathcal{I}_1 \cup \mathcal{I}_2} \log(2n_I)$. □

Lemma A.4. *The following inequalities holds:*

$$\left| \max_{\boldsymbol{x}} f(\boldsymbol{x}) - \max_{\boldsymbol{y}} g(\boldsymbol{y}) \right| \leq \max_{\boldsymbol{x}} |f(\boldsymbol{x}) - g(\boldsymbol{x})|, \tag{1}$$

$$\left| \min_{\boldsymbol{x}} f(\boldsymbol{x}) - \min_{\boldsymbol{y}} g(\boldsymbol{y}) \right| \leq \max_{\boldsymbol{x}} |f(\boldsymbol{x}) - g(\boldsymbol{x})|. \tag{2}$$

Proof. Consider the following inequality:

$$f(\boldsymbol{x}) \leq |f(\boldsymbol{x}) - g(\boldsymbol{x})| + g(\boldsymbol{x})$$

Applying the max operator to both sides of the previous equation and observing that $\max_{\boldsymbol{x}}(|f(\boldsymbol{x})-g(\boldsymbol{x})|+g(\boldsymbol{x})) \leq \max_{\boldsymbol{x}} |f(\boldsymbol{x}) - g(\boldsymbol{x})| + \max_{\boldsymbol{x}} g(\boldsymbol{x}))$ we can rearrange the inequality to obtain:

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}) - \max_{\boldsymbol{x}} g(\boldsymbol{x}) \leq \max_{\boldsymbol{x}} |f(\boldsymbol{x}) - g(\boldsymbol{x})|.$$

We can in the same way obtain that:

$$\max_{\boldsymbol{x}} g(\boldsymbol{x}) - \max_{\boldsymbol{x}} f(\boldsymbol{x}) \leq \max_{\boldsymbol{x}} |f(\boldsymbol{x}) - g(\boldsymbol{x})|,$$

which combined with the above let us conclude Equation (1). Now Equation (2) follows from considering Equation (1) with $-f(\boldsymbol{x})$ instead of $f(\boldsymbol{x})$ and $-g(\boldsymbol{x})$ instead of $g(\boldsymbol{x})$. □

Theorem 4.1. *Given an EFG and two parameters $\lambda, \epsilon > 0$, the unique NE $\boldsymbol{z}^\star_{\lambda,\epsilon}$ of $\mathcal{G}(\lambda, \epsilon)$ constitutes a $O\left(\frac{1}{\lambda}\right)$-approximate ϵ-EFPE of the EFG.*

Proof. Let us consider the function $\tilde{f}_\epsilon : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ such that, for every $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{y} \in \mathcal{Y}$, it holds:

$$\tilde{f}_\epsilon(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \boldsymbol{x}^\top \boldsymbol{U} \boldsymbol{y} - \mathbb{I}_{\mathcal{X}^\epsilon}(\boldsymbol{x}) + \mathbb{I}_{\mathcal{Y}^\epsilon}(\boldsymbol{y}),$$

where $\mathbb{I}_{\mathcal{W}}(\boldsymbol{w})$ is 0 if $\boldsymbol{w} \in \mathcal{W}$, while it is $+\infty$ if $\boldsymbol{w} \notin \mathcal{W}$.

First, it is easy to check that any $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$ such that $\tilde{f}_\epsilon(\boldsymbol{x}, \boldsymbol{y}) = \max_{\tilde{\boldsymbol{x}} \in \mathcal{X}} \min_{\tilde{\boldsymbol{y}} \in \mathcal{Y}} \tilde{f}_\epsilon(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}})$ is an ϵ-EFPE of the original EFG (see Definition 3.1).

Moreover, for any $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X}^\epsilon \times \mathcal{Y}^\epsilon$ it is easy to show that $|f_{\lambda,\epsilon} - \tilde{f}_\epsilon| = O(1/\lambda)$:

$$|f_{\lambda,\epsilon}(\boldsymbol{x}, \boldsymbol{y}) - \tilde{f}_\epsilon(\boldsymbol{x}, \boldsymbol{y})| \leq \frac{|d_1^\epsilon(\boldsymbol{x}) - d_2^\epsilon(\boldsymbol{y})|}{\lambda} \leq \frac{2C}{\lambda},$$

where C is defined in Lemma A.3.

Finally, let us consider the unique NE $\boldsymbol{z}^\star_{\lambda,\epsilon} = (\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}^\star_{\lambda,\epsilon})$ of $\mathcal{G}(\lambda, \epsilon)$. By Lemma A.4, we have that:

$$\begin{aligned}
\max_{\boldsymbol{x} \in \mathcal{X}^\epsilon} \tilde{f}_\epsilon(\boldsymbol{x}, \boldsymbol{y}^\star_{\lambda,\epsilon}) - \min_{\boldsymbol{y} \in \mathcal{Y}^\epsilon} \tilde{f}_\epsilon(\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}) &\leq \left| \max_{\boldsymbol{x} \in \mathcal{X}^\epsilon} \tilde{f}_\epsilon(\boldsymbol{x}, \boldsymbol{y}^\star_{\lambda,\epsilon}) - \max_{\boldsymbol{x} \in \mathcal{X}^\epsilon} f_{\lambda,\epsilon}(\boldsymbol{x}, \boldsymbol{y}^\star_{\lambda,\epsilon}) \right| \\
&\quad + \left| \min_{\boldsymbol{y} \in \mathcal{Y}^\epsilon} \tilde{f}_\epsilon(\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}) - \min_{\boldsymbol{y} \in \mathcal{Y}^\epsilon} f_{\lambda,\epsilon}(\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}) \right| \\
&\leq \frac{4C}{\lambda},
\end{aligned}$$

which concludes the proof. □

Proposition 4.1. *There exists an EFG for which $\lim_{\lambda \to \infty} \lim_{\epsilon \to 0} \boldsymbol{z}^\star_{\lambda,\epsilon} \neq \lim_{\epsilon \to 0} \lim_{\lambda \to \infty} \boldsymbol{z}^\star_{\lambda,\epsilon}$ and, additionally, $\lim_{\lambda \to \infty} \lim_{\epsilon \to 0} \boldsymbol{z}^\star_{\lambda,\epsilon}$ is not an EFPE.*

Proof. Let us consider a game in which the two players play simultaneously and only once, having each of them three different actions available (*i.e.*, a 3×3 game in normal form). The following matrix encodes the first player's payoffs for all the possible combinations of players' actions:

$$\begin{bmatrix} 0.3 & 0.5 & 0.3 \\ 0.7 & 0.3 & 0.7 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}$$

By using the fact that any quantal equilibrium enjoys the "independence of irrelevant alternatives" property McKelvey and Palfrey (1995), we can prove that it holds:

$$\lim_{\lambda \to \infty} \left[\lim_{\epsilon \to 0} \boldsymbol{z}^\star_{\lambda,\epsilon} \right] = \left(\begin{bmatrix} 2/3 \\ 1/3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/6 \\ 2/3 \\ 1/6 \end{bmatrix} \right),$$

where we expressed player's strategies as the probability distributions that they induce over the three actions for ease of presentation. Moreover, the only EFPE of the game is:

$$\boldsymbol{z}^\star = \left(\begin{bmatrix} 2/3 \\ 1/3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2/3 \\ 1/3 \end{bmatrix} \right),$$

since this is the unique NE that eliminates weakly dominated strategies, and, thus, it is an (Van Damme, 1991, Corollary 2.2.6). This proves the proposition. □

Theorem 4.3. *Given any EFG, there exists a function $\mu : (0,1] \to \mathbb{R}_+$ such that, given any sequence $\{(\lambda_k, \epsilon_k)\}_{k\in\mathbb{N}}$ defined so that $\lambda_k \to +\infty$, $\epsilon_k \to 0$, and $1/\lambda_k < \mu(\epsilon_k)$ as $k \to +\infty$, the limit $\lim_{k\to+\infty} \boldsymbol{z}^\star_{\lambda_k,\epsilon_k}$ identifies an EFPE of the given EFG.*

Proof. By Corollary 4.2, we have that

$$\boldsymbol{z}^\star \coloneqq \lim_{\epsilon\to 0} \left[\lim_{\lambda\to\infty} \boldsymbol{z}^\star_{\lambda,\epsilon} \right]$$

is an EFPE. Let $\boldsymbol{z}^\star_\epsilon \coloneqq \lim_{\lambda\to\infty} \boldsymbol{z}^\star_{\lambda,\epsilon}$ for all $\epsilon > 0$. By definition of limit, for every $\tau > 0$ there exists $R_\epsilon(\tau) \in \mathbb{R}_+$ such that $||\boldsymbol{z}^\star_{\lambda,\epsilon} - \boldsymbol{z}^\star_\epsilon||_2 \leq \tau$ for all $\lambda \in \mathbb{R} : \lambda > R_\epsilon(\tau)$. Moreover, by looking at the outer limit in the definition of $\boldsymbol{z}^\star$ we have that, for every $\tau' > 0$, there exists $H(\tau') \in \mathbb{R}_+$ such that $||\boldsymbol{z}^\star_\epsilon - \boldsymbol{z}^\star||_2 \leq \tau'$ for all $\epsilon \in \mathbb{R} : |\epsilon| \leq H(\tau')$.

By using the triangular inequality, for every $\tau > 0$:

$$||\boldsymbol{z}^\star_{\epsilon,\lambda} - \boldsymbol{z}^\star||_2 \leq ||\boldsymbol{z}^\star_{\epsilon,\lambda} - \boldsymbol{z}^\star_\epsilon||_2 + ||\boldsymbol{z}^\star_\epsilon - \boldsymbol{z}^\star||_2 \leq \tau$$

for all $\lambda \in \mathbb{R} : \lambda > R_\epsilon(\tau/2)$ and $\epsilon \in \mathbb{R} : |\epsilon| \leq H(\tau/2)$.

Moreover, if $\epsilon_k \leq H(\tau/2)$, then $\lambda_k > R_{\epsilon_k}(\tau/2)$. This follows from the following inequalities:

$$\lambda_k > \frac{1}{\mu(\epsilon_k)} \geq \frac{1}{\mu(H(\tau/2))} \geq R_{\epsilon_k}(\tau/2),$$

which holds for a suitable choice of the function $\mu : (0,1] \to \mathbb{R}_+$.

As a result, for every $\tau > 0$ and for all $k \in \mathbb{N}$ such that $\epsilon_k \leq H(\tau/2)$, it holds:

$$||\boldsymbol{z}^\star_{\epsilon_k,\lambda_k} - \boldsymbol{z}^\star||_2 \leq \tau,$$

which concludes the proof. □

B Proofs omitted from Section 5

In this section is convenient to consider the join updates for the first and second player. If we define $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{Y}$, then it is easy to verify that updates of the form:

$$\hat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \left\{ \boldsymbol{x}^\top \boldsymbol{U} \boldsymbol{y}_0 - \frac{1}{\lambda} d_1^\epsilon(\boldsymbol{x}) - \frac{1}{\eta} D_1^\epsilon(\boldsymbol{x}|\tilde{\boldsymbol{x}}) \right\} \tag{3}$$

$$\hat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}\in\mathcal{Y}} \left\{ \boldsymbol{x}_0^\top \boldsymbol{U} \boldsymbol{y} + \frac{1}{\lambda} d_2^\epsilon(\boldsymbol{y}) + \frac{1}{\eta} D_2^\epsilon(\boldsymbol{y}|\tilde{\boldsymbol{y}}) \right\}, \tag{4}$$

can be jointly expressed as:

$$\hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}\in\mathcal{X}\times\mathcal{Y}} \left\{ G(\boldsymbol{z}_0)^\top \boldsymbol{z} + \frac{1}{\lambda} H^\epsilon(\boldsymbol{z}) + \frac{1}{\eta} D^\epsilon(\boldsymbol{z}|\tilde{\boldsymbol{z}}) \right\},$$

where we define $H^\epsilon(\boldsymbol{z}) := d_1^\epsilon(\boldsymbol{x}) + d_2^\epsilon(\boldsymbol{y})$, $D^\epsilon(\boldsymbol{z}|\tilde{\boldsymbol{z}}) := D_1^\epsilon(\boldsymbol{x}|\tilde{\boldsymbol{x}}) + D_2^\epsilon(\boldsymbol{y}|\tilde{\boldsymbol{y}})$ and $G(\boldsymbol{z}) := (-\boldsymbol{U}\boldsymbol{y}, \boldsymbol{U}^\top \boldsymbol{x})$.

B.1 Convergence Analysis

Lemma B.1. *Let $\boldsymbol{z}^\star_{\lambda,\epsilon}$ the unique NE of the game $\mathcal{G}_{\lambda,\epsilon}$, then for all $\boldsymbol{z}$:*

$$G(\boldsymbol{z})^\top(\boldsymbol{z}^\star_{\lambda,\epsilon} - \boldsymbol{z}) + \frac{1}{\lambda}\left[H^\epsilon(\boldsymbol{z}^\star_{\lambda,\epsilon}) - H^\epsilon(\boldsymbol{z})\right] \leq 0$$

Proof. Since $\boldsymbol{z}^\star_{\lambda,\epsilon}$ is an equilibrium of $\mathcal{G}_{\lambda,\epsilon}$ we have that:

$$f_{\lambda,\epsilon}(\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}^\star_{\lambda,\epsilon}) \leq f_{\lambda,\epsilon}(\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}), \quad \forall \boldsymbol{y} \in \mathcal{Y}$$

which implies that:

$$\boldsymbol{x}^{\star,\top}_{\lambda,\epsilon}\boldsymbol{U}(\boldsymbol{y}^\star_{\lambda,\epsilon} - \boldsymbol{y}) + \frac{1}{\lambda}\left[d^\epsilon_2(\boldsymbol{y}^\star_{\lambda,\epsilon}) - d^\epsilon_2(\boldsymbol{y})\right] \leq 0. \tag{5}$$

Similarly we can observe that:

$$f_{\lambda,\epsilon}(\boldsymbol{x}, \boldsymbol{y}^\star_{\lambda,\epsilon}) \leq f_{\lambda,\epsilon}(\boldsymbol{x}^\star_{\lambda,\epsilon}, \boldsymbol{y}^\star_{\lambda,\epsilon}), \quad \forall \boldsymbol{x} \in \mathcal{X}$$

which implies that:

$$(\boldsymbol{x} - \boldsymbol{x}^\star_{\lambda,\epsilon})^\top\boldsymbol{U}\boldsymbol{y}^\star_{\lambda,\epsilon} + \frac{1}{\lambda}\left[d^\epsilon_1(\boldsymbol{x}^\star_{\lambda,\epsilon}) - d^\epsilon_1(\boldsymbol{x})\right] \leq 0. \tag{6}$$

By summing Equation (5) and Equation (6) and observing that $\boldsymbol{x}^\top\boldsymbol{U}\boldsymbol{y}^\star_{\lambda,\epsilon} - \boldsymbol{x}^{\star,\top}_{\lambda,\epsilon}\boldsymbol{U}\boldsymbol{y} = G(\boldsymbol{z})^\top(\boldsymbol{z}^\star_{\lambda,\epsilon} - \boldsymbol{z})$ we can conclude the statement of the lemma. □

Lemma B.2. *If $\hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}\in\mathcal{X}\times\mathcal{Y}}\left\{G(\boldsymbol{z}_0)^\top\boldsymbol{z} + \frac{1}{\lambda}H^\epsilon(\boldsymbol{z}) + \frac{1}{\eta}D^\epsilon(\boldsymbol{z}|\tilde{\boldsymbol{z}})\right\}$ then:*

$$\eta(\hat{\boldsymbol{z}} - \boldsymbol{z})^\top G(\boldsymbol{z}_0) + \frac{\eta}{\lambda}H^\epsilon(\hat{\boldsymbol{z}}) - \frac{\eta}{\lambda}H^\epsilon(\boldsymbol{z}) \leq D^\epsilon(\boldsymbol{z}|\tilde{\boldsymbol{z}}) - D^\epsilon(\hat{\boldsymbol{z}}|\tilde{\boldsymbol{z}}) - \frac{\lambda+\eta}{\lambda}D^\epsilon(\boldsymbol{z}|\hat{\boldsymbol{z}}) \tag{7}$$

Proof. By the fact that for any Bregman divergence we have $\nabla_{\boldsymbol{z}}D^\epsilon(\boldsymbol{z}|\tilde{\boldsymbol{z}}) = \nabla H^\epsilon(\boldsymbol{z}) - \nabla H^\epsilon(\tilde{\boldsymbol{z}})$, and rearranging the optimality condition of $\hat{\boldsymbol{z}}$ we get that:

$$\left[G(\boldsymbol{z}_0) + \left(\frac{1}{\lambda} + \frac{1}{\eta}\right)\nabla H^\epsilon(\hat{\boldsymbol{z}}) - \frac{1}{\eta}\nabla H^\epsilon(\tilde{\boldsymbol{z}})\right]^\top(\boldsymbol{z} - \hat{\boldsymbol{z}}) \geq 0,$$

which implies that:

$$\frac{\lambda\eta}{\lambda+\eta}G(\boldsymbol{z}_0)^\top(\hat{\boldsymbol{z}} - \boldsymbol{z}) \leq \left[\nabla H^\epsilon(\hat{\boldsymbol{z}}) - \frac{\lambda}{\lambda+\eta}\nabla H^\epsilon(\tilde{\boldsymbol{z}})\right]^\top(\boldsymbol{z} - \hat{\boldsymbol{z}}). \tag{8}$$

It is then easy to check that for the right hand side of Equation (8) the following equality holds:

$$\begin{aligned}\left[\nabla H^\epsilon(\hat{\boldsymbol{z}}) - \frac{\lambda}{\lambda+\eta}\nabla H^\epsilon(\tilde{\boldsymbol{z}})\right]^\top(\boldsymbol{z} - \hat{\boldsymbol{z}}) = \frac{\lambda}{\lambda+\eta}D^\epsilon(\boldsymbol{z}|\tilde{\boldsymbol{z}}) - \frac{\lambda}{\lambda+\eta}D^\epsilon(\hat{\boldsymbol{z}}|\tilde{\boldsymbol{z}}) + D^\epsilon(\boldsymbol{z}|\hat{\boldsymbol{z}}) \\ + \left[\frac{\lambda}{\lambda+\eta} - 1\right]H^\epsilon(\hat{\boldsymbol{z}}) - \left[\frac{\lambda}{\lambda+\eta} - 1\right]H^\epsilon(\boldsymbol{z}),\end{aligned}$$

which concludes the proof by rearranging the terms. □

Next we prove a similar result to Wei et al. (2020, Lemma 11) that relates the distance between the gradients of two updates to the distance in the outputs. Formally:

Lemma B.3. *Let*

$$\hat{\boldsymbol{z}}_1 = \arg\min_{\boldsymbol{z}\in\mathcal{X}\times\mathcal{Y}}\left\{G(\boldsymbol{z}_{0,1})^\top\boldsymbol{z} + \frac{1}{\lambda}H^\epsilon(\boldsymbol{z}) + \frac{1}{\eta}D^\epsilon(\boldsymbol{z} \mid \boldsymbol{z}_1)\right\}$$

and

$$\hat{\boldsymbol{z}}_2 = \arg\min_{\boldsymbol{z}\in\mathcal{X}\times\mathcal{Y}}\left\{G(\boldsymbol{z}_{0,2})^\top\boldsymbol{z} + \frac{1}{\lambda}H^\epsilon(\boldsymbol{z}) + \frac{1}{\eta}D^\epsilon(\boldsymbol{z} \mid \boldsymbol{z}_1)\right\}.$$

Then:

$$\|\hat{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}_2\|_2 \leq \frac{\eta\lambda}{\eta+\lambda}\|G(\boldsymbol{z}_{0,1}) - G(\boldsymbol{z}_{0,2})\|_2.$$

Proof. By summing the first order conditions of the two equations above we obtain:

$$\eta \left[G(\boldsymbol{z}_{0,1}) - G(\boldsymbol{z}_{0,2})\right]^\top (\hat{\boldsymbol{z}}_2 - \hat{\boldsymbol{z}}_1) \geq \left(1 + \frac{\eta}{\lambda}\right) \left[\nabla H^\epsilon(\hat{\boldsymbol{z}}_1) - \nabla H^\epsilon(\hat{\boldsymbol{z}}_2)\right]^\top (\hat{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}_2). \tag{9}$$

By strong convexity of $H^\epsilon(\cdot)$, given by Lemma A.1, the right hand side is lower bounded by $\left(1 + \frac{\eta}{\lambda}\right) \|\hat{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}_2\|_2^2$, while by Cauchy-Schwartz inequality we have that the wight hand side is upper bounded by $\eta \|G(\boldsymbol{z}_{0,1}) - G(\boldsymbol{z}_{0,2})\|_2 \|\hat{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}_1\|_2$. Thus:

$$\eta \|G(\boldsymbol{z}_{0,1}) - G(\boldsymbol{z}_{0,2})\|_2 \|\hat{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}_1\|_2 \geq \left(1 + \frac{\eta}{\lambda}\right) \|\hat{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}_2\|_2^2,$$

which, after dividing by $\|\hat{\boldsymbol{z}}_1 - \hat{\boldsymbol{z}}_2\|_2$, concludes the proof. □

Theorem 5.1. *Define* $D^{\epsilon_k}(\boldsymbol{z}|\tilde{\boldsymbol{z}}) := D_1^{\epsilon_k}(\boldsymbol{x}|\tilde{\boldsymbol{x}}) + D_2^{\epsilon_k}(\boldsymbol{y}|\tilde{\boldsymbol{y}})$*, then for all* $\eta \leq 1/\sqrt{2}\|\boldsymbol{U}\|_2$*,* $k \in \mathbb{N}$ *and round* $t \in \{1, \dots, T_k\}$*, Algorithm 2 guarantees:*

$$\|\boldsymbol{z}_k^\star - \boldsymbol{z}_t^{(k)}\|_2 \leq 2 \left[2D^{\epsilon_k}\left(\boldsymbol{z}_k^\star | \boldsymbol{z}_0^{(k)}\right)\right]^{1/2} \left(\frac{\lambda_k}{\lambda_k + \eta}\right)^{t/2}.$$

Proof. Notice that the update of Algorithm 1 can be jointly written as:

$$\begin{aligned}
\boldsymbol{z}_{t+1}^{(k)} &= \arg\min_{\boldsymbol{z}} \left\{ \boldsymbol{z}^\top G(\boldsymbol{z}_t^{(k)}) + \frac{1}{\lambda_k} H^{\epsilon_k}(\boldsymbol{z}) + \frac{1}{\lambda_k} D^{\epsilon_k}\left(\boldsymbol{z} | \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}\right) \right\}, \\
\boldsymbol{z}_{t+\frac{1}{2}}^{(k)} &= \arg\min_{\boldsymbol{z}} \left\{ \boldsymbol{z}^\top G(\boldsymbol{z}_t^{(k)}) + \frac{1}{\lambda_k} H^{\epsilon_k}(\boldsymbol{z}) + \frac{1}{\lambda_k} D^{\epsilon_k}\left(\boldsymbol{z} | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}\right) \right\}.
\end{aligned}$$

Then we can use Lemma B.2 for the update of $\boldsymbol{z}_{t+\frac{1}{2}}^{(k)}$ which gives:

$$\begin{aligned}
\eta(\boldsymbol{z}_{t+\frac{1}{2}}^{(k)} - \boldsymbol{z})^\top G(\boldsymbol{z}_t) + \frac{\eta}{\lambda_k} H^{\epsilon_k}(\boldsymbol{z}_{t+\frac{1}{2}}^{(k)}) - \frac{\eta}{\lambda_k} H^{\epsilon_k}(\boldsymbol{z}) \leq \\
D^{\epsilon_k}(\boldsymbol{z} | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) - D^{\epsilon_k}(\boldsymbol{z}_{t+\frac{1}{2}}^{(k)} | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) - \frac{\lambda_k + \eta}{\lambda_k} D^{\epsilon_k}(\boldsymbol{z} | \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}),
\end{aligned} \tag{10}$$

which holds for any $\boldsymbol{z}$ and thus also for $\boldsymbol{z}_k^\star$.

On the other hand using Lemma B.2 for the update of $\boldsymbol{z}_t^{(k)}$ which gives:

$$\begin{aligned}
\eta(\boldsymbol{z}_t^{(k)} - \boldsymbol{z})^\top G(\boldsymbol{z}_{t-1}^{(k)}) + \frac{\eta}{\lambda_k} H^{\epsilon_k}(\boldsymbol{z}_t^{(k)}) - \frac{\eta}{\lambda_k} H^{\epsilon_k}(\boldsymbol{z}) \leq \\
D^{\epsilon_k}(\boldsymbol{z} | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) - D^{\epsilon_k}(\boldsymbol{z}_t^{(k)} | \boldsymbol{z}_{t-\frac{1}{2}}) - \frac{\lambda_k + \eta}{\lambda_k} D^{\epsilon_k}(\boldsymbol{z} | \boldsymbol{z}_t^{(k)}),
\end{aligned} \tag{11}$$

which holds fro any $\boldsymbol{z}$ and thus also for $\boldsymbol{z}_{t+\frac{1}{2}}^{(k)}$. Summing Equation (10) with $\boldsymbol{z} = \boldsymbol{z}_k^\star$ to Equation (11) with $\boldsymbol{z} = \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}$, and rearranging results in the following inequality:

$$\begin{aligned}
&\eta(\boldsymbol{z}_t^{(k)} - \boldsymbol{z}^\star)^\top G(\boldsymbol{z}_t^{(k)}) + \frac{\eta}{\lambda_k} \left[H^{\epsilon_k}(\boldsymbol{z}_t^{(k)}) - H^{\epsilon_k}(\boldsymbol{z}^\star)\right] \\
&\quad \leq D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) - \frac{\lambda_k + \eta}{\lambda_k} D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}) - D^{\epsilon_k}(\boldsymbol{z}_t^{(k)} | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) \\
&\qquad - \frac{\lambda_k + \eta}{\lambda_k} D^{\epsilon_k}(\boldsymbol{z}_{t+\frac{1}{2}}^{(k)} | \boldsymbol{z}_t^{(k)}) + \eta(\boldsymbol{z}_t^{(k)} - \boldsymbol{z}_{t+\frac{1}{2}}^{(k)})^\top (G(\boldsymbol{z}_t^{(k)}) - G(\boldsymbol{z}_{t-1}^{(k)}))
\end{aligned} \tag{12}$$

Now consider the last term of Equation (12) and the following inequalities:

$$\begin{aligned}
(\boldsymbol{z}_t^{(k)} - \boldsymbol{z}_{t+\frac{1}{2}}^{(k)})^\top (G(\boldsymbol{z}_t^{(k)}) - G(\boldsymbol{z}_{t-1}^{(k)})) &\leq \|\boldsymbol{z}_t^{(k)} - \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}\|_2 \|G(\boldsymbol{z}_t^{(k)}) - G(\boldsymbol{z}_{t-1}^{(k)})\|_2 \\
&\leq \frac{\lambda_k \eta}{\lambda_k + \eta} \|G(\boldsymbol{z}_{t-1}^{(k)}) - G(\boldsymbol{z}_t^{(k)})\|_2^2
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{\lambda_k \eta}{\lambda_k + \eta} L_{\boldsymbol{U}}^2 \|\boldsymbol{z}_{t-1}^{(k)} - \boldsymbol{z}_t^{(k)}\|_2^2 \\
&\leq \frac{2\lambda_k \eta}{\lambda_k + \eta} L_{\boldsymbol{U}}^2 \left[\|\boldsymbol{z}_{t-1}^{(k)} - \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}\|_2^2 + \|\boldsymbol{z}_{t-\frac{1}{2}}^{(k)} - \boldsymbol{z}_{t-1}^{(k)}\|_2^2\right] \\
&\leq \frac{4\lambda_k \eta}{\lambda_k + \eta} L_{\boldsymbol{U}}^2 \left[D^{\epsilon_k}(\boldsymbol{z}_t^{(k)} | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) + D^{\epsilon_k}(\boldsymbol{z}_{t-\frac{1}{2}}^{(k)} | \boldsymbol{z}_{t-1}^{(k)})\right],
\end{aligned}
$$

where the first inequality is the Cauchy Schwartz inequality and the second inequality follows from Lemma B.3. Then we used the fact that the operator $G(\cdot)$ is linear with matrix $\boldsymbol{A}$ defined as:

$$
\boldsymbol{A} = \begin{bmatrix} \boldsymbol{0} & -\boldsymbol{U}^\top \\ \boldsymbol{U} & \boldsymbol{0} \end{bmatrix},
$$

and $L_{\boldsymbol{U}} \coloneqq \|\boldsymbol{U}\|_2 = \|\boldsymbol{A}\|_2$. Thus, continuing from Equation (12) we get:

$$
\begin{aligned}
&\eta(\boldsymbol{z}_t^{(k)} - \boldsymbol{z}_k^\star)^\top G(\boldsymbol{z}_t^{(k)}) + \frac{\eta}{\lambda_k}\left[H^{\epsilon_k}(\boldsymbol{z}_t^{(k)}) - H^{\epsilon_k}(\boldsymbol{z}_k^\star)\right] \\
&\leq D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) - \frac{\lambda_k + \eta}{\lambda_k} D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}) - D^{\epsilon_k}(\boldsymbol{z}_t^{(k)} | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) - \frac{\lambda_k + \eta}{\lambda_k} D^{\epsilon_k}(\boldsymbol{z}_{t+\frac{1}{2}}^{(k)} | \boldsymbol{z}_t^{(k)}) \\
&+ \frac{4\lambda_k \eta^2}{\lambda_k + \eta} L_{\boldsymbol{U}}^2 \left[D^{\epsilon_k}(\boldsymbol{z}_t^{(k)} | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) + D^{\epsilon_k}(\boldsymbol{z}_{t-\frac{1}{2}}^{(k)} | \boldsymbol{z}_{t-1}^{(k)})\right]
\end{aligned}
$$

By assuming that $\lambda_k \geq \eta$ and $\eta \leq \frac{1}{\sqrt{2} L_{\boldsymbol{U}}}$ we have that $\frac{4\lambda_k \eta^2}{\lambda_k + \eta} L_{\boldsymbol{U}}^2 \leq 1$. Moreover, thanks to Lemma B.1 we have that:

$$
\eta(\boldsymbol{z}_t^{(k)} - \boldsymbol{z}_k^\star)^\top G(\boldsymbol{z}_t^{(k)}) + \frac{\eta}{\lambda_k}\left[H^{\epsilon_k}(\boldsymbol{z}_t^{(k)}) - H^{\epsilon_k}(\boldsymbol{z}_k^\star)\right] \geq 0.
$$

Thus, by rearranging Equation(12) and recalling that $D^{\epsilon_k}(\cdot|\cdot) \geq 0$ we obtain:

$$
\frac{\lambda_k + \eta}{\lambda_k}\left[D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}) + D^{\epsilon_k}(\boldsymbol{z}_{t+\frac{1}{2}}^{(k)} | \boldsymbol{z}_t^{(k)})\right] \leq D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_{t-\frac{1}{2}}^{(k)}) + D^{\epsilon_k}(\boldsymbol{z}_{t-\frac{1}{2}}^{(k)} | \boldsymbol{z}_{t-1}^{(k)}).
$$

Thus, by iterating the above expression, we get that:

$$
D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}) + D^{\epsilon_k}(\boldsymbol{z}_{t+\frac{1}{2}}^{(k)} | \boldsymbol{z}_t^{(k)}) \leq \left(\frac{\lambda_k}{\lambda_k + \eta}\right)^t D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_0^{(k)}),
$$

and since $D^{\epsilon_k}(\boldsymbol{z}_1, \boldsymbol{z}_2) \geq \frac{1}{2}\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|_2^2 \geq 0$ we have that:

$$
\|\boldsymbol{z}_k^\star - \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}\|_2^2 \leq 2\left(\frac{\lambda_k}{\lambda_k + \eta}\right)^t D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_0^{(k)}),
$$

and

$$
\|\boldsymbol{z}_t^{(k)} - \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}\|_2^2 \leq 2\left(\frac{\lambda_k}{\lambda_k + \eta}\right)^t D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_0^{(k)}),
$$

Thus we can conclude that:

$$
\|\boldsymbol{z}_k^\star - \boldsymbol{z}_t^{(k)}\|_2 \leq \|\boldsymbol{z}_k^\star - \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}\|_2 + \|\boldsymbol{z}_t^{(k)} - \boldsymbol{z}_{t+\frac{1}{2}}^{(k)}\|_2 \leq 2\sqrt{2D^{\epsilon_k}(\boldsymbol{z}_k^\star | \boldsymbol{z}_0^{(k)})}\left(\frac{\lambda_k}{\lambda_k + \eta}\right)^{t/2}. \tag{13}
$$

which concludes the proof. □

B.2 Exploitability

In this section we will prove all the results needed to prove the convergence rate of Algorithm 1 in terms of Nash gap.

We first need the following lemma.

Lemma B.4. *For all $x \geq a > 0$ we have the following inequality:*

$$\left(\frac{x}{a+x}\right)^{x^2/a} \leq \frac{2}{x}.$$

Proof. Consider $(1-y)^r$ with $r > 0$ and $y < 1$ and define $b = \log\left(\frac{1}{1-y}\right)$. Then the following inequality holds:

$$(1-y)^r = e^{-rb} \leq \frac{1}{1+rb} \leq \frac{1}{1+r\log\left(\frac{1}{1-y}\right)}. \tag{14}$$

Define $\phi = a + x$. Thanks to Equation (14) we get:

$$\left(\frac{x}{a+x}\right)^{x^2/a} = \left(1-\frac{a}{\phi}\right)^{\frac{(\phi-a)^2}{a}} \leq \frac{1}{1+\frac{(\phi-a)^2}{a}\log\left(\frac{\phi}{\phi-a}\right)}.$$

Substituting back $x = \phi - a$ we can notice that $\log(1 + a/x) \geq \frac{a/x}{1-a/x} = \frac{a}{x-a} \geq \frac{a}{2x}$, where the last inequality follows from $x \geq a$. Thus we obtain:

$$\left(\frac{x}{a+x}\right)^{x^2/a} \leq \frac{1}{1+\frac{x^2}{a}\log\left(1+\frac{a}{x}\right)} \leq \frac{1}{1+\frac{x^2}{a}\frac{a}{x-a}} = \frac{1}{1+x/2} \leq \frac{2}{x},$$

thus concluding the proof. □

Theorem 5.2. *Given any sequence $\{(\lambda_k, \epsilon_k)\}_{k\in\mathbb{N}}$ defined as in Theorem 4.3 and satisfying $\eta \leq \lambda_k^2 \leq \eta\beta^k/2$ for every $k \in \mathbb{N}$, Algorithm 1 grantees that $\lim_{k\to\infty} \boldsymbol{z}_{T_k}^{(k)}$ is an EFPE.*

Proof. Consider now the following chain of inequalities:

$$\|\boldsymbol{z}_k^\star - \boldsymbol{z}_{T_k}^{(k)}\|_2 \overset{(i)}{=} O\left(\left(\frac{\lambda_k}{\lambda_k+\eta}\right)^{T_k/2}\right) \overset{(ii)}{=} O(1/\lambda_k),$$

where (i) is due to Theorem 5.1 and (ii) is due to Lemma B.4 and the assumption that $T_k := \beta^k \geq \frac{2\lambda_k^2}{\eta}$. Thus $\lim_{k\to\infty} \boldsymbol{z}_{T_k}^{(k)} = \boldsymbol{z}_k^\star$. Moreover from Theorem 4.3 we know that $\boldsymbol{z}_k^\star \to \boldsymbol{z}^\star$ and thus by triangular inequality we can show that:

$$\|\boldsymbol{z}^\star - \boldsymbol{z}_{T_k}^{(k)}\|_2 \leq \|\boldsymbol{z}^\star - \boldsymbol{z}_k^\star\|_2 + \|\boldsymbol{z}_k^\star - \boldsymbol{z}_{T_k}^{(k)}\|_2,$$

and both terms on the right hand side goes to zero with $k \to \infty$, thus concluding the proof. □

Theorem 5.3. *Let $\tilde{\boldsymbol{z}} := (\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \in \mathcal{X} \times \mathcal{Y}$ be such that it holds $\|\boldsymbol{z}_{\lambda,\epsilon}^\star - \tilde{\boldsymbol{z}}\|_2 \leq \nu$ for some $\nu > 0$, then:*

$$\mathit{Gap}(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}) \leq \nu|\Sigma|^{3/2} + 8C/\lambda,$$

where $|\Sigma| := \max\{|\Sigma_1|, |\Sigma_2|\}$ and $C > 0$ is a constant that depends polynomially in $\max_{i\in\{1,2\}} |\mathcal{I}_i|$, $\max_{I\in\mathcal{I}_1\cup\mathcal{I}_2} \alpha_I$, and $\max_{I\in\mathcal{I}_1\cup\mathcal{I}_2} \log n_I$.

Proof. We will first consider the saddle point gap with respect the utility function $f_{\lambda,\epsilon}$ of the regularized-perturbed game:

$$f_{\lambda,\epsilon}(\boldsymbol{x}, \tilde{\boldsymbol{y}}) - f_{\lambda,\epsilon}(\tilde{\boldsymbol{x}}, \boldsymbol{y}) := \boldsymbol{x}^\top \boldsymbol{U}\tilde{\boldsymbol{y}} - \frac{1}{\lambda}d_1^\epsilon(\boldsymbol{x}) + \frac{1}{\lambda}d_2^\epsilon(\tilde{\boldsymbol{y}}) - \tilde{\boldsymbol{x}}^\top \boldsymbol{U}\boldsymbol{y} + \frac{1}{\lambda}d_1^\epsilon(\tilde{\boldsymbol{x}}) - \frac{1}{\lambda}d_2^\epsilon(\boldsymbol{y}) \tag{15}$$

Now we can add and subtract the following quantities $\boldsymbol{x}^\top \boldsymbol{U}\boldsymbol{y}_{\lambda,\epsilon}^\star, \boldsymbol{x}_{\lambda,\epsilon}^{\star,\top}\boldsymbol{U}\boldsymbol{y}$, $\frac{1}{\lambda}d_1^\epsilon(\boldsymbol{x}_{\lambda,\epsilon}^\star)$ and $\frac{1}{\lambda}d_2^\epsilon(\boldsymbol{y}_{\lambda,\epsilon}^\star)$ to obtain:

$$f_{\lambda,\epsilon}(\boldsymbol{x}, \tilde{\boldsymbol{y}}) - f_{\lambda,\epsilon}(\tilde{\boldsymbol{x}}, \boldsymbol{y}) = \underbrace{\boldsymbol{x}^\top \boldsymbol{U}(\tilde{\boldsymbol{y}} - \boldsymbol{y}_{\lambda,\epsilon}^\star) - (\tilde{\boldsymbol{x}} - \boldsymbol{x}_{\lambda,\epsilon}^\star)^\top \boldsymbol{U}\boldsymbol{y}}_{①}$$

$$+\underbrace{\frac{1}{\lambda}d_1^\epsilon(\tilde{\boldsymbol{x}})-\frac{1}{\lambda}d_1^\epsilon(\boldsymbol{x}^\star_{\lambda,\epsilon})+\frac{1}{\lambda}d_2^\epsilon(\tilde{\boldsymbol{y}})-\frac{1}{\lambda}d_2^\epsilon(\boldsymbol{y}^\star_{\lambda,\epsilon})}_{②}$$
$$+\underbrace{f_{\lambda,\epsilon}(\boldsymbol{x},\boldsymbol{y}^\star_{\lambda,\epsilon})-f_{\lambda,\epsilon}(\boldsymbol{x}^\star_{\lambda,\epsilon},\boldsymbol{y})}_{③}.$$

We can observe that by Cauchy-Schwartz inequality we have that the ① is can be upper bounded by:

$$\begin{aligned}\nu L_{\boldsymbol{U}}(\|\boldsymbol{x}\|_2+\|\boldsymbol{y}\|_2)&\le \nu L_{\boldsymbol{U}}\sqrt{2|\Sigma_1|+2|\Sigma_2|}\\&\le 2\nu|\Sigma|^{3/2},\end{aligned}$$

where $L_{\boldsymbol{U}}:=\|\boldsymbol{U}\|_2\le\sqrt{|\Sigma_1|\cdot|\Sigma_2|}\le|\Sigma|$.

On the other hand ② is upper bounded by $4C/\lambda$ (see Lemma A.3 for the definition of the constant C). Finally ③ is negative by definition of the equilibria $\boldsymbol{z}^\star_{\lambda,\epsilon}=(\boldsymbol{x}^\star_{\lambda,\epsilon},\boldsymbol{y}^\star_{\lambda,\epsilon})$. Thus we have:

$$f_{\lambda,\epsilon}(\boldsymbol{x},\tilde{\boldsymbol{y}})-f_{\lambda,\epsilon}(\tilde{\boldsymbol{x}},\boldsymbol{y})\le\nu|\Sigma|^{3/2}+\frac{4C}{\lambda}.$$

On the other hand, following the same argument as in the proof of Theorem 4.1 we know that $|f_{\lambda,\epsilon}(\boldsymbol{x},\boldsymbol{y})-\boldsymbol{x}^\top\boldsymbol{U}\boldsymbol{y}|\le\frac{2C}{\lambda}$ which directly implies that:

$$\boldsymbol{x}^\top\boldsymbol{U}\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{x}}^\top\boldsymbol{U}\boldsymbol{y}\le\nu|\Sigma|^{3/2}+\frac{8C}{\lambda}.$$

This concludes the proof. □

Corollary 5.4. *Given any sequence $\{(\lambda_k,\epsilon_k)\}_{k\in\mathbb{N}}$ as in Theorem 5.2, at the end of every phase $k\in\mathbb{N}$ of Algorithm 1, the following holds:*

$$Gap\left(\boldsymbol{x}^{(k)}_{T_k},\boldsymbol{y}^{(k)}_{T_k}\right)\le\frac{4}{\lambda_k}|\Sigma|^{\frac{3}{2}}\left[2D^{\epsilon_k}(\boldsymbol{z}^\star_k|\boldsymbol{z}^{(k)}_0)\right]^{\frac{1}{2}}+\frac{8C}{\lambda_k}.$$

Proof. From Theorem 5.1 we have that at the end of the k-th phase of Algorithm 1 we can guarantee:

$$\|\boldsymbol{z}^\star_k-\boldsymbol{z}^{(k)}_{T_k}\|_2\le 2\sqrt{2D^{\epsilon_k}\left(\boldsymbol{z}^\star_k|\boldsymbol{z}^{(k)}_0\right)}\left(\frac{\lambda_k}{\lambda_k+\eta}\right)^{T_k/2}.$$

Moreover, by the assumption on λ_k that $T_k:=\beta^k\ge\frac{2\lambda_k^2}{\eta}$, and combining Theorem 5.1 with Lemma B.4 we have that:

$$\left(\frac{\lambda_k}{\lambda_k+\eta}\right)^{T_k/2}\le\frac{2}{\lambda_k}$$

This, thanks to Theorem 5.1, let us conclude that:

$$\|\boldsymbol{z}^\star_k-\boldsymbol{z}^{(k)}_{T_k}\|_2\le\frac{4}{\lambda_k}\left[2D^{\epsilon_k}(\boldsymbol{z}^\star_k|\boldsymbol{z}^{(k)}_0)\right]^{1/2}$$

and thanks to Theorem 5.3 we can readily conclude that:

$$\boldsymbol{x}^\top\boldsymbol{U}\boldsymbol{y}^{(k)}_{T_k}-\boldsymbol{x}^{(k),\top}_{T_k}\boldsymbol{U}\boldsymbol{y}\le\frac{4}{\lambda_k}|\Sigma|^{3/2}\left[2D^{\epsilon_k}(\boldsymbol{z}^\star_k|\boldsymbol{z}^{(k)}_0)\right]^{1/2}+\frac{8C}{\lambda_k}.$$

This concludes the proof. □

B.3 Decomposition

Here we prove the results on the decomposition of Algorithm 1 on the tree. We also report here for completeness of exposition the decomposition of in Farina et al. (2019, 2021) that shows how to compute the conjugate gradient $\nabla d_i^{\epsilon,*}(\cdot)$ and the gradient $\nabla d_i^{\epsilon}(\cdot)$ over the entire tree by computing local conjugate gradient $\nabla d_{\Delta_I}^{\epsilon,*}(\cdot)$ and gradient $\nabla d_{\Delta_I}^{\epsilon}(\cdot)$ at each infoset $I \in \mathcal{I}_i$.

Algorithm 3 Prox-mapping decomposition (Farina et al., 2021)

function $\nabla d_i^{\epsilon,*}(\boldsymbol{g})$ ▷ Conjugate gradient computation.
  $\boldsymbol{v} \leftarrow \mathbf{0} \in \mathbb{R}^{|\Sigma_i|}$
  $\boldsymbol{v}[\varnothing] \leftarrow 1$
  for $I \in \mathcal{I}_i$ in bottom-up order **do** ▷ Compute the behavioral strategy $\boldsymbol{v}$ in bottom-up fashion.
    $\boldsymbol{v}[\sigma_i(I)a] \leftarrow \nabla d_{\Delta_I}^{\epsilon,*}(\boldsymbol{g}[I]/\alpha_I)[a]$
    $\boldsymbol{g}[\sigma_I] \leftarrow \boldsymbol{g}[\sigma_I] - \alpha_I d_{\Delta_I}^{\epsilon}(\boldsymbol{v}[I]) - \boldsymbol{g}[I]^\top \boldsymbol{v}[I]$
  for $I \in \mathcal{I}_i$ in top-down order **do** ▷ Convert the strategy $\boldsymbol{v} \in \mathbb{R}^{|\Sigma_i|}$ from behavioral to sequence form.
    $\boldsymbol{v}[\sigma(I)a] \leftarrow \boldsymbol{v}[\sigma(I)a]\vec{[}\sigma(I)]$
  return $\boldsymbol{v}$

function $\nabla d_i^{\epsilon}(\boldsymbol{g})$ ▷ Gradient computation.
  $\boldsymbol{v} \leftarrow \mathbf{0} \in \mathbb{R}^{|\Sigma_i|}$
  for $I \in \mathcal{I}_i$ in bottom-up order **do** ▷ Compute the gradient $\boldsymbol{v}$ in bottom-up fashion.
    $\boldsymbol{v}[\sigma_i(I)a] \leftarrow \boldsymbol{v}[\sigma_i(I)a] + \alpha_I \nabla d_{\Delta_I}^{\epsilon}\left(\frac{\boldsymbol{g}[\sigma_i(I)a]}{\boldsymbol{g}[\sigma_i(I)]}\right)$
    $\boldsymbol{v}[\sigma_i(I)] \leftarrow \boldsymbol{v}[\sigma_i(I)] + \alpha_I d_{\Delta_I}^{\epsilon}\left(\frac{\boldsymbol{g}[\sigma_i(I)a]}{\boldsymbol{g}[\sigma_i(I)]}\right) - \alpha_I \left(\frac{\boldsymbol{g}[\sigma_i(I)a]}{\boldsymbol{g}[\sigma_i(I)]}\right)^\top \nabla d_{\Delta_I}^{\epsilon}\left(\frac{\boldsymbol{g}[\sigma_i(I)a]}{\boldsymbol{g}[\sigma_i(I)]}\right)$
  return $\boldsymbol{v}$

Theorem 5.5. *The following hold:*

(i) *For any pair of vectors $\boldsymbol{g}, \tilde{\boldsymbol{x}} \in \mathbb{R}^{|\Sigma_1|}$ and $\epsilon > 0$, the updates of Algorithm 2 of the form $\arg\max_{\boldsymbol{x}\in\mathcal{X}} \left\{\boldsymbol{x}^\top \boldsymbol{g} - \frac{d_1^{\epsilon}(\boldsymbol{x})}{\lambda} - \frac{1}{\eta} D_1^{\epsilon}(\boldsymbol{x}|\tilde{\boldsymbol{x}})\right\}$ can be formulated as the computation of the conjugate gradient $\nabla d_1^{\epsilon,*}(\tilde{\boldsymbol{g}})$ for a suitably-defined, efficiently-computable vector $\tilde{\boldsymbol{g}} \in \mathbb{R}^{|\Sigma_1|}$.*

(ii) *For every $\epsilon > 0$, $I \in \mathcal{I}_1$, and $\tilde{\boldsymbol{g}} \in \mathbb{R}^{n_I}$, $\nabla d_{\Delta_I}^{\epsilon,*}(\tilde{\boldsymbol{g}})[a] = (1 - \epsilon\, n_I)\frac{e^{\tilde{\boldsymbol{g}}[a]}}{\|e^{\tilde{\boldsymbol{g}}}\|_1} + \epsilon$, and the local gradient can be computed as $\nabla d_{\Delta_I}^{\epsilon}(\tilde{\boldsymbol{g}})[a] = 1 + \log(\tilde{\boldsymbol{g}}[a] - \epsilon)$ for all $a \in A(I)$.*

Proof. **(i)** Let us firs prove the first statement, and let us consider the following chain of equations:

$$
\begin{aligned}
\hat{\boldsymbol{x}} &:= \arg\max_{\boldsymbol{x}\in\mathcal{X}} \left\{\boldsymbol{x}^\top \boldsymbol{g} - \frac{d_1^{\epsilon}(\boldsymbol{x})}{\lambda} - \frac{1}{\eta} D_1^{\epsilon}(\boldsymbol{x}|\tilde{\boldsymbol{x}})\right\} \\
&= \arg\max_{\boldsymbol{x}\in\mathcal{X}} \left\{\boldsymbol{x}^\top \boldsymbol{g} - \left(\frac{1}{\eta} + \frac{1}{\lambda}\right) d_1^{\epsilon}(\boldsymbol{x}) + \frac{1}{\eta} d_1^{\epsilon}(\tilde{\boldsymbol{x}}) + \frac{1}{\eta} \nabla d_1^{\epsilon}(\tilde{\boldsymbol{x}})^\top (\boldsymbol{x} - \tilde{\boldsymbol{x}})\right\} \\
&= \arg\max_{\boldsymbol{x}\in\mathcal{X}} \left\{\boldsymbol{x}^\top \left(\boldsymbol{g} + \frac{1}{\eta} \nabla d_1^{\epsilon}(\tilde{\boldsymbol{x}})\right) - \left(\frac{1}{\eta} + \frac{1}{\lambda}\right) d_1^{\epsilon}(\boldsymbol{x})\right\} \\
&= \arg\max_{\boldsymbol{x}\in\mathcal{X}} \left\{\boldsymbol{x}^\top \left(\gamma \boldsymbol{g} + \frac{\gamma}{\eta} \nabla d_1^{\epsilon}(\tilde{\boldsymbol{x}})\right) - d_1^{\epsilon}(\boldsymbol{x})\right\} \\
&:= \nabla d_1^{\epsilon,*}(\tilde{\boldsymbol{g}}),
\end{aligned}
$$

where we defined $\tilde{\boldsymbol{g}} := \gamma \boldsymbol{g} + \frac{\gamma}{\eta} \nabla d_1^{\epsilon}(\tilde{\boldsymbol{x}})$ and $\frac{1}{\eta} + \frac{1}{\lambda} = \frac{1}{\gamma}$, which concludes the first statement whenever $\nabla d_1^{\epsilon}(\tilde{\boldsymbol{x}})$ is efficiently computable, which, thanks to Algorithm 3 happens whenever the local gradient $\nabla d_{\Delta_I}^{\epsilon}(\tilde{\boldsymbol{x}})$ have a closed formula (see below the proof for statement **(ii)**).

(ii) Now let us turn to the second statement, which concerns the closed formula updates of the local conjugate gradients of $d^{\epsilon}_{\Delta_I}$. It is well known that $\nabla d^{0,*}_{\Delta_I}(\tilde{\boldsymbol{g}})$ is solved by $\frac{e^{-\tilde{\boldsymbol{g}}[a]}}{\sum_{b\in A(I)} e^{-\tilde{\boldsymbol{g}}[b]}}$. Moreover, it is straightforward to verify that:

$$\begin{aligned}\nabla d^{\epsilon,*}_{\Delta_I}(\tilde{\boldsymbol{g}}) &:= \arg\max_{\boldsymbol{w}\in\Delta_I} \left\{\boldsymbol{w}^\top \tilde{\boldsymbol{g}} - d^{\epsilon}_{\Delta_I}(\boldsymbol{w})\right\} \\ &= \mathbf{1}\epsilon + \arg \max_{\substack{\boldsymbol{w}\geq 0,\\ \sum_{k=1}^{n_I} \boldsymbol{w}[k]=1-\epsilon\cdot n_I}} \left\{\boldsymbol{w}^\top \tilde{\boldsymbol{g}} - d^{0}_{\Delta_I}(\boldsymbol{w})\right\}\end{aligned}$$

where we used the change of variable $\boldsymbol{w} \mapsto \boldsymbol{w} - \mathbf{1}\epsilon$. Clearly a similar statement holds for the updates of the second player. Now we employ the change of variable $\boldsymbol{w} \mapsto \boldsymbol{w}/(1-\epsilon\cdot n_I)$ which gives:

$$\begin{aligned}\nabla d^{\epsilon,*}_{\Delta_I}(\tilde{\boldsymbol{g}}) &= \mathbf{1}\epsilon + (1-\epsilon\cdot n_I)\arg\max_{\boldsymbol{w}\in\Delta_I} \left\{\boldsymbol{w}^\top \tilde{\boldsymbol{g}}\nabla - d^{0}_{\Delta_I}(\boldsymbol{w})\right\} \\ &= \mathbf{1}\epsilon + (1-\epsilon\cdot n_I)\arg\max_{\boldsymbol{w}\in\Delta_I} \nabla d^{0,*}_{\Delta_I}(\tilde{\boldsymbol{g}}) \\ &= \mathbf{1}\epsilon + (1-\epsilon\cdot n_I)\frac{e^{\tilde{\boldsymbol{g}}}}{\sum_{b\in A(I)} e^{\tilde{\boldsymbol{g}}[b]}},\end{aligned}$$

which proves the second statement. Finally the computation of the gradient of $d^{\epsilon}_{\Delta_I}(\boldsymbol{w}) := (\boldsymbol{w}-\mathbf{1}\epsilon)^\top \log(\boldsymbol{w}-\mathbf{1}\epsilon)$ follows from direct computation. □

C Additional Experiments

C.1 Games Description

	$\lvert\mathcal{I}\rvert$	$\lvert\Sigma\rvert$
Kuhn	6	13
Leduc 3	114	337
Leduc 5	390	911
Goofspiel	57	118
dRPS	3	10

Figure 3: Games sizes with respect to the number of infosets $|\mathcal{I}|$ and sequences $|\Sigma|$.

First we are going to describe in details the games used in the evaluation.

Kuhn Poker Is a simplified poker game Kuhn (1950) in which is played with 3 cards. Each player then pays one blind to the pot and is dealt a private card. The first player then decide to either check or to bet (places an additional blind on the pot). In the first case the second player can either check or bet and fold or call in the second case. In the case in which the second player has placed a bet, the first player still has to decide weather to call or to fold. In the case no one has folded, there is a showdown phase in which the player with the hand of higher value wins the pot. In the case one of the player folded, then the other takes the pot.

Leduc Poker (n) Leduc is a card game played first introduced in Southey et al. (2005). It is played with $2n$ cards, where n is called the rank of the game. Each player is dealt a private card an there is a common, unknown card. At the start of every hand, each player places one blind in the pot. There are two betting phases, which are identical and work as in the Kuhn poker, one before the revel of the common card, and one after. After the two betting stages, if no player folded, the player whose hand is of higher value wins the pot.

Goofspiel It is a card game introduce by Ross (1971). Each of the two players has 3 ordered cards, which are used to privately bet on a community card (also of value from 1 to 3) revealed et each of the 3 turns. The player who has bet the highest card at a specific turn, wins the amount represented by the community card. In case of ties the community card is disregarded, otherwise it has the value of the card itself.

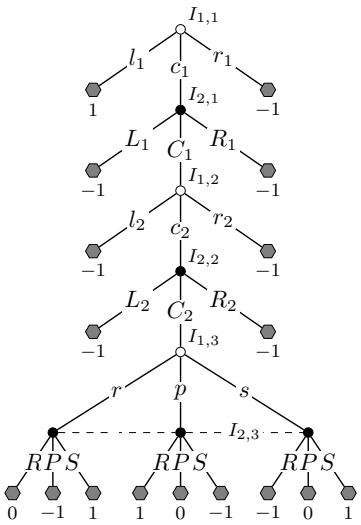

Figure 4: Game tree of the dRPS game described in this section. White (black) nodes are nodes of the infoset of the first (second) player. $I_{i,j}$ is the i-th player's j-th infoset.

dRPS We designed dRPS which is a "deep" version of Rock Paper Scissor in which each player has 3 actions at each infoset for two alternating times. Then, after the second move of the second player, begins a turn of Rock Paper Scissor. The payoffs of such a game are designed so that all the NE of the game prescribe the first player to end the game straight away. This renders irrelevant all the moves played on the other infoset. Indeed every strategy that assign probability 1 to action l_1 of the first player is a NE. The game is designed so that at deep infosets the chance of an algorithm which is *not* designed to explore the tree would have low probability of visiting such infosets due to random chance.

Matrix Game We also consider a 3×3 normal form game (that was also used in the proof of Proposition 4.1). The utility matrix, in which element i, j is the utility received by the first player when it plays i and the second player plays 2, is:

$$\begin{bmatrix} 0.3 & 0.5 & 0.3 \\ 0.7 & 0.3 & 0.7 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}$$

In this game we can easily compute the unique perfect equilibria that is $\boldsymbol{z}^\star = \left([2/3, 1/3, 0]^\top, [0, 2/3, 1/3]^\top\right)$.

C.2 Additional Results

We can see in Figure 5 that in both Leduc 5 and dRPS, Algorithm 1 outperforms the baselines in term if average infoset regret R_I. This is the case for both the baselines which are agnostic to refinements (CFR and EGT) and those who are build to work for refinements, but only find approximate ones (EGT($\epsilon = 0.01$), EGT($\epsilon = 0.001$), CFR($\epsilon = 0.01$) and CFR($\epsilon = 0.001$)). On the other hand, as seen in Section 6, this is not always the case for what concern the Nash gap, even if we can see that in Leduc 5, Algorithm 1 still manages to outperform all the baselines by 2 orders of magnitude. However, this is not always the case. Indeed, in dRPS, CFR outperforms all the other methods by 2 orders of magnitude. As already remarked in Section 6, we think that investigating such behavior requires further investigation.

Finally we used the Matrix Game introduce in the previous section, of which we know the only perfect equilibria, to compute the distance $\|\boldsymbol{z}_t - \boldsymbol{z}^\star\|_2$ of the iterates. In Figure 6 we reported the evolution of the ℓ_2-distance to the unique EFPE of the game. We can see that Algorithm 1 outperforms all the other benchmark in terms of this metric. Indeed it achieves a distance more then one order of magnitude smaller then the second best algorithms that are EGT and CFR. This further corroborates the theoretical findings of Section 4. Finally we can see that CFR, CFR($\epsilon = 0.01$), CFR($\epsilon = 0.001$), EGT, EGT($\epsilon = 0.01$) and EGT($\epsilon = 0.001$) are ordered as expected, with the one instantiated with the lower, fixed, ϵ are closer to $\boldsymbol{z}^\star$.

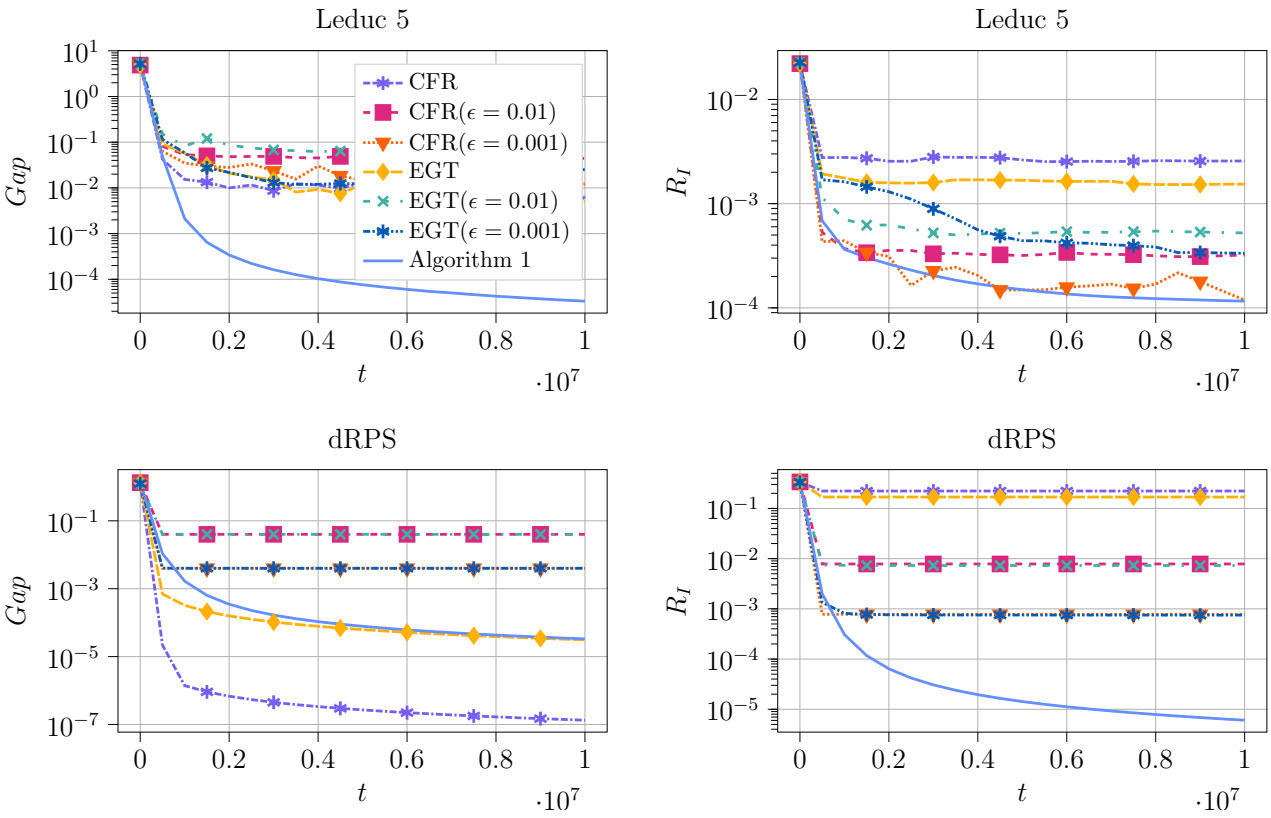

Figure 5: Results of the experimental evaluation. Algorithm 1 is compared with the baselines in terms of Nash gap (*Left*) and average infoset regret (*Right*).

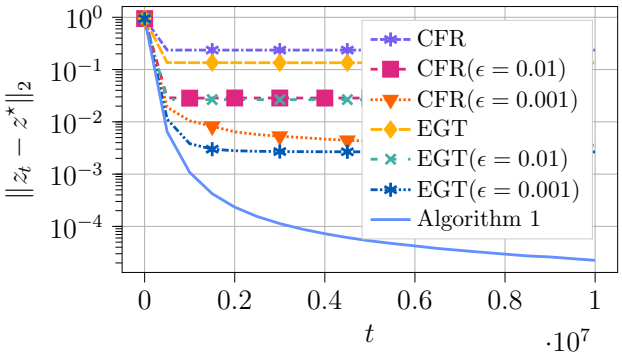

Figure 6: Distance to the unique EFPE of the Matrix Game.