CONVERGENCE OF REGRET MATCHING IN POTENTIAL GAMES AND CONSTRAINED OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Regret matching (RM)—and its modern variants—is a foundational online algorithm that has been at the heart of many AI breakthrough results in solving benchmark zero-sum games, such as poker. Yet, surprisingly little is known so far in theory about its convergence beyond two-player zero-sum games. For example, whether regret matching converges to Nash equilibria in *potential games* has been an open problem for two decades. Even beyond games, one could try to use RM variants for general constrained optimization problems. Recent empirical evidence suggests that they—particularly *regret matching*$^+$ *(RM*$^+$*)*—attain strong performance on benchmark constrained optimization problems, outperforming traditional gradient descent-type algorithms.

We show that alternating RM$^+$ converges to an ϵ-KKT point after $O_\epsilon(1/\epsilon^4)$ iterations, establishing for the first time that it is a sound and fast first-order optimizer. Our argument relates the KKT gap to the accumulated *regret*, two quantities that are entirely disparate in general but interact in an intriguing way in our setting, so much so that when regrets are bounded, our complexity bound improves all the way to $O_\epsilon(1/\epsilon^2)$. From a technical standpoint, while RM$^+$ does *not* have the usual one-step improvement property in general, we show that it does in a certain region that the algorithm will quickly reach and remain in thereafter. In sharp contrast, our second main result establishes a lower bound: RM, with or without alternation, can take an exponential number of iterations to reach a crude approximate solution even in two-player potential games. This represents the first worst-case separation between RM and RM$^+$. Our lower bound shows that convergence to coarse correlated equilibria in potential games is exponentially faster than convergence to Nash equilibria.

1 INTRODUCTION

Regret matching is a foundational online algorithm for minimizing *regret*. It was famously introduced by Hart & Mas-Colell (2000), although its conception can be traced much further back to the seminal *approachability* framework of Blackwell (1956), which lay the groundwork for online learning and regret minimization. As the name suggests, regret matching prescribes playing each action with probability proportional to the (nonnegative) regret accumulated by that action. Its appeal lies in its simplicity and scalability, being both *parameter free* and *scale invariant*.

Regret matching—and modern versions thereof—has been at the forefront of equilibrium computation in massive two-player zero-sum games. A notable variant with strong empirical performance is *regret matching*$^+$, introduced by Tammelin (2014); the only difference is that it truncates all negative coordinates of the regret vector to zero in every iteration. Even so, this variant is typically far superior than its predecessor, and was a central component in AI poker breakthroughs (Bowling et al., 2015; Brown & Sandholm, 2017; 2019b; Moravčík et al., 2017) and a more recent superhuman agent for dark chess (Zhang & Sandholm, 2025).

As such, the regret matching family of algorithms has rightfully been the subject of intense study in contemporary research. Much of this focus has been confined to two-player zero-sum games, where minimizing regret translates to convergence—of the *average* strategies—to minimax (equivalently, Nash) equilibria (Freund & Schapire, 1999). More broadly, in general-sum games, no-regret algo-

rithms guarantee convergence to the set of *coarse correlated equilibria* (Moulin & Vial, 1978)—a more permissive concept than Nash equilibria.

In this paper, we examine the convergence of regret matching and its variants in the seminal class of *potential games*, and, more broadly, nonconvex optimization constrained over a product of simplices. Surprisingly little is known about this question even though it was identified early on as an important open question in this space (Kleinberg et al., 2009; Marden et al., 2007). Recent empirical evidence brings this question to the fore again: Tewolde et al. (2025) showed that the regret matching family—and especially regret matching$^+$—attains strong performance on a benchmark suite of constrained optimization problems, significantly outperforming gradient descent-type algorithms. Yet, there is no theory to suggest that regret matching will even asymptotically converge to approximate KKT points in constrained optimization, which are tantamount to Nash equilibria when dealing specifically with potential games. We fill this gap in this paper.

1.1 Our results

We analyze the convergence of regret matching (RM) and regret matching (RM$^+$) in the general class of (nonconvex) optimization problems constrained over a product of probability simplices. This encompasses as a special case Nash equilibria in potential games when the objective is multilinear; more broadly, to have a unifying treatment of both settings, we think of each probability simplex as being controlled by a single player who is observing the corresponding part of the gradient.

We mostly focus on the alternating version of RM$^+$, whereby players update their strategies one after the other, akin to coordinate descent. Our main result for RM$^+$ is summarized below.

Theorem 1.1. *Alternating RM$^+$ converges to an ϵ-KKT point of any optimization problem over a product of simplices after $O_\epsilon(1/\epsilon^4)$ iterations.*

This theorem confirms that RM$^+$ is a sound and efficient first-order optimizer, lending further credence to the empirical results of Tewolde et al. (2025). We hope that Theorem 1.1 will help cement RM$^+$ in the optimization arsenal going forward.

Our argument proceeds by parameterizing the rate of convergence of RM$^+$ as a function of the accumulated regret, so much so that if the regret with respect to each individual simplex remains bounded, the rate is improved all the way to $T^{-1/2}$.

Theorem 1.2. *Suppose that the regret of RM$^+$ on each individual simplex grows as at most T^α for some $\alpha \in [0, 1/2]$. Then RM$^+$ converges to an ϵ-KKT point after $O_\epsilon(1/\epsilon^{2/1-\alpha})$ iterations.*

RM$^+$ always guarantees regret growing as $\sqrt{T}$, so Theorem 1.1 is implied by Theorem 1.2. What makes the latter theorem surprising is that, in general, regret is a fundamentally disparate property compared to KKT gap: as we point out in Proposition 3.2, a sequence can incur zero regret while having an $\Omega(1)$ KKT gap in each iteration. Even so, Theorem 1.2 directly relates the KKT gap in terms of the regret. In particular, the non-asymptotic rate of Theorem 1.1 is a consequence of the fact that RM$^+$ has the no-regret property! In the special case of potential games, regret is known to bound the rate of convergence to *coarse correlated equilibria*; Theorem 1.2 shows for the first time that regret can also dictate the rate of convergence to Nash equilibria.

On a similar vein, a further important consequence of Theorem 1.2 is that, in *symmetric* potential games, *simultaneous* RM$^+$ converges under a symmetric initialization.

Corollary 1.3. *Simultaneous RM$^+$ converges to an ϵ-Nash equilibrium of any symmetric potential game. Furthermore, if convergence to CCE happens at a rate of $T^{-(1-\alpha)}$, for some $\alpha \in [0, 1/2]$, the rate of convergence to Nash equilibria is no slower than $T^{-\frac{1-\alpha}{2}}$.*

From a technical standpoint, the key challenge is that RM$^+$ does *not* have a one-step improvement property: even if one initializes RM$^+$ close to a KKT point, RM$^+$ can still grossly overshoot. And, of course, it is a parameter-free algorithm, so the usual treatment of gradient descent-type algorithms that relies on appropriately tuning the learning rate falls short. In this context, our starting observation is that, at least when the utility function is linear, RM$^+$ is bound to improve the utility, although the improvement is inversely proportional to the norm of the regret vector (Lemma 3.3). This key property already suffices to show that alternating RM$^+$ will converge to Nash equilibria in potential games. For the more challenging setting where the objective is not multilinear, we first

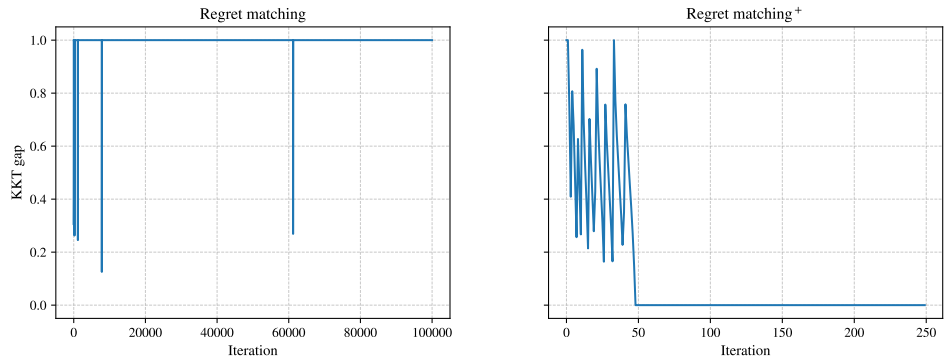


Figure 1: Illustration of our main results: $\texttt{RM}^+$ always converges fast to a KKT point while $\texttt{RM}$ can take exponential time even in two-player identical-interest games, constructed in Section 4.

show that one-step improvement holds conditional on the norm of the regret vector being *sufficiently large* (Lemma 3.7). To conclude the argument, we combine this property with the crucial insight that the ℓ_2 norm of the regret vector is *monotonically increasing* proportionally to the KKT gap (Lemma 3.8). This means that $\texttt{RM}^+$ will never get stuck in a cycle: the regret vector would quickly grow in norm, at which point the one-step improvement promised by Lemma 3.7 kicks in.

Does $\texttt{RM}$ share the same convergence properties as $\texttt{RM}^+$? As a reminder, the only difference is that $\texttt{RM}$ refrains from truncating negative regrets to zero. Even so, we find that this seemingly innocuous difference gives rise to an exponential gap in the performance of $\texttt{RM}$ *vis-à-vis* $\texttt{RM}^+$, manifested even in two-player identical-interest games—a special case of potential games (Figure 1).

Theorem 1.4. *There is a two-player $m \times m$ identical-interest game where RM, with or without alternation, requires $m^{\Omega(m)}$ iterations to converge to an $m^{-\Theta(1)}$-approximate Nash equilibrium.*

This is the first worst-case separation—let alone an exponential one—between $\texttt{RM}$ and $\texttt{RM}^+$. Indeed, in zero-sum games, it is known that $\texttt{RM}$ and $\texttt{RM}^+$ both attain a rate no faster than $T^{-1/2}$ (Farina et al., 2023), even though $\texttt{RM}^+$ typically performs much better in practice. Theorem 1.4 provides further justification for opting for $\texttt{RM}^+$ instead of $\texttt{RM}$, albeit in a fundamentally different setting.

The basic flaw of $\texttt{RM}$ that underpins Theorem 1.4 is that, even with a linear utility, it is not guaranteed to improve the utility even when it has a large best-response gap; specifically, as we show in Lemma 3.6, the improvement is conditional on a good-enough action having nonnegative regret. But herein lies the problem: it could take many iterations before the regret resurfaces to being positive. What happens in the construction behind Theorem 1.4 is that it takes longer and longer—exponentially so—for the regret of the unique good-enough action to be positive; before then, $\texttt{RM}$ is entirely stalled without making any progress. At the same time, $\texttt{RM}$ is bound to converge to the set of coarse correlated equilibria (CCE) at a rate of $T^{-1/2}$, simply because it always has the no-regret property. This leads to the following interesting consequence.

Corollary 1.5. *There is a class of potential games in which RM converges to an ϵ-CCE in $O_\epsilon(1/\epsilon^2)$ rounds but it takes exponentially many rounds to converge to an approximate Nash equilibrium.*

We defer further discussion on related work in Appendix A.

2 Preliminaries

Normal-form games Our first key focus in this paper is on *potential games*, which we represent in the usual normal form. Here, we have n players, each of whom is to select an action a_i from a finite set $\mathcal{A}_i$, with $m_i := |\mathcal{A}_i|$ and $m = \max_{1 \le i \le n} m_i$. Under a joint action profile $(a_1, \dots, a_n) \in \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$, each player $i \in [n]$ receives a payoff given by a *utility function* $u_i : (a_1, \dots, a_n) \mapsto \mathbb{R}$ with range bounded by 1. A player $i \in [n]$ can randomize by specifying a *mixed strategy* $\boldsymbol{x}_i \in \Delta(\mathcal{A}_i) := \{\boldsymbol{x}_i \in \mathbb{R}^{\mathcal{A}_i}_{\ge 0} : \sum_{a_i \in \mathcal{A}_i} \boldsymbol{x}_i[a_i] = 1\}$. Player i strives to maximize its *expected* utility, given by $u_i(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) := \sum_{(a_1, \dots, a_n)} u_i(a_1, \dots, a_n) \prod_{i'=1}^n \boldsymbol{x}_{i'}[a_{i'}]$. A key

fact is that the expected utility is *multilinear*, in that $u_i(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \langle \boldsymbol{x}_i, \boldsymbol{u}_i(\boldsymbol{x}_{-i}) \rangle$ for some utility vector $\boldsymbol{u}_i(\boldsymbol{x}_{-i}) \in \mathbb{R}^{\mathcal{A}_i}$ that does not depend on $\boldsymbol{x}_i$; here and throughout, we use the shorthand notation $\boldsymbol{x}_{-i} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \dots, \boldsymbol{x}_n)$, while $\langle \cdot, \cdot \rangle$ denotes the inner product. Further, $\mathsf{BRGap}_i(\boldsymbol{u}_i) \coloneqq \max_{\boldsymbol{x}'_i \in \Delta(\mathcal{A}_i)} \langle \boldsymbol{x}'_i - \boldsymbol{x}_i, \boldsymbol{u}_i \rangle$.

The predominant solution concept in game theory is the *Nash equilibrium* (Nash, 1950).

Definition 2.1. A strategy profile $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \in \Delta(\mathcal{A}_1) \times \cdots \times \Delta(\mathcal{A}_n)$ is an *ϵ-Nash equilibrium* if for any player $i \in [n]$ and unilateral deviation $\boldsymbol{x}'_i \in \Delta(\mathcal{A}_i)$, $u_i(\boldsymbol{x}'_i, \boldsymbol{x}_{-i}) \le u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i}) + \epsilon$.

A standard relaxation of the Nash equilibrium is the *coarse correlated equilibrium* (Definition B.1), which can be attained by no-regret algorithms (Proposition B.2). While finding a Nash equilibrium is hard even in two-player general-sum games (Daskalakis et al., 2008; Chen et al., 2009), our focus is on *potential games*—equivalently, *congestion games* (Monderer & Shapley, 1996).

Potential games This is a seminal class that goes back to the work of Rosenthal (1973). The defining property is the admission of a global, player-independent function—the *potential*—whose difference reflects the benefit of any unilateral deviation.

Definition 2.2 (Potential game)**.** An n-player game is a *potential game* if there exists a function $\Phi : \Delta(\mathcal{A}_1) \times \cdots \times \Delta(\mathcal{A}_n) \to \mathbb{R}$ such that for any player $i \in [n]$ and strategies $\boldsymbol{x}_i, \boldsymbol{x}'_i \in \Delta(\mathcal{A}_i)$, $\Phi(\boldsymbol{x}'_i, \boldsymbol{x}_{-i}) - \Phi(\boldsymbol{x}_i, \boldsymbol{x}_{-i}) = u_i(\boldsymbol{x}'_i, \boldsymbol{x}_{-i}) - u_i(\boldsymbol{x}_i, \boldsymbol{x}_{-i})$.

A special case of a potential game worth noting is an *identical-interest* game, which means that $u_1(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \cdots = u_n(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ for all $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$; in the presence of only two players, this simplifies to $u_1(\boldsymbol{x}_1, \boldsymbol{x}_2) = \langle \boldsymbol{x}_1 \mathbf{A} \boldsymbol{x}_2 \rangle = u_2(\boldsymbol{x}_1, \boldsymbol{x}_2)$ for a common payoff matrix $\mathbf{A} \in \mathbb{R}^{\mathcal{A}_1 \times \mathcal{A}_2}$.

A (mixed) Nash equilibrium in potential games is amenable to (projected) gradient descent, but is likely hard to compute when the precision $\epsilon > 0$ is exponentially small (Babichenko & Rubinstein, 2021). Our focus will be on algorithms whose complexity is polynomial in $1/\epsilon$.

Constrained optimization More broadly, beyond potential games, we are interested in computing *Karush-Kuhn-Tucker (KKT) points* of a function $u : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} \coloneqq \Delta(\mathcal{A}_1) \times \cdots \times \Delta(\mathcal{A}_n)$. We assume that u, which is to be maximized, is differentiable over an open set $\hat{\mathcal{X}} \supset \mathcal{X}$ and *L-smooth*, meaning that $\|\nabla u(\boldsymbol{x}) - \nabla u(\boldsymbol{x}')\|_2 \le L \|\boldsymbol{x} - \boldsymbol{x}'\|_2$ for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$; $\|\boldsymbol{x}\|_2 \coloneqq \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$ denotes the (Euclidean) ℓ_2 norm. We make the normalization assumption $|\langle \boldsymbol{x}_i - \boldsymbol{x}'_i, \nabla_{\boldsymbol{x}_i} u(\boldsymbol{x}) \rangle| \le 1$ for all $i \in [n]$ and $\boldsymbol{x}_i, \boldsymbol{x}'_i \in \Delta(\mathcal{A}_i)$. The goal is to minimize *KKT gap*, which we measure by

$$\mathsf{KKTGap} : \mathcal{X} \ni \boldsymbol{x} \mapsto \max_{\boldsymbol{x}' \in \mathcal{X}} \langle \boldsymbol{x}' - \boldsymbol{x}, \nabla u(\boldsymbol{x}) \rangle = \sum_{i=1}^{n} \mathsf{BRGap}_i(\nabla_{\boldsymbol{x}_i} u(\boldsymbol{x})). \tag{1}$$

A point with small KKT gap per (1) is also referred to as an approximate *first-order stationary point*, which is an approximate fixed point of the (constrained) gradient descent mapping $\boldsymbol{x} \mapsto \Pi_{\mathcal{X}}(\boldsymbol{x} + \eta \nabla u(\boldsymbol{x}))$, where $\eta \le 1/L$ and $\Pi_{\mathcal{X}}(\cdot)$ is (Euclidean) projection mapping. A potential game can be seen as the special case in which u is multilinear.[1] One class of problems that fits in this framework are *imperfect-recall* games; we point to Tewolde et al. (2025) and the references therein.

Online learning and regret matching Moving on, we now introduce RM and RM$^+$ within the framework of online learning. Here, a *learner* interacts with an *environment* over a sequence of T rounds. In each round $t \in [T]$, the learner first elects a mixed strategy $\boldsymbol{x} \in \Delta(\mathcal{A})$. The environment in turn specifies a linear utility function $u^{(t)} : \boldsymbol{x} \mapsto \langle \boldsymbol{x}, \boldsymbol{u}^{(t)} \rangle$ for some utility vector $\boldsymbol{u}^{(t)} \in \mathbb{R}^{\mathcal{A}}$; $u^{(t)}$ has a range bounded by 1. In the full-feedback setting, $\boldsymbol{u}^{(t)}$ is revealed to the learner at the end of the round. The performance of the learner in this online environment is evaluated through *regret*,

$$\mathsf{Reg}^{(T)} \coloneqq \max_{\boldsymbol{x}' \in \Delta(\mathcal{A})} \sum_{t=1}^{T} \langle \boldsymbol{x}' - \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle. \tag{2}$$

Two algorithms for minimizing regret on the simplex are regret matching (RM) and regret matching$^+$ (RM$^+$), formally defined in Algorithms 1 and 2. They both prescribe playing an action with probability proportional to the nonnegative regret accumulated by that action. Their *only* difference is

[1] We caution that if we use the KKT gap per (1) in the special case of potential games we get the *sum* of the players' deviation benefits, while the approximation in the Nash equilibrium is defined with respect to the *max*.

that RM$^+$ always truncates the regret to 0 (Line 11); in that line, $\mathbf{1}$ denotes the all-ones vector, whose dimension is omitted as it is clear from the context, and $[\cdot]^+ := \max(\mathbf{0}, \cdot)$ is the nonnegative part.

Proposition 2.3 (Zinkevich et al., 2007; Farina et al., 2021)**.** *For any sequence of utilities $(\boldsymbol{u}^{(t)})_{t=1}^T$, both RM and RM$^+$ guarantee that the ℓ_2 norm of $[\boldsymbol{r}^{(T)}]^+$ is at most $\sqrt{mT}$.*

In particular, for both RM and RM$^+$, $\mathsf{Reg}^{(T)} \le \|[\boldsymbol{r}^{(T)}]^+\|_\infty \le \|[\boldsymbol{r}^{(T)}]^+\|_2 \le \sqrt{mT}$.

Algorithm 1: Regret matching (RM)

1 Initialize cumulative regrets $\boldsymbol{r}^{(0)} \leftarrow \mathbf{0}$;
2 Initialize strategy $\boldsymbol{x}^{(0)} \in \Delta(\mathcal{A})$;
3 **for** $t = 1, \dots, T$ **do**
4 Set $\boldsymbol{\theta}^{(t)} \leftarrow [\boldsymbol{r}^{(t-1)}]^+$;
5 **if** $\boldsymbol{\theta}^{(t)} \neq \mathbf{0}$ **then**
6 Compute $\boldsymbol{x}^{(t)} \leftarrow {\boldsymbol{\theta}^{(t)}}/{\|\boldsymbol{\theta}^{(t)}\|_1}$;
7 **else**
8 $\boldsymbol{x}^{(t)} \leftarrow \boldsymbol{x}^{(t-1)}$;
9 Output strategy $\boldsymbol{x}^{(t)} \in \Delta(\mathcal{A})$;
10 Observe utility $\boldsymbol{u}^{(t)} \in \mathbb{R}^{\mathcal{A}}$;
11 $\boldsymbol{r}^{(t)} \leftarrow \boldsymbol{r}^{(t-1)} + \boldsymbol{u}^{(t)} - \langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle \mathbf{1}$;

Algorithm 2: Regret matching$^+$ (RM$^+$)

1 Initialize cumulative regrets $\boldsymbol{r}^{(0)} := \mathbf{0}$;
2 Initialize strategy $\boldsymbol{x}^{(1)} \in \Delta(\mathcal{A})$;
3 **for** $t = 1, \dots, T$ **do**
4 Set $\boldsymbol{\theta}^{(t)} \leftarrow \boldsymbol{r}^{(t-1)}$;
5 **if** $\boldsymbol{\theta}^{(t)} \neq \mathbf{0}$ **then**
6 Compute $\boldsymbol{x}^{(t)} \leftarrow {\boldsymbol{\theta}^{(t)}}/{\|\boldsymbol{\theta}^{(t)}\|_1}$;
7 **else**
8 $\boldsymbol{x}^{(t)} \leftarrow \boldsymbol{x}^{(t-1)}$;
9 Output strategy $\boldsymbol{x}^{(t)} \in \Delta(\mathcal{A})$;
10 Observe utility $\boldsymbol{u}^{(t)} \in \mathbb{R}^{\mathcal{A}}$;
11 $\boldsymbol{r}^{(t)} \leftarrow [\boldsymbol{r}^{(t-1)} + \boldsymbol{u}^{(t)} - \langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle \mathbf{1}]^+$;

Simultaneous and alternating updates We are interested in the convergence of RM and RM$^+$ when used by all players; in the constrained optimization setting, we think of having one player acting on each simplex, in direct correspondence with potential games. In this setting, the sequence of utilities $(\boldsymbol{u}_i^{(t)})_{t=1}^T$ given as input to player $i \in [n]$ is determined by the strategies of the other players. If the updates are *simultaneous*, we have $\boldsymbol{u}_i^{(t)} = \nabla_{\boldsymbol{x}_i} u(\boldsymbol{x}^{(t)})$ for each player $i \in [n]$. In the alternating setting, we first fix a precision $\epsilon > 0$. We go through the players in a round-robin fashion $i = 1, \dots, n$. For each $i \in [n]$, we first compute $\boldsymbol{u}_i^{(t)} = \nabla_{\boldsymbol{x}_i} u(\boldsymbol{x}_{i'<i}^{(t+1)}, \boldsymbol{x}_{i' \ge i}^{(t)})$. If the best-response gap is already at most ϵ, we refrain from updating that player. That is, $\boldsymbol{x}_i^{(t+1)} := \boldsymbol{x}_i^{(t)}$; this is a lazy version of the update. Otherwise, the player updates its strategy to $\boldsymbol{x}_i^{(t+1)}$ using $\boldsymbol{u}_i^{(t)}$. Alternation speeds up performance, at least in zero-sum games (Tammelin, 2014), and has been the subject of much recent research (Wibisono et al., 2022; Cevher et al., 2023).

3 Convergence of Regret Matching$^+$

In this section, we analyze the convergence of RM$^+$ in potential games (Section 3.1), and more broadly, constrained optimization (Section 3.2). A central theme in our analysis of RM and RM$^+$ is a recurring connection between regret and convergence to KKT points.

Before we proceed, it is worth highlighting that, in general, the no-regret property is fundamentally different from convergence to KKT points in *nonconvex problems*. To begin with, we point out that when the underlying function to be maximized, u, is concave, then the no-regret property does imply convergence to a global optimum, from Jensen's inequality.

Proposition 3.1 (Under concavity, no-regret implies convergence)**.** *Let u be a smooth concave function. If an online algorithm observes the sequence of utilities $(\nabla u(\boldsymbol{x}^{(t)}))_{t=1}^T$, then $\frac{1}{T}\sum_{t=1}^T u(\boldsymbol{x}^{(t)}) \ge \max_{\boldsymbol{x} \in \mathcal{X}} u(\boldsymbol{x}) - \frac{1}{T}\mathsf{Reg}^{(T)}$, where $\mathsf{Reg}^{(T)}$ is the regret of the algorithm.*

Thus, if the algorithm has vanishing average regret, $u(\boldsymbol{x}^{(t)}) \to \max_{\boldsymbol{x} \in \mathcal{X}} u(\boldsymbol{x})$. But beyond concave problems, no-regret algorithms do not necessarily guarantee convergence even to a KKT point.

Proposition 3.2. *For any $T \in \mathbb{N}$ with $T = 0 \mod 4$, there exists a polynomial function u in $[0, 1]$ and a sequence of points $(x^{(t)})_{t=1}^T$ such that*

- *the regret of the sequence with respect to $(\nabla u(x^{(t)}))_{t=1}^T$ is zero, while*

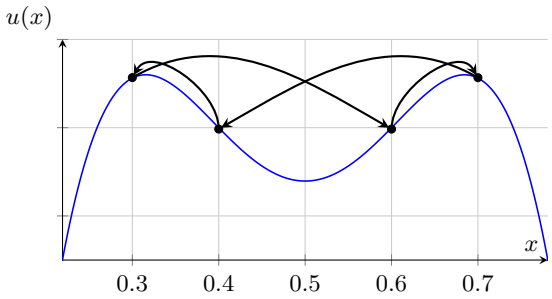


Figure 2: The example corresponding to Proposition 3.2, demonstrating that having zero regret, let alone sublinear, has no implications concerning convergence in terms of KKT gap.

- *every point in the sequence has an $\Omega(1)$ KKT gap with respect to the function u.*

This is based on the 4-cycle $0.6 \to 0.7 \to 0.4 \to 0.3 \to 0.6$. If the gradients observed at those points are $0.6 \mapsto 2, 0.7 \mapsto -1, 0.4 \mapsto -2, 0.3 \mapsto 1$, it follows that i) $\sum_{t=1}^{T} \nabla u(x^{(t)}) = 0$ and ii) $\sum_{t=1}^{T} x^{(t)} \nabla u(x^{(t)}) = 0$, which in turn implies that this sequence incurs zero regret. But, by construction, the gradients at those interior points have a large magnitude, which in turn implies that the KKT gap is large. (That the average is a local minimum is coincidental.) A polynomial consistent with the above gradients is $90x - 298.\bar{3}x^2 + 416.\bar{6}x^3 - 208.\bar{3}x^4$, leading to Proposition 3.2; we note that the above sequence of iterates is not realizable through an algorithm such as gradient descent.

3.1 Potential Games

We first analyze convergence in potential games. A key property, which paves the way for Theorem 3.4, is that, for a fixed utility vector, $\texttt{RM}^+$ has a one-step improvement property; the lemma below takes the perspective of a single, arbitrary player in the game.

Lemma 3.3 (One-step improvement for $\texttt{RM}^+$)**.** *For any $\boldsymbol{r} \in \mathbb{R}^{\mathcal{A}}_{\geq 0}$ and $\boldsymbol{u} \in \mathbb{R}^{\mathcal{A}}$, we define $\boldsymbol{x} := \boldsymbol{r}/\|\boldsymbol{r}\|_1$; if $\boldsymbol{r} = \boldsymbol{0}$, $\boldsymbol{x} \in \Delta(\mathcal{A})$ can be arbitrary. If $\boldsymbol{r}' := [\boldsymbol{r} + \boldsymbol{u} - \langle \boldsymbol{x}, \boldsymbol{u} \rangle \boldsymbol{1}]^+ \neq \boldsymbol{0}$ and $\boldsymbol{x}' := \boldsymbol{r}'/\|\boldsymbol{r}'\|_1$,*

$$\langle \boldsymbol{x}' - \boldsymbol{x}, \boldsymbol{u} \rangle \geq \frac{1}{\|\boldsymbol{r}'\|_1} \left(\max_{a \in \mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u} \rangle \right)^2 = \frac{1}{\|\boldsymbol{r}'\|_1} \mathsf{BRGap}(\boldsymbol{u})^2. \tag{3}$$

If $\boldsymbol{r}' = \boldsymbol{0}$, then $\langle \boldsymbol{x}, \boldsymbol{u} \rangle = \langle \boldsymbol{x}', \boldsymbol{u} \rangle \geq \max_{a \in \mathcal{A}} \boldsymbol{u}[a]$.

The left-hand side of (3) reflects the improvement in utility obtained by updating $\boldsymbol{x}$ to $\boldsymbol{x}'$. $\max_{a \in \mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u} \rangle$ is the best-response gap of $\boldsymbol{x}$ with respect to $\boldsymbol{u}$. Lemma 3.3 implies that the utility is monotonically increasing—unless the current strategy is already a best response to $\boldsymbol{u}$. Furthermore, so long as the regret vector is *small enough*, the improvement is bound to be substantial, being proportional to the squared best-response gap. It is worth noting that Lemma 3.3 holds no matter the initial regret vector $\boldsymbol{r}$, subject to $\boldsymbol{r} \in \mathbb{R}^{\mathcal{A}}_{\geq 0}$; this invariance always holds for $\texttt{RM}^+$ (by definition of the algorithm in Line 11), but that is not so for $\texttt{RM}$ (*cf.* Lemma 3.6).

The proof of Lemma 3.3 proceeds by expressing (3) in terms of the regret vectors, and appears in Appendix C.1. Furthermore, as we point out in Lemma C.4, Lemma 3.3 is in a certain sense tight.

Convergence in potential games We now employ Lemma 3.3 to show that alternating $\texttt{RM}^+$ quickly converges to approximate Nash equilibria in potential games. Using the fact that the game admits a potential function (per Definition 2.2), we have that for any round $t \in [T]$, $\Phi(\boldsymbol{x}_1^{(t+1)}, \dots, \boldsymbol{x}_n^{(t+1)}) - \Phi(\boldsymbol{x}_1^{(t)}, \dots, \boldsymbol{x}_n^{(t)}) \geq \sum_{i=1}^{n} \frac{1}{\|\boldsymbol{r}_i^{(t)}\|_1} \mathsf{BRGap}_i(\boldsymbol{u}_i^{(t)})^2 \mathbb{1}\{\mathsf{BRGap}_i(\boldsymbol{u}_i^{(t)}) > \epsilon\}$, where we used Lemma 3.3 together with the assumption that only players with more than ϵ best-response gap update their strategies. The telescopic summation over $t = 1, \dots, T$ yields

$$\Phi_{\mathsf{range}} \geq \sum_{t=1}^{T} \sum_{i=1}^{n} \frac{1}{\|\boldsymbol{r}_i^{(t)}\|_1} \mathsf{BRGap}_i(\boldsymbol{u}_i^{(t)})^2 \mathbb{1}\{\mathsf{BRGap}_i(\boldsymbol{u}_i^{(t)}) > \epsilon\}, \tag{4}$$

where Φ_{range} denotes the range of the potential function. If in every round $t \in [T]$ there is a player $i \in [n]$ such that $\mathsf{BRGap}_i(\boldsymbol{u}_i^{(t)}) > \epsilon$, we have $\Phi_{\mathsf{range}} \geq \sum_{t=1}^{T} \frac{1}{m\sqrt{t}}\epsilon^2 \geq \frac{1}{m}\epsilon^2\sqrt{T}$, where we used that $\|\boldsymbol{r}_i^{(t)}\|_1 \leq \sqrt{m}\|\boldsymbol{r}_i^{(t)}\|_2 \leq m\sqrt{T}$ (Proposition 2.3). We thus arrive at the following result.

Theorem 3.4. *In any potential game, alternating* RM$^+$ *requires at most* $1 + {}^{(m\Phi_{\mathsf{range}})^2}/_{\epsilon^4}$ *rounds to converge to an* ϵ*-Nash equilibrium. More broadly, if* $\|\boldsymbol{r}_i^{(t)}\|_1 \leq C(n,m)t^\alpha$ *for all* $i \in [n]$ *and some* $\alpha \in [0, 1/2]$*, it requires* $1 + {}^{(C(n,m)\Phi_{\mathsf{range}})^\beta}/_{\epsilon^{2\beta}}$ *rounds, where* $\beta := {}^1/_{1-\alpha}$.

This provides a convergence rate of $T^{-1/4}$ to Nash equilibria. Notwithstanding Proposition 3.2, an intriguing aspect of Theorem 3.4 is that it *connects convergence to Nash equilibria to the regret*. In particular, if RM$^+$ did not have the no-regret property, meaning that $\|\boldsymbol{r}_i^{(t)}\|_1 = \Omega(t)$, we could only prove an exponential bound since $\sum_{t=1}^{T} {}^1/_t = \Theta(\log T)$. At the other end of the spectrum, when each player accumulates constant regret, Theorem 3.4 implies an improved convergence rate of $T^{-1/2}$. It is an open question whether RM$^+$ can experience $\Omega(\sqrt{T})$ regret in potential games.

Faster rate using discounting Next, we refine Theorem 3.4 through the use of *discounted* RM$^+$, which means that the regret vector is multiplied by a discount factor $\alpha^{(t)} \in (0, 1]$ in each round; we spell out DRM$^+$ in Algorithm 3. This class of algorithms was introduced by Brown & Sandholm (2019a), who showed that discounting drastically improves empirical performance in zero-sum games. Our next result shows that DRM$^+$ with geometric discounting, $\alpha^{(t)} = 1 - \gamma$ for some $\gamma > 0$, attains a rate of $T^{-1/2}$ to Nash equilibria in potential games; the basic reason is that DRM$^+$ maintains the norm of the regret vector bounded by $\sqrt{{}^m/_\gamma}$ (Lemma C.2 and Corollary C.3), while still enjoying the one-step improvement property of Lemma 3.3.

Corollary 3.5. *In any potential game, alternating* DRM$^+$ *with discount factor* $1 - \gamma < 1$ *requires at most* $1 + {}^{m\Phi_{\mathsf{range}}}/_{\epsilon^2\sqrt{\gamma}}$ *rounds to converge to an* ϵ*-Nash equilibrium.*

Regret matching Before we switch gears to the more general constrained optimization setting, it is instructive to examine the behavior of RM. It turns out that one can adjust Lemma 3.3, but with a crucial caveat: the one-step improvement property is now only *conditional*, as specified below.

Lemma 3.6. *For any* $\boldsymbol{r} \in \mathbb{R}^{\mathcal{A}}_{\geq 0}$ *and* $\boldsymbol{u} \in \mathbb{R}^{\mathcal{A}}$*, we define* $\boldsymbol{x} := {}^{\boldsymbol{\theta}}/_{\|\boldsymbol{\theta}\|_1}$*, where* $\boldsymbol{\theta} := \max(\boldsymbol{r}, \boldsymbol{0})$*; if* $\boldsymbol{\theta} = \boldsymbol{0}$*,* $\boldsymbol{x} \in \Delta(\mathcal{A})$ *can be arbitrary. If* $\boldsymbol{r}' := \boldsymbol{r} + \boldsymbol{u} - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \boldsymbol{1}$ *and* $\boldsymbol{x}' := {}^{\boldsymbol{\theta}'}/_{\|\boldsymbol{\theta}'\|_1}$*, where* $\boldsymbol{\theta}' = \max(\boldsymbol{r}', \boldsymbol{0}) \neq \boldsymbol{0}$*, we have* $\langle \boldsymbol{x}' - \boldsymbol{x}, \boldsymbol{u}\rangle \geq \frac{1}{\|\boldsymbol{\theta}'\|_1}\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|_2^2 \geq \frac{1}{\|\boldsymbol{\theta}'\|_1}\left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2 \mathbb{1}\{\boldsymbol{r}[a] \geq 0\}$*, where* $a \in \arg\max_{a'\in\mathcal{A}} \boldsymbol{u}[a']$*. If* $\boldsymbol{\theta}' = \boldsymbol{0}$*, then* $\langle \boldsymbol{x}, \boldsymbol{u}\rangle = \langle \boldsymbol{x}', \boldsymbol{u}\rangle \geq \max_{a\in\mathcal{A}} \boldsymbol{u}[a]$.

We see that RM's one-step improvement is conditional on the regret accumulated thus far by a best-response action to be nonnegative. This is not an artifact of our analysis; it alludes to a fundamental discrepancy between RM and RM$^+$ that will be formally established later on (Theorem 4.4). The main issue with RM can be seen as follows. If we consider a utility vector $\boldsymbol{u} = (1, 0)$ and the initial regret vector is, say, $(-R, R)$, it will take RM many iterations—proportionally to the magnitude of $R > 0$—to finally change strategies, although this will eventually happen with a stationary utility.

3.2 Nonlinear optimization and simultaneous updates

We now treat the more general setting where we are maximizing an L-smooth function u.

Single simplex We begin with the special case of a single probability simplex, $\mathcal{X} = \Delta(\mathcal{A})$. Our first goal is to adapt Lemma 3.3. The key challenge is that RM$^+$ does *not* have a one-step improvement, unlike algorithms such as gradient descent (for a small enough learning rate), even if one initializes RM$^+$ close to a KKT point. But we observe that if the norm of the regret vector is *large enough*—having small regrets is an obstacle here, in contrast to Section 3.1—we are guaranteed a one-step improvement in terms of the value of the function (Lemma 3.7).

To do so, we will use the basic quadratic bound, which yields $u(\boldsymbol{x}') \geq u(\boldsymbol{x}) + \langle \nabla u(\boldsymbol{x}), \boldsymbol{x}' - \boldsymbol{x}\rangle - \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2$; we think of $\boldsymbol{x}'$ as the updated strategy starting from $\boldsymbol{x}$. Using a slight refinement of Lemma 3.3, we first have the lower bound $\langle \boldsymbol{x}' - \boldsymbol{x}, \nabla u(\boldsymbol{x})\rangle \geq \frac{1}{\|\boldsymbol{r}'\|_1}\|\boldsymbol{r} - \boldsymbol{r}'\|_2^2$ (Lemma C.5).

Also, we observe that $\|\boldsymbol{x} - \boldsymbol{x}'\|_1 \le \|\boldsymbol{r} - \boldsymbol{r}'\|_1 \left(\frac{1}{\|\boldsymbol{r}\|_1} + \frac{1}{\|\boldsymbol{r}'\|_1}\right)$ (Lemma C.6). We are now ready to establish a *conditional* one-step improvement when the regret vector has a *sufficiently large* norm.

Lemma 3.7. *Let u be an L-smooth function over $\Delta(\mathcal{A})$. For any $\boldsymbol{r} \in \mathbb{R}^{\mathcal{A}}_{\ge 0}$ with $\boldsymbol{r} \neq \boldsymbol{0}$, we define $\boldsymbol{x} := \boldsymbol{r}/\|\boldsymbol{r}\|_1$. Further, let $\boldsymbol{r}' := [\boldsymbol{r} + \nabla u(\boldsymbol{x}) - \langle \boldsymbol{x}, \nabla u(\boldsymbol{x})\rangle \boldsymbol{1}]^+ \neq \boldsymbol{0}$ and $\boldsymbol{x}' := \boldsymbol{r}'/\|\boldsymbol{r}'\|_1$. If $\|\boldsymbol{r}'\|_2 \ge \max\{2m, 9mL\}$, then $u(\boldsymbol{x}') - u(\boldsymbol{x}) \ge \frac{1}{2\|\boldsymbol{r}'\|_1} \left(\max_{\boldsymbol{x}^\star \in \Delta(\mathcal{A})} \langle \boldsymbol{x}^\star - \boldsymbol{x}, \nabla u(\boldsymbol{x})\rangle\right)^2$.*

Lemma 3.7 only shows a one-step improvement so long as the norm of the regret vector is large enough. But how can we guarantee that? It would seem possible that RM^+ ends up cycling in perpetuity under a regret vector with small norm. The following lemma shows that cannot happen.

Lemma 3.8. *For any t, RM^+ guarantees $\|\boldsymbol{r}^{(t)}\|_2^2 \ge \|\boldsymbol{r}^{(t-1)}\|_2^2 + \|[\boldsymbol{g}^{(t)}]^+\|_2^2$, where $\boldsymbol{g}^{(t)} := \nabla u(\boldsymbol{x}) - \langle \nabla u(\boldsymbol{x}), \boldsymbol{x}^{(t)}\rangle \boldsymbol{1}$ is the instantaneous regret at round t.*

In particular, $\|\boldsymbol{r}^{(t)}\|_2^2 \ge \|\boldsymbol{r}^{(t-1)}\|_2^2 + \|[\boldsymbol{g}^{(t)}]_+\|_2^2 \ge \|\boldsymbol{r}^{(t-1)}\|_2^2 + \mathsf{KKTGap}(\boldsymbol{x}^{(t)})^2$ since $\|[\boldsymbol{g}^{(t)}]^+\|_2^2 \ge \mathsf{KKTGap}(\boldsymbol{x}^{(t)})^2$. Not only is the ℓ_2 norm of the regret vector nondecreasing, but the increase is at least $\mathsf{KKTGap}(\boldsymbol{x}^{(t)})^2$ at each round $t \in [T]$. Combining with Lemma 3.7 yields the following.

Theorem 3.9. *Let u be an L-smooth function in $\Delta(\mathcal{A}) \subset \mathbb{R}^m$ with range u_{range} and $R := \max\{2m, 9mL\}$. RM^+ requires at most $1 + {}^{(m(2u_{\mathsf{range}}+R^2))^2}/_{\epsilon^4}$ rounds to reach an ϵ-KKT point.*

Simultaneous updates in symmetric potential games We now use Theorem 3.9 to prove convergence of *simultaneous* RM^+ in *symmetric* potential games; our earlier result in Theorem 3.4 shows convergence for arbitrary potential games but for the alternating version. The symmetry assumption here means that $\mathcal{A}_1 = \mathcal{A}_1 = \cdots = \mathcal{A}_n$ and $\boldsymbol{u}_1(\boldsymbol{x}_{-1}) = \boldsymbol{u}_2(\boldsymbol{x}_{-2}) = \cdots = \boldsymbol{u}_n(\boldsymbol{x}_{-n})$ when $\boldsymbol{x}_1 = \boldsymbol{x}_2 = \cdots = \boldsymbol{x}_n$. It is further assumed that all players initialize from the same strategy, so that the previous property implies that, inductively, it will be the case that $\boldsymbol{x}_1^{(t)} = \boldsymbol{x}_2^{(t)} = \cdots = \boldsymbol{x}_n^{(t)}$ for all t under simultaneous updates because players observe exactly the same utilities. A simple example of this is a two-player game with a common, symmetric payoff matrix $\mathbf{A} = \mathbf{A}^\top$. Then $\boldsymbol{u}_1(\boldsymbol{x}_2) = \mathbf{A}\boldsymbol{x}_2$ and $\boldsymbol{u}_2(\boldsymbol{x}_1) = \mathbf{A}\boldsymbol{x}_1$, so the previous assumption is satisfied.

Corollary 3.10. *In any symmetric potential game, simultaneous RM^+ converges to an ϵ-Nash equilibrium after $O_\epsilon(1/\epsilon^4)$ rounds. In particular, if convergence to the set of CCE happens at a rate of $T^{-\alpha}$, for some $\alpha \in [0, 1/2]$, the rate of convergence to Nash equilibria is at least $T^{-\frac{1-\alpha}{2}}$.*

Multiple simplices We now have the necessary tools to analyze the general case where we maximize u over a product of simplices. Similarly to Theorem 3.4, we run alternating RM^+, thinking of every individual simplex as being controlled by a single player; this is akin to coordinate descent.

Theorem 3.11. *Let u be an L-smooth function in $\Delta(\mathcal{A}_1) \times \cdots \times \Delta(\mathcal{A}_n)$ with range u_{range} and $R := \sqrt{\sum_{i=1}^n \max\{2m_i, 9m_iL\}^2}$. Alternating RM^+ requires at most $1 + {}^{(mn^2(2u_{\mathsf{range}}+R^2))^2}/_{\epsilon^4}$ rounds to reach an ϵ-KKT point of u.*

4 Exponential lower bounds for regret matching

In stark contrast, we show that RM, with or without alternation, can take exponentially many rounds to reach an approximate Nash equilibrium even in two-player identical-interest games. The underlying class of games is based on the one considered by Panageas et al. (2023), who treated fictitious play. Specifically, for $m = 4, 6, \ldots$ and $k \in \mathbb{N}$ we define the matrix $\mathbf{A}_{m,k}$ per the recursion

$$\mathbb{R}^{m\times m} \ni \mathbf{A}_{m,k} := \begin{bmatrix} k+1 & 0 & \cdots & 0 & 0 \\ 0 & & & & k+4 \\ \vdots & & \mathbf{A}_{m-2,k+4} & & \vdots \\ 0 & & & & 0 \\ k+2 & 0 & \cdots & 0 & k+3 \end{bmatrix}, \text{ where } \mathbf{A}_{2,k} := \begin{bmatrix} k+1 & 0 \\ k+2 & k+3 \end{bmatrix}.$$

(An illustrative example appears in Appendix C.2.) For any even dimension m, we define $\mathbf{A} := \mathbf{A}_{m,0}$, with maximum entry $2m-1$. Further, we define, for $1 \le a_1 \le m+1$ and $1 \le a_2 \le m+1$,

$$\mathbf{B}[a_1, a_2] := \begin{cases} \mathbf{A}[a_1, a_2] & \text{if } a_1 \le m \text{ and } a_2 \le m; \\ 1/2 & \text{if } (a_1 = m+1 \text{ and } a_2 = 1) \text{ or } (a_1 = 1 \text{ and } a_2 = m+1); \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

The action sets of the two players are $\mathcal{A}_1 = [m+1] = \mathcal{A}_2$. We assume that RM is initialized to the pure strategy $(m+1, m+1)$. We recall that one round includes one update from each player, which for now is assumed to be made in a simultaneous fashion. For a payoff $k \in \mathbb{N}$, we denote by $a_1(k), a_2(k) \in [m]$ the row and column index, respectively, corresponding to k in the matrix $\mathbf{A}$.

We begin by stating a basic invariance concerning the behavior of RM when executed on the game (5).

Property 4.1. *After the first round both players play the first action. Thereupon, either the players play with probability* 1 $(a_1(k), a_2(k))$*, or, when k is odd, only Player 1 (respectively, Player 2 when k is even) mixes between $a_1(k)$ and $a_1(k+1)$ (respectively, $a_2(k)$ and $a_2(k+1)$). If a row or a column stops being played, it will never be played henceforth. An action profile $(a_1(k+1), a_2(k+1))$ is played with positive probability only if $(a_1(k), a_2(k))$ was played at some previous round.*

We prove this property inductively in Appendix C.2. We take it for granted in what follows.

In accordance with Property 4.1, for $k \ge 2$, we define $\underline{t_k}$ to be the first round in which the action profile corresponding to payoff k is played with positive probability and $\overline{t_k}$ the last round before the action profile corresponding to payoff $k+1$ is played with positive probability. We then define $T_k := \overline{t_k} - \underline{t_k} + 1$ to be the number of rounds corresponding to the period $[\underline{t_k}, \overline{t_k}]$.

We also define $\mathcal{A}_1(k) := \{a_1(k') : 2m-1 \ge k' \ge k\}$ and $\mathcal{A}_2(k) := \{a_2(k') : 2m-1 \ge k' \ge k\}$. These are the rows and columns, respectively, that will be played after stop playing the action profile corresponding to k. The next crucial lemma shows that before an action becomes desirable, it will have accumulated very negative regret in the previous rounds.

Lemma 4.2. *For any even $k \ge 4$, let $\boldsymbol{r}_1^{(\overline{t_{k-2}})}[a_1]$ be the regret of Player 1 with respect to any action $a_1 \in \mathcal{A}_1(k)$. Then $\boldsymbol{r}_1^{(\overline{t_{k-2}})}[a_1] \le -\sum_{l=2}^{k-2}(l-1)T_l$. Similarly, for any odd $k \ge 5$, if $\boldsymbol{r}_2^{(\overline{t_{k-2}})}[a_2]$ is the regret of Player 2 with respect to any action $a_2 \in \mathcal{A}_2(k)$, $\boldsymbol{r}_2^{(\overline{t_{k-2}})}[a_2] \le -\sum_{l=2}^{k-2}(l-1)T_l$.*

At the same time, when an action has very negative regret, it will take a long time before that action gets played with positive probability, as formalized below.

Lemma 4.3. *For any even $k \ge 4$, $T_k \ge -\frac{1}{2}\boldsymbol{r}_2^{(\overline{t_{k-1}})}[a_2(k+1)]$. Similarly, for every odd $k \ge 5$, $T_k \ge -\frac{1}{2}\boldsymbol{r}_1^{(\overline{t_{k-1}})}[a_1(k+1)]$.*

By Lemmas 4.2 and 4.3, it follows that $T_k \ge \sum_{l=2}^{k-1}\frac{l-1}{2}T_l$ for any $k \ge 4$. By the inductive basis, we know that $T_3 \ge 1$. As a result, $T_k \ge \frac{k-2}{2}T_{k-1} \ge \frac{k-2}{2}\frac{k-3}{2}\cdots\frac{2}{2}T_3 \ge \frac{(k-2)!}{2^{k-3}}$ for all $k \ge 4$.

Moreover, it takes as least T_{2m-2} rounds to converge to an NE with approximation gap at most $1/2m+2$ (Lemma C.10). We thus arrive at the following exponential lower bound.

Theorem 4.4. *Simultaneous RM requires $m^{\Omega(m)}$ rounds to converge to a $\frac{1}{2m+2}$-Nash equilibrium.*

The same reasoning applies to alternating RM; unlike Section 3, here we update each player even if the best-response gap is arbitrarily small, although this does not qualitatively affect the lower bound.

Corollary 4.5. *Alternating RM requires $m^{\Omega(m)}$ rounds to converge to a $\frac{1}{2m+2}$-Nash equilibrium.*

5 Future Research

Our paper sheds new light on the convergence properties of regret matching(+) in constrained optimization problems, and potential games in particular. We showed that alternating RM$^+$ is a sound and fast first-order optimizer, while, on the flip side, RM can be exponentially slow even in potential games. Several interesting questions remain open. Does *simultaneous* RM$^+$ always achieve fast convergence? And does RM asymptotically converge even under alternating updates?

REFERENCES

Robert Aumann. Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1:67–96, 1974.

Yakov Babichenko and Aviad Rubinstein. Settling the complexity of Nash equilibrium in congestion games. In *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2021.

David Blackwell. An analog of the minmax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

Michael Bowling, Neil Burch, Michael Johanson, and Oskari Tammelin. Heads-up limit hold'em poker is solved. *Science*, 347(6218), January 2015.

Noam Brown and Tuomas Sandholm. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, pp. eaao1733, Dec. 2017.

Noam Brown and Tuomas Sandholm. Solving imperfect-information games via discounted regret minimization. In *Conference on Artificial Intelligence (AAAI)*, 2019a.

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

Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Weiqiang Zheng. Fast last-iterate convergence of learning in games requires forgetful algorithms. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

Yang Cai, Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo, and Weiqiang Zheng. Last-iterate convergence properties of regret-matching algorithms in games. In *International Conference on Learning Representations (ICLR)*, 2025.

Volkan Cevher, Ashok Cutkosky, Ali Kavis, Georgios Piliouras, Stratis Skoulakis, and Luca Viano. Alternation makes the adversary weaker in two-player games. In *Neural Information Processing Systems*, 2023.

Darshan Chakrabarti, Julien Grand-Clément, and Christian Kroer. Extensive-form game solving via blackwell approachability on treeplexes. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2024.

Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM*, 2009.

Constantinos Daskalakis, Paul Goldberg, and Christos Papadimitriou. The complexity of computing a Nash equilibrium. *SIAM Journal on Computing*, 2008.

Gabriele Farina, Christian Kroer, and Tuomas Sandholm. Faster game solving via predictive Blackwell approachability: Connecting regret matching and mirror descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, and Haipeng Luo. Regret matching+:(in) stability and fast convergence in games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

Yoav Freund and Robert Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

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

Sergiu Hart and Andreu Mas-Colell. Regret-based continuous-time dynamics. *Games and Economic Behavior*, 45(2):375–394, 2003.

Robert Kleinberg, Georgios Piliouras, and Éva Tardos. Multiplicative updates outperform generic no-regret learning in congestion games: extended abstract. In Michael Mitzenmacher (ed.), *Proceedings of the Annual Symposium on Theory of Computing (STOC)*, 2009.

Tai-Yu Ma and Philippe Gerber. Distributed regret matching algorithm for dynamic congestion games with information provision. *Transportation Research Procedia*, 3:3–12, 2014.

Jason R. Marden, Gürdal Arslan, and Jeff S. Shamma. Regret based dynamics: convergence in weakly acyclic games. In *Autonomous Agents and Multi-Agent Systems*, 2007.

Linjian Meng, Youzhi Zhang, Zhenxing Ge, Tianpei Yang, and Yang Gao. Asynchronous predictive counterfactual regret minimization$^+$ algorithm in solving extensive-form games. *arXiv:2503.12770*, 2025.

Dov Monderer and Lloyd S Shapley. Potential games. *Games and Economic Behavior*, 14(1): 124–143, 1996.

Matej Moravčík, Martin Schmid, Neil Burch, Viliam Lisý, Dustin Morrill, Nolan Bard, Trevor Davis, Kevin Waugh, Michael Johanson, and Michael Bowling. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, May 2017.

H. Moulin and J.-P. Vial. Strategically zero-sum games: The class of games whose completely mixed equilibria cannot be improved upon. *International Journal of Game Theory*, 7(3-4):201–221, 1978.

John Nash. Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36:48–49, 1950.

Ioannis Panageas, Nikolas Patris, Stratis Skoulakis, and Volkan Cevher. Exponential lower bounds for fictitious play in potential games. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

Robert W Rosenthal. A class of games possessing pure-strategy nash equilibria. *International Journal of Game Theory*, 2(1):65–67, 1973.

Oskari Tammelin. Solving large imperfect information games using CFR+. *arXiv:1407.5042*, 2014.

Emanuel Tewolde, Brian Hu Zhang, Ioannis Anagnostides, Tuomas Sandholm, and Vince Conitzer. Decision making under imperfect recall: Algorithms and benchmarks. In *Uncertainty in Artificial Intelligence (UAI)*, 2025.

Andre Wibisono, Molei Tao, and Georgios Piliouras. Alternating mirror descent for constrained min-max games. In *Neural Information Processing Systems*, 2022.

Hang Xu, Kai Li, Haobo Fu, Qiang Fu, Junliang Xing, and Jian Cheng. Dynamic discounted counterfactual regret minimization. In *International Conference on Learning Representations (ICLR)*, 2024a.

Hang Xu, Kai Li, Bingyun Liu, Haobo Fu, Qiang Fu, Junliang Xing, and Jian Cheng. Minimizing weighted counterfactual regret with optimistic online mirror descent. *arXiv:2404.13891*, 2024b.

Brian Hu Zhang and Tuomas Sandholm. General search techniques without common knowledge for imperfect-information games, and application to superhuman fog of war chess. *arXiv:2506.01242*, 2025.

Naifeng Zhang, Stephen McAleer, and Tuomas Sandholm. Faster game solving via hyperparameter schedules. *arXiv:2404.09097*, 2024.

Martin Zinkevich, Michael Bowling, Michael Johanson, and Carmelo Piccione. Regret minimization in games with incomplete information. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NIPS)*, 2007.

A Further related work

Much of the existing research on regret matching revolves around zero-sum games. Many variants have been proposed over the years to speed up its convergence (Xu et al., 2024b; Cai et al., 2025; Chakrabarti et al., 2024; Meng et al., 2025; Farina et al., 2021; Tammelin, 2014; Brown & Sandholm, 2019a). Some notable variations that have considerably improved performance are *predictive* RM and RM$^+$ (Farina et al., 2021), which rely on predicting the next utility, and *discounted* RM and RM$^+$ (Brown & Sandholm, 2019a; Zhang et al., 2024; Xu et al., 2024a), where one dynamically discounts the accumulated regret; in a similar vein, our work shows that a discounted variant of RM$^+$ achieves a better convergence upper bound than RM$^+$ in our setting (Corollary 3.5). It must be stressed that the focus of all that prior work was on zero-sum games. Constrained optimization is a fundamentally different problem. For one, in zero-sum games, it is only the average strategy of RM and RM$^+$ that converges, not the last iterate (Farina et al., 2023).

The recent paper of Tewolde et al. (2025) demonstrated that the regret matching family is a formidable first-order optimizer in constrained optimization problems. In particular, their focus was on (single-player) imperfect-recall problems, which are tantamount to general polynomial optimization problems over a product of simplices. Interestingly, many of the trends observed in zero-sum games are actually reversed in constrained optimization. For example, the predictive versions of RM and RM$^+$ generally performed worse than their non-predictive counterparts. One trend that did persist was the superiority of RM$^+$ over RM. It is also worth mentioning an earlier work by Ma & Gerber (2014) that also reported fast empirical convergence in a certain class of congestion games. Yet, there was hitherto no theoretical understanding of those algorithms in this setting. The main precursors of our work are the paper of Hart & Mas-Colell (2003), which established asymptotic convergence in discrete time but for a somewhat artificial variant of regret matching, and the paper of Marden et al. (2007), which analyzed asymptotically a certain variant of regret matching that aggressively discounts the regrets.

An interesting result that sheds light on RM and RM$^+$ is by Farina et al. (2021), who showed that RM can be obtained by running *follow the regularized leader (FTRL)* in a certain lifted space, whereas RM$^+$ can be obtained through *mirror descent (MD)* in the same space; this is despite the fact that, unlike FTRL and MD, RM and RM$^+$ are both parameter free. On a related note, Cai et al. (2024) showed that only forgetful algorithms—closer to MD than to FTRL—can attain fast last-iterate convergence. Our exponential separation of RM and RM$^+$ echoes their finding, although in a different setting and class of algorithms.

B Further background

Coarse correlated equilibria For completeness, we provide the definition of a coarse correlated equilibrium (Moulin & Vial, 1978), which is a relaxation of correlated equilibria (Aumann, 1974). The key connection that relates to our results is that if all players in a normal-form game have sublinear regret, the average correlated distribution of play converges to the set of coarse correlated equilibria. In particular, the rate of convergence is driven by the maximum of the players' regrets (Proposition B.2).

Definition B.1 (Coarse correlated equilibrium)**.** Consider an n-player game in normal form. A correlated distribution $\mu \in \Delta(\mathcal{A}_1 \times \cdots \times \mathcal{A}_n)$ is an *ϵ-coarse correlated equilibrium (CCE)* if for any player $i \in [n]$ and deviation $a_i' \in \mathcal{A}_i$,

$$\mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu} u_i(a_1,\ldots,a_n) \geq \mathop{\mathbb{E}}_{(a_1,\ldots,a_n)\sim\mu} u_i(a_i', a_{-i}) - \epsilon.$$

Proposition B.2. *If each player $i \in [n]$ observes the sequence of utilities $(\boldsymbol{u}_i(\boldsymbol{x}_{-i}^{(t)}))_{t=1}^T$, the average correlated distribution of play is an ϵ-CCE with $\epsilon \leq \frac{1}{T}\max_{1\leq i\leq n} \mathsf{Reg}_i^{(T)}$, where $\mathsf{Reg}_i^{(T)}$ is the regret of the ith player.*

This connection holds for simultaneous updates; it is unclear if and how it can be extended under alternating updates.

Discounting Next, we spell out regret matching$^+$ with discounting (DRM$^+$; Algorithm 3). The only difference from RM$^+$ is that the regret vector is multiplied by a discounting coefficient $\alpha^{(t)} \in (0, 1]$ in every round $t \in [T]$ (Line 12); the special case where $\alpha^{(t)} = 1$ for all $t \in [T]$ is RM$^+$.

Algorithm 3: Regret matching$^+$ with discounting (DRM$^+$)

1 **Input**: discounting coefficients $(\alpha^{(1)}, \dots, \alpha^{(T)}) \in (0, 1]^T$;
2 Initialize cumulative regrets $\boldsymbol{r}^{(0)} \coloneqq \mathbf{0}$;
3 Initialize strategy $\boldsymbol{x}^{(1)} \in \Delta(\mathcal{A})$;
4 **for** $t = 1, \dots, T$ **do**
5 Set $\boldsymbol{\theta}^{(t)} \leftarrow \boldsymbol{r}^{(t-1)}$;
6 **if** $\boldsymbol{\theta}^{(t)} \neq \mathbf{0}$ **then**
7 Compute $\boldsymbol{x}^{(t)} \leftarrow \boldsymbol{\theta}^{(t)} / \|\boldsymbol{\theta}^{(t)}\|_1$;
8 **else**
9 $\boldsymbol{x}^{(t)} \leftarrow \boldsymbol{x}^{(t-1)}$;
10 Output strategy $\boldsymbol{x}^{(t)} \in \Delta(\mathcal{A})$;
11 Observe utility $\boldsymbol{u}^{(t)} \in \mathbb{R}^{\mathcal{A}}$;
12 $\boldsymbol{r}^{(t)} \leftarrow \alpha^{(t)}[\boldsymbol{r}^{(t-1)} + \boldsymbol{u}^{(t)} - \langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle \mathbf{1}]^+$;

C Omitted proofs

This section provides the proofs missing from the main body. We begin by stating a simple lemma that bounds the regret of RM$^+$, implying Proposition 2.3; we will then adapt it to account for discounting per Algorithm 3.

Lemma C.1 (Regret vector upper bound)**.** *For any time $t \in [T]$, RM$^+$ guarantees $\|\boldsymbol{r}^{(t)}\|_2^2 \leq \|\boldsymbol{r}^{(t-1)}\|_2^2 + \|\boldsymbol{g}^{(t)}\|_2^2$, where $\boldsymbol{g}^{(t)} \coloneqq \boldsymbol{u}^{(t)} - \langle \boldsymbol{x}^{(t)}, \boldsymbol{u}^{(t)} \rangle$ is the instantaneous regret at time t.*

Proof. By definition of RM$^+$, $\langle \boldsymbol{r}^{(t-1)}, \boldsymbol{g}^{(t)} \rangle = \langle \boldsymbol{x}^{(t)}, \boldsymbol{g}^{(t)} \rangle = 0$ since $\boldsymbol{x}^{(t)} \propto \boldsymbol{r}^{(t-1)}$. Thus,

$$\|\boldsymbol{r}^{(t)}\|_2^2 = \|[\boldsymbol{r}^{(t-1)} + \boldsymbol{g}^{(t)}]^+\|_2^2 \leq \|\boldsymbol{r}^{(t-1)} + \boldsymbol{g}^{(t)}\|_2^2 = \|\boldsymbol{r}^{(t-1)}\|_2^2 + \|\boldsymbol{g}^{(t)}\|_2^2,$$

by orthogonality. □

As a result, the telescopic summation yields $\|\boldsymbol{r}^{(T)}\|_2^2 \leq \sum_{t=1}^T \|\boldsymbol{g}^{(t)}\|_2^2 \leq mT$ since $\|\boldsymbol{g}^{(t)}\|_\infty \leq 1$ (by the assumption that the range of the utilities is bounded by 1). A similar proof works for RM. We now adapt Lemma C.1 for DRM$^+$.

Lemma C.2. *For any time $t \in [T]$, DRM$^+$ guarantees $\|\boldsymbol{r}^{(t)}\|_2^2 \leq (\alpha^{(t)})^2(\|\boldsymbol{r}^{(t-1)}\|_2^2 + \|\boldsymbol{g}^{(t)}\|_2^2)$.*

Proof. As before, $\langle \boldsymbol{r}^{(t-1)}, \boldsymbol{g}^{(t)} \rangle = \langle \boldsymbol{x}^{(t)}, \boldsymbol{g}^{(t)} \rangle = 0$ since $\boldsymbol{x}^{(t)} \propto \boldsymbol{r}^{(t-1)}$. Thus,

$$\|\boldsymbol{r}^{(t)}\|_2^2 = (\alpha^{(t)})^2\|[\boldsymbol{r}^{(t-1)} + \boldsymbol{g}^{(t)}]^+\|_2^2 \leq (\alpha^{(t)})^2\|\boldsymbol{r}^{(t-1)} + \boldsymbol{g}^{(t)}\|_2^2 \leq (\alpha^{(t)})^2(\|\boldsymbol{r}^{(t-1)}\|_2^2 + \|\boldsymbol{g}^{(t)}\|_2^2).$$

□

A direct consequence is the following bound on the regret vector.

Corollary C.3. *For any time $t \in [T]$, DRM$^+$ guarantees*

$$\|\boldsymbol{r}^{(t)}\|_2^2 \leq (\alpha^{(t)})^2\|\boldsymbol{g}^{(t)}\|_2^2 + (\alpha^{(t)}\alpha^{(t-1)})^2\|\boldsymbol{g}^{(t-1)}\|_2^2 + \cdots + \left(\prod_{\tau=1}^{t} \alpha^{(\tau)}\right)^2 \|\boldsymbol{g}^{(1)}\|_2^2.$$

In particular, if $\alpha^{(t)} = 1 - \gamma$ for some constant $\gamma \in (0, 1)$, it follows that $\|\boldsymbol{r}^{(T)}\|_2 \leq \sqrt{m/\gamma}$.

C.1 Proofs from Section 3

We continue with the proofs from Section 3. We first establish that RM$^+$ enjoys a one-step improvement property when the utility is linear.

Lemma 3.3 (One-step improvement for RM$^+$)**.** *For any $\boldsymbol{r} \in \mathbb{R}^{\mathcal{A}}_{\geq 0}$ and $\boldsymbol{u} \in \mathbb{R}^{\mathcal{A}}$, we define $\boldsymbol{x} \coloneqq \boldsymbol{r}/\|\boldsymbol{r}\|_1$; if $\boldsymbol{r} = \boldsymbol{0}$, $\boldsymbol{x} \in \Delta(\mathcal{A})$ can be arbitrary. If $\boldsymbol{r}' \coloneqq [\boldsymbol{r} + \boldsymbol{u} - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \boldsymbol{1}]^+ \neq \boldsymbol{0}$ and $\boldsymbol{x}' \coloneqq \boldsymbol{r}'/\|\boldsymbol{r}'\|_1$,*

$$\langle \boldsymbol{x}' - \boldsymbol{x}, \boldsymbol{u}\rangle \geq \frac{1}{\|\boldsymbol{r}'\|_1}\left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2 = \frac{1}{\|\boldsymbol{r}'\|_1}\mathsf{BRGap}(\boldsymbol{u})^2. \tag{3}$$

If $\boldsymbol{r}' = \boldsymbol{0}$, then $\langle \boldsymbol{x}, \boldsymbol{u}\rangle = \langle \boldsymbol{x}', \boldsymbol{u}\rangle \geq \max_{a\in\mathcal{A}} \boldsymbol{u}[a]$.

Proof. First, if $\boldsymbol{r}' = \boldsymbol{0}$, it follows that $\boldsymbol{r} + \boldsymbol{u} - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \boldsymbol{1} \leq \boldsymbol{0}$, where the inequality is to be taken coordinate-wise. Since $\boldsymbol{r} \geq \boldsymbol{0}$, we have $\langle \boldsymbol{x}, \boldsymbol{u}\rangle \geq \boldsymbol{u}[a]$ for all $a \in \mathcal{A}$, as claimed.

We now assume $\boldsymbol{r}' \neq \boldsymbol{0}$. If $\boldsymbol{r} = \boldsymbol{0}$, we have $\boldsymbol{r}' = [\boldsymbol{u} - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \boldsymbol{1}]^+$. (3) can then be equivalently expressed as

$$\sum_{a\in\mathcal{A}} \boldsymbol{r}'[a](\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle) \geq \left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2,$$

which holds since $\boldsymbol{r}' = [\boldsymbol{u} - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \boldsymbol{1}]^+$. So we can assume $\boldsymbol{r} \neq \boldsymbol{0}$. We define $\boldsymbol{\delta} \coloneqq \boldsymbol{r}' - \boldsymbol{r}$. (3) can be expressed as

$$\frac{\sum_{a\in\mathcal{A}}(\boldsymbol{r}[a] + \boldsymbol{\delta}[a])\boldsymbol{u}[a]}{\sum_{a'\in\mathcal{A}}(\boldsymbol{r}[a'] + \boldsymbol{\delta}[a'])} \geq \frac{\sum_{a\in\mathcal{A}} \boldsymbol{r}[a]\boldsymbol{u}[a]}{\sum_{a'\in\mathcal{A}} \boldsymbol{r}[a']} + \frac{(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle)^2}{\sum_{a'\in\mathcal{A}}(\boldsymbol{r}[a'] + \boldsymbol{\delta}[a'])}.$$

Equivalently,

$$\begin{aligned}\sum_{a'\in\mathcal{A}} \boldsymbol{r}[a'] \sum_{a\in\mathcal{A}}(\boldsymbol{r}[a] + \boldsymbol{\delta}[a])\boldsymbol{u}[a] \geq \sum_{a\in\mathcal{A}} \boldsymbol{r}[a] \sum_{a'\in\mathcal{A}}(\boldsymbol{r}[a'] + \boldsymbol{\delta}[a'])\boldsymbol{u}[a] \\ + \sum_{a'\in\mathcal{A}} \boldsymbol{r}[a']\left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2.\end{aligned}$$

This in turn equivalent to

$$\begin{aligned}\sum_{a'\in\mathcal{A}} \boldsymbol{r}[a'] \sum_{a\in\mathcal{A}} \boldsymbol{\delta}[a]\boldsymbol{u}[a] &\geq \sum_{a\in\mathcal{A}} \boldsymbol{r}[a] \sum_{a'\in\mathcal{A}} \boldsymbol{\delta}[a']\boldsymbol{u}[a] + \sum_{a'\in\mathcal{A}} \boldsymbol{r}[a']\left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2 \\ &= \sum_{a'\in\mathcal{A}} \boldsymbol{\delta}[a'] \sum_{a\in\mathcal{A}} \boldsymbol{r}[a]\langle \boldsymbol{x}, \boldsymbol{u}\rangle + \sum_{a'\in\mathcal{A}} \boldsymbol{r}[a']\left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2.\end{aligned}$$

Rearranging,

$$\sum_{a'\in\mathcal{A}} \boldsymbol{r}[a'] \sum_{a\in\mathcal{A}} \boldsymbol{\delta}[a](\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle) \geq \sum_{a'\in\mathcal{A}} \boldsymbol{r}[a']\left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2.$$

Now, for any $a \in \mathcal{A}$ such that $\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \geq 0$, it follows that $\boldsymbol{\delta}[a] = \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \geq 0$; on the other hand, for $a \in \mathcal{A}$ such that $\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle < 0$, we have $\boldsymbol{\delta}[a] \leq 0$. That is, $\boldsymbol{\delta}[a](\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle) \geq 0$, and the claim follows. □

We will now show that Lemma 3.3 is, in a certain sense, tight. We consider a simple linear maximization over the simplex. If the regret vector of RM$^+$ can be initialized arbitrarily, as is the premise in Lemma 3.3, we make the following observation.

Lemma C.4. *Consider a utility vector $\boldsymbol{u} \in \mathbb{R}^{\mathcal{A}}$ and some initial regret vector $\mathbb{R}^{\mathcal{A}}_{\geq 0} \ni \boldsymbol{r}^{(1)} \neq \boldsymbol{0}$. If $\boldsymbol{x}^{(1)} = \boldsymbol{r}^{(1)}/\|\boldsymbol{r}^{(1)}\|_1$ and $\epsilon = \max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}^{(1)}, \boldsymbol{u}\rangle$ is the initial best-response gap, it takes at least $\|\boldsymbol{r}^{(1)}\|_1/2\epsilon$ iterations for RM$^+$ to reach a point $\boldsymbol{x}^{(t)}$ with best-response gap at most $\epsilon/2$.*

Indeed, we consider the two-dimensional problem in which $\boldsymbol{u} = (1-\epsilon, 1)$ and $\boldsymbol{r} = (\|\boldsymbol{r}^{(1)}\|_1, 0)$. To incur a best-response gap of at most $\epsilon/2$, the player needs to allot a probability mass of at least $1/2$ to the second action. In the meantime, the decrement of the first coordinate of $\boldsymbol{r}^{(t)}$ will be at most ϵ while the increment of the second coordinate of $\boldsymbol{r}^{(t)}$ will be at most ϵ. But it must be the case that the second coordinate of $\boldsymbol{r}$ is at least as large as the first coordinate of $\boldsymbol{r}$, leading to Lemma C.4. Given that $\langle \boldsymbol{x}^{(t)} - \boldsymbol{x}^{(1)}, \boldsymbol{u}\rangle \le \epsilon$, this matches the bound obtained for this problem through Lemma 3.3 in the regime where $\|\boldsymbol{r}\|_1$ is at least as large as $1/\epsilon$ (so that the norm of $\boldsymbol{r}^{(t)}$ is within a constant factor of $\boldsymbol{r}^{(1)}$, by Lemma C.1).

Unlike $\texttt{RM}^+$, RM only has a *conditional* one-step improvement because the regret vector can have negative coordinates.

Lemma 3.6. *For any $\boldsymbol{r} \in \mathbb{R}^{\mathcal{A}}_{\ge 0}$ and $\boldsymbol{u} \in \mathbb{R}^{\mathcal{A}}$, we define $\boldsymbol{x} \coloneqq \boldsymbol{\theta}/\|\boldsymbol{\theta}\|_1$, where $\boldsymbol{\theta} \coloneqq \max(\boldsymbol{r}, \boldsymbol{0})$; if $\boldsymbol{\theta} = \boldsymbol{0}$, $\boldsymbol{x} \in \Delta(\mathcal{A})$ can be arbitrary. If $\boldsymbol{r}' \coloneqq \boldsymbol{r} + \boldsymbol{u} - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \boldsymbol{1}$ and $\boldsymbol{x}' \coloneqq \boldsymbol{\theta}'/\|\boldsymbol{\theta}'\|_1$, where $\boldsymbol{\theta}' = \max(\boldsymbol{r}', \boldsymbol{0}) \neq \boldsymbol{0}$, we have $\langle \boldsymbol{x}' - \boldsymbol{x}, \boldsymbol{u}\rangle \ge \frac{1}{\|\boldsymbol{\theta}'\|_1}\|\boldsymbol{\theta}' - \boldsymbol{\theta}\|_2^2 \ge \frac{1}{\|\boldsymbol{\theta}'\|_1}\left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2 \mathbb{1}\{\boldsymbol{r}[a] \ge 0\}$, where $a \in \arg\max_{a'\in\mathcal{A}} \boldsymbol{u}[a']$. If $\boldsymbol{\theta}' = \boldsymbol{0}$, then $\langle \boldsymbol{x}, \boldsymbol{u}\rangle = \langle \boldsymbol{x}', \boldsymbol{u}\rangle \ge \max_{a\in\mathcal{A}} \boldsymbol{u}[a]$.*

Proof. We define $\boldsymbol{\delta} \coloneqq \boldsymbol{\theta}' - \boldsymbol{\theta}$. Following the proof of Lemma 3.3, it suffices to show that

$$\sum_{a'\in\mathcal{A}} \boldsymbol{\theta}[a'] \sum_{a\in\mathcal{A}} \boldsymbol{\delta}[a](\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle) \ge \sum_{a'\in\mathcal{A}} \boldsymbol{\theta}[a'] \left(\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle\right)^2 \mathbb{1}\{\boldsymbol{r}[a] \ge 0\}. \tag{6}$$

For an action $a \in \mathcal{A}$, we consider the following cases.

- If $\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle \ge 0$,
  - if $\boldsymbol{r}[a] \ge 0$, we have $\boldsymbol{\delta}[a] = \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle$.
  - If $\boldsymbol{r}[a] < 0$, it follows that $\boldsymbol{\delta}[a] \ge 0$; in particular, $\boldsymbol{\delta}[a] = 0$ if $\boldsymbol{r}'[a] \le 0$ and $\boldsymbol{\delta}[a] > 0$ otherwise. As a result, $\boldsymbol{\delta}[a](\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle) \ge 0$.
- If $\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle < 0$,
  - if $\boldsymbol{r}[a] \le 0$, we have $\boldsymbol{\delta}[a] = 0$ since $\boldsymbol{\theta}[a] = 0 = \boldsymbol{\theta}'[a]$.
  - if $\boldsymbol{r}[a] > 0$, it follows that $\boldsymbol{\delta}[a] < 0$. Again, we have $\boldsymbol{\delta}[a](\boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle) \ge 0$.

Combining those items, (6) follows. □

We next state a direct refinement of Lemma 3.3 that we rely on in the more general setting of constrained optimization.

Lemma C.5 (Refinement of Lemma 3.3)**.** *Under the preconditions of Lemma 3.3,*

$$\langle \boldsymbol{x}' - \boldsymbol{x}, \boldsymbol{u}\rangle \ge \frac{1}{\|\boldsymbol{r}'\|_1}\|\boldsymbol{r} - \boldsymbol{r}'\|_2^2. \tag{7}$$

In particular, (7) implies (3) since $\|\boldsymbol{r} - \boldsymbol{r}'\|_2^2 \ge (\max_{a\in\mathcal{A}} \boldsymbol{u}[a] - \langle \boldsymbol{x}, \boldsymbol{u}\rangle)^2$, by definition of $\boldsymbol{r}'$. The proof of Lemma C.5 is identical to that of Lemma 3.3.

The next elementary lemma shows that, so long as the norm of the regret vector is not too small, closeness in regrets implies closeness in strategies.

Lemma C.6. *For $\mathbb{R}^{\mathcal{A}}_{\ge 0} \ni \boldsymbol{r}, \boldsymbol{r}' \neq \boldsymbol{0}$, let $\boldsymbol{x} \coloneqq \boldsymbol{r}/\|\boldsymbol{r}\|_1$ and $\boldsymbol{x}' \coloneqq \boldsymbol{r}'/\|\boldsymbol{r}'\|_1$. Then*

$$\|\boldsymbol{x} - \boldsymbol{x}'\|_1 \le \|\boldsymbol{r} - \boldsymbol{r}'\|_1 \left(\frac{1}{\|\boldsymbol{r}\|_1} + \frac{1}{\|\boldsymbol{r}'\|_1}\right).$$

Proof. The term $\boldsymbol{x}[a] - \boldsymbol{x}'[a]$ can be expressed, for any $a \in \mathcal{A}$, as

$$\begin{aligned}\frac{\boldsymbol{r}[a]}{\sum_{a'\in\mathcal{A}} \boldsymbol{r}[a']} - \frac{\boldsymbol{r}'[a]}{\sum_{a'\in\mathcal{A}} \boldsymbol{r}'[a']} &= \frac{\sum_{a'\in\mathcal{A}} (\boldsymbol{r}[a]\boldsymbol{r}'[a'] - \boldsymbol{r}'[a]\boldsymbol{r}[a'])}{\|\boldsymbol{r}\|_1\|\boldsymbol{r}'\|_1} \\ &= \frac{\sum_{a'\in\mathcal{A}} (\boldsymbol{r}[a](\boldsymbol{r}'[a'] - \boldsymbol{r}[a']) + \boldsymbol{r}[a'](\boldsymbol{r}[a] - \boldsymbol{r}'[a]))}{\|\boldsymbol{r}\|_1\|\boldsymbol{r}'\|_1},\end{aligned}$$

and the claim follows. □

Combining Lemmas C.5 and C.6, we now formally show that $\texttt{RM}^+$ improves the value of the underlying function when the norm of the regret vector is not too small.

Lemma 3.7. *Let u be an L-smooth function over $\Delta(\mathcal{A})$. For any $\boldsymbol{r} \in \mathbb{R}^{\mathcal{A}}_{\geq 0}$ with $\boldsymbol{r} \neq \mathbf{0}$, we define $\boldsymbol{x} := \boldsymbol{r}/\|\boldsymbol{r}\|_1$. Further, let $\boldsymbol{r}' := [\boldsymbol{r} + \nabla u(\boldsymbol{x}) - \langle \boldsymbol{x}, \nabla u(\boldsymbol{x})\rangle \mathbf{1}]^+ \neq \mathbf{0}$ and $\boldsymbol{x}' := \boldsymbol{r}'/\|\boldsymbol{r}'\|_1$. If $\|\boldsymbol{r}'\|_2 \geq \max\{2m, 9mL\}$, then $u(\boldsymbol{x}') - u(\boldsymbol{x}) \geq \frac{1}{2\|\boldsymbol{r}'\|_1} \left(\max_{\boldsymbol{x}^\star \in \Delta(\mathcal{A})} \langle \boldsymbol{x}^\star - \boldsymbol{x}, \nabla u(\boldsymbol{x})\rangle\right)^2$.*

Proof. Using the quadratic bound for u, we have

$$\begin{aligned} u(\boldsymbol{x}') - u(\boldsymbol{x}) &\geq \langle \nabla u(\boldsymbol{x}), \boldsymbol{x}' - \boldsymbol{x}\rangle - \frac{L}{2}\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2 \\ &\geq \frac{1}{\|\boldsymbol{r}'\|_1}\|\boldsymbol{r} - \boldsymbol{r}'\|_2^2 - \frac{L}{2}\|\boldsymbol{r} - \boldsymbol{r}'\|_1^2 \left(\frac{1}{\|\boldsymbol{r}\|_1} + \frac{1}{\|\boldsymbol{r}'\|_1}\right)^2 \qquad (8) \\ &\geq \frac{1}{\|\boldsymbol{r}'\|_1}\|\boldsymbol{r} - \boldsymbol{r}'\|_2^2 - \frac{9mL}{2\|\boldsymbol{r}'\|_1^2}\|\boldsymbol{r} - \boldsymbol{r}'\|_2^2 \qquad (9) \\ &\geq \frac{1}{2\|\boldsymbol{r}'\|_1}\|\boldsymbol{r} - \boldsymbol{r}'\|_2^2, \qquad (10) \end{aligned}$$

where (8) uses the one-step improvement property (Lemma C.5) applied for $\boldsymbol{u} := \nabla u(\boldsymbol{x})$ together with Lemma C.6; (9) follows from the fact that $\|\boldsymbol{r}\|_1 \geq \|\boldsymbol{r}'\|_1 - m \geq \frac{1}{2}\|\boldsymbol{r}'\|_1$ since $\|\boldsymbol{r}'\|_1 \geq \|\boldsymbol{r}'\|_2 \geq 2m$ and the $|\langle \boldsymbol{x} - \boldsymbol{x}', \nabla u(\boldsymbol{x})\rangle| \leq 1$ for all $\boldsymbol{x}' \in \Delta(\mathcal{A})$ (per our normalization assumption); and (10) follows from the assumption that $\|\boldsymbol{r}'\|_1 \geq 9mL$. □

To make use of Lemma 3.7, we next establish that the ℓ_2 norm of the regret vector is nondecreasing.

Lemma 3.8. *For any t, $\texttt{RM}^+$ guarantees $\|\boldsymbol{r}^{(t)}\|_2^2 \geq \|\boldsymbol{r}^{(t-1)}\|_2^2 + \|[\boldsymbol{g}^{(t)}]^+\|_2^2$, where $\boldsymbol{g}^{(t)} := \nabla u(\boldsymbol{x}) - \langle \nabla u(\boldsymbol{x}), \boldsymbol{x}^{(t)}\rangle \mathbf{1}$ is the instantaneous regret at round t.*

Proof. We have $\boldsymbol{r}^{(t)} - \boldsymbol{r}^{(t-1)} = \max(\boldsymbol{g}^{(t)}, -\boldsymbol{r}^{(t-1)})$ (element-wise), so

$$\|\boldsymbol{r}^{(t)} - \boldsymbol{r}^{(t-1)}\|_2 = \|\max(\boldsymbol{g}^{(t)}, -\boldsymbol{r}^{(t-1)})\|_2 \geq \|[\boldsymbol{g}^{(t)}]^+\|_2.$$

Further, $\langle \boldsymbol{r}^{(t-1)}, \boldsymbol{r}^{(t)} - \boldsymbol{r}^{(t-1)}\rangle = \langle \boldsymbol{r}^{(t-1)}, \max(\boldsymbol{g}^{(t)}, -\boldsymbol{r}^{(t-1)})\rangle \geq \langle \boldsymbol{r}^{(t-1)}, \boldsymbol{g}^{(t)}\rangle = 0$, where we used the fact that $\boldsymbol{r}^t \geq \mathbf{0}$, element-wise. Therefore,

$$\begin{aligned} \|\boldsymbol{r}^{(t)}\|_2^2 = \|\boldsymbol{r}^{(t)} - \boldsymbol{r}^{(t-1)} + \boldsymbol{r}^{(t-1)}\|_2^2 &= \|\boldsymbol{r}^{(t)} - \boldsymbol{r}^{(t-1)}\|_2^2 + \|\boldsymbol{r}^{(t-1)}\|_2^2 + 2\langle \boldsymbol{r}^{(t-1)}, \boldsymbol{r}^{(t)} - \boldsymbol{r}^{(t-1)}\rangle \\ &\geq \|\boldsymbol{r}^{(t-1)}\|_2^2 + \|[\boldsymbol{g}^{(t)}]^+\|_2^2, \end{aligned}$$

as claimed. □

Armed with Lemmas 3.7 and 3.8, we can now prove Theorem 3.9.

Theorem 3.9. *Let u be an L-smooth function in $\Delta(\mathcal{A}) \subset \mathbb{R}^m$ with range u_{range} and $R := \max\{2m, 9mL\}$. $\texttt{RM}^+$ requires at most $1 + (m(2u_{\text{range}} + R^2))^2/\epsilon^4$ rounds to reach an ϵ-KKT point.*

Proof. Let $t_c \in [T]$ be the largest t such that $\|\boldsymbol{r}^{(t)}\|_2 < \max\{2m, 9mL\} = R$. By Lemma 3.8,

$$\|\boldsymbol{r}^{(t)}\|_2^2 \geq \|\boldsymbol{r}^{(t-1)}\|_2^2 + \|[\boldsymbol{g}^{(t)}]_+\|_2^2 \geq \|\boldsymbol{r}^{(t-1)}\|_2^2 + \mathsf{KKTGap}(\boldsymbol{x}^{(t)})^2;$$

so,

$$\sum_{t=1}^{t_c} \mathsf{KKTGap}(\boldsymbol{x}^{(t)})^2 \leq \sum_{t=1}^{t_c} (\|\boldsymbol{r}^{(t)}\|_2^2 - \|\boldsymbol{r}^{(t-1)}\|_2^2) = \|\boldsymbol{r}^{(t_c)}\|_2^2 \leq R^2. \qquad (11)$$

Further, for any $t \geq t_c + 1$, we have $\|\boldsymbol{r}^{(t)}\|_2 \geq R$ since the ℓ_2 norm of the regret vector is nondecreasing (Lemma 3.8) and $\|\boldsymbol{r}^{(t_c+1)}\|_2 \geq R$. Thus, by Lemma 3.7,

$$\sum_{t=t_c+1}^{T} \frac{1}{2\|\boldsymbol{r}^{(t)}\|_1} \mathsf{KKTGap}(\boldsymbol{x}^{(t)})^2 \leq u(\boldsymbol{x}^{(T+1)}) - u(\boldsymbol{x}^{(t_c+1)}) \leq u_{\text{range}}. \qquad (12)$$

Combining (11) and (12), together with the fact that $\|\boldsymbol{r}^{(t)}\|_1 \le m\sqrt{t}$,

$$\sum_{t=1}^{T} \frac{1}{m\sqrt{t}} \mathsf{KKTGap}(\boldsymbol{x}^{(t)})^2 \le 2u_{\mathsf{range}} + R^2.$$

□

Theorem 3.11. *Let u be an L-smooth function in $\Delta(\mathcal{A}_1) \times \cdots \times \Delta(\mathcal{A}_n)$ with range u_{range} and $R := \sqrt{\sum_{i=1}^{n} \max\{2m_i, 9m_iL\}^2}$. Alternating $\mathtt{RM}^+$ requires at most $1 + {}^{(mn^2(2u_{\mathsf{range}}+R^2))^2}/_{\epsilon^4}$ rounds to reach an ϵ-KKT point of u.*

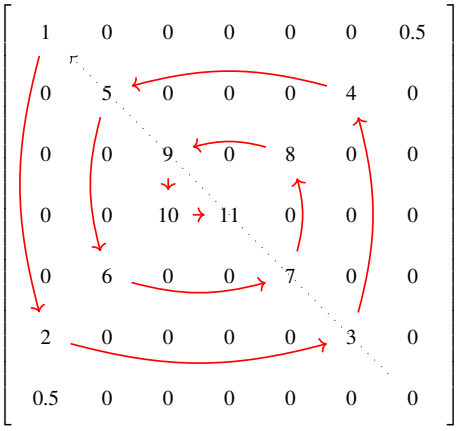
Proof. Following the previous argument in Theorem 3.9,

$$\sum_{t=1}^{T}\sum_{i=1}^{n} \frac{1}{\|\boldsymbol{r}_i^{(t)}\|_1} \mathsf{BRGap}_i(\boldsymbol{u}_i^{(t)})^2 \mathbb{1}\{\mathsf{BRGap}_i(\boldsymbol{u}_i^{(t)}) > \epsilon\} \le 2u_{\mathsf{range}} + \sum_{i=1}^{n} \max\{2m_i, 9m_iL\}^2.$$

Since $\|\boldsymbol{r}_i^{(t)}\|_1 \le m\sqrt{t}$ for all $i \in [n]$ and $t \in [T]$, it will take at most $1 + {}^{(m(2u_{\mathsf{range}}+R^2))^2}/_{\epsilon^4}$ rounds to converge to a point in which all players have at most an ϵ best-response gap, which in turn implies that the KKT gap is at most $n\epsilon$. Rescaling ϵ concludes the proof. □

C.2 Proofs from Section 4

We conclude with the proofs from Section 4. Below, we provide an illustrative example of the matrix $\mathbf{B}$, defined earlier in (5), for $m = 6$. The plots in Figure 1 are obtained by running simultaneous RM (left) and alternating $\mathtt{RM}^+$ (right) on this exact game.



Our main goal is to prove the following invariance.

Property 4.1. *After the first round both players play the first action. Thereupon, either the players play with probability 1 $(a_1(k), a_2(k))$, or, when k is odd, only Player 1 (respectively, Player 2 when k is even) mixes between $a_1(k)$ and $a_1(k+1)$ (respectively, $a_2(k)$ and $a_2(k+1)$). If a row or a column stops being played, it will never be played henceforth. An action profile $(a_1(k+1), a_2(k+1))$ is played with positive probability only if $(a_1(k), a_2(k))$ was played at some previous round.*

It is possible to check the claim for $k = 1, 2, 3, 4$ by executing RM for a number of rounds. In particular, we find that $T_3 \ge 5$ and $T_4 \ge 20$. We proceed by induction in k. Suppose that it holds for all payoffs $1, \ldots, \kappa$. We will show that it holds for $\kappa + 1$.

Lemma C.7. *For any even $\kappa + 2 \ge k \ge 4$, let $\boldsymbol{r}_1^{(t_{k-2})}[a_1]$ be the regret of Player 1 with respect to any action $a_1 \in \mathcal{A}_1(k)$. Then $\boldsymbol{r}_1^{(t_{k-2})}[a_1] \le -\sum_{l=2}^{k-2}(l-1)T_l$. Similarly, for any odd $\kappa + 2 \ge k \ge 5$, if $\boldsymbol{r}_2^{(t_{k-2})}[a_2]$ is the regret of Player 2 with respect to any action $a_2 \in \mathcal{A}_2(k)$, $\boldsymbol{r}_2^{(t_{k-2})}[a_2] \le -\sum_{l=2}^{k-2}(l-1)T_l$.*

Proof. Let $a_1 \in \mathcal{A}_1(k)$ and $l \in [k-2]$ for an even k. Playing $a_1 \in \mathcal{A}_1(k)$ during $[\underline{t_l}, \overline{t_l}]$ gives Player 1 a utility of 0; this follows from the fact that for any column $a_2 \in \{a_2(1), a_2(3), \ldots, a_2(k-3)\} = \{a_2(1), a_2(2), a_2(3), \ldots, a_2(k-3), a_2(k-2)\}$, it holds that $\mathbf{A}[a_1(k), a_2] = 0$, by construction of $\mathbf{A}$. At the same time, Player 1 actually got a utility of at least $l-1$ for each round in $[\underline{t_l}, \overline{t_l}]$. This means that every time Player 1 updates its regret vector within the time period $[\underline{t_l}, \overline{t_l}]$, the regret of a_1 decreases by at least $l-1$. The same reasoning applies for Player 2 when k is odd. □

Lemma C.8. *For any even $\kappa \geq k \geq 4$, $T_k \geq -\frac{1}{2}\boldsymbol{r}_2^{(\overline{t_{k-1}})}[a_2(k+1)]$. Similarly, for every odd $\kappa \geq k \geq 5$, $T_k \geq -\frac{1}{2}\boldsymbol{r}_1^{(\overline{t_{k-1}})}[a_1(k+1)]$.*

Proof. T_k is at least as large as the number of rounds it takes for $a_2(k+1)$ to have nonnegative regret. But in every round in $[\underline{t_k}, \overline{t_k}]$ the regret of $a_2(k)$ can increase additively by at most 2. The same reasoning applies when k is odd. □

The following upper bound on the regret is crude, but will suffice for our purposes.

Lemma C.9 (Regret upper bound)**.** *For any even $\kappa \geq k \geq 4$, $\|[\boldsymbol{r}_1^{(\overline{t_k})}]^+\|_\infty \leq 2\|[\boldsymbol{r}_1^{(\overline{t_{k-2}})}]^+\|_\infty + 2 \leq \frac{5}{3}2^{k/2}$ since $\|\boldsymbol{r}_1^{(\overline{t_2})}\|_\infty \leq \frac{4}{3}$. Similarly, for any odd $k \geq 5$, $\|[\boldsymbol{r}_2^{(\overline{t_k})}]^+\|_\infty \leq \max\{2\|[\boldsymbol{r}_2^{(\overline{t_{k-2}})}]^+\|_\infty, 2\} \leq \frac{5}{3}2^{(k-1)/2}$ since $\|\boldsymbol{r}_2^{(\overline{t_3})}\|_\infty \leq \frac{4}{3}$.*

Proof. The fact that $\|\boldsymbol{r}_1^{(\overline{t_2})}\|_\infty, \|\boldsymbol{r}_2^{(\overline{t_3})}\|_\infty \leq \frac{4}{3}$ can be shown as part of the basis of the induction. We make the argument for an even k. From round $\underline{t_k}$ until Player 1 plays $a_1(k)$ with probability 1, the regret of $a_1(k)$ increases by $k - (k\boldsymbol{x}_1^{(t)}[a_1(k)] + (k-1)\boldsymbol{x}_1^{(t)}[a_1(k-2)]) = \boldsymbol{x}_1^{(t)}[a_1(k-2)]$ and the regret of $a_1(k-2)$ increases by $k-1-(k\boldsymbol{x}_1^{(t)}[a_1(k)]+(k-1)\boldsymbol{x}_1^{(t)}[a_1(k-2)]) = -1+\boldsymbol{x}_1^{(t)}[a_1(k-2)]$; that is, it decreases by $1 - \boldsymbol{x}_1^{(t)}[a_1(k-2)]$. Let t' be the first round for which $\boldsymbol{r}^{(t')}[a_1(k)] \geq \boldsymbol{r}^{(t')}[a_1(k-2)]$. It holds that $\boldsymbol{r}^{(t')}[a_1(k)] \leq \|[\boldsymbol{r}_1^{(\overline{t_{k-2}})}]^+\|_\infty + 1$ since the regret of $a_1(k)$ is increasing by at most 1 in each round and $\boldsymbol{r}^{(t')}[a_1(k-2)] \leq \|[\boldsymbol{r}_1^{(\overline{t_{k-2}})}]^+\|_\infty$. From then onward, the regret of $a_1(k)$ is increasing by at most $1/2$ while the regret of $a_1(k-2)$ is decreasing by at least $1/2$. Thus, it will take at most $\lceil 2|\boldsymbol{r}^{(t')}[a_1(k-2)]|\rceil \leq 2|\boldsymbol{r}^{(t')}[a_1(k-2)]| + 1 \leq 2\|[\boldsymbol{r}_1^{(\overline{t_{k-2}})}]^+\|_\infty + 1$ rounds for the regret of $a_1(k-2)$ to be nonpositive. During that time, the regret of $a_1(k)$ can increase by at most $\|[\boldsymbol{r}_1^{(\overline{t_{k-2}})}]^+\|_\infty + 1$. □

Proof of Property 4.1. If κ is odd, it suffices to prove that in every round Player 1 mixes between $a_1(\kappa)$ and $a_1(\kappa+1)$, Player 2 plays $a_2(\kappa) = a_2(\kappa+1)$ with probability 1. Similarly, if κ is even it suffices to prove that in every round Player 2 mixes between $a_2(\kappa)$ and $a_2(\kappa+1)$, Player 1 plays $a_1(\kappa) = a_1(\kappa+1)$ with probability 1. Let us analyze the case where κ is even; the odd case is similar. When Player 2 starts mixing more and more to $a_2(\kappa+1)$, it makes the row $a_1(\kappa+2)$ more attractive for Player 2. By Lemmas C.7 and C.8,

$$\boldsymbol{r}_1^{(\overline{t_\kappa})}[a_1(\kappa+2)] \leq -\frac{\kappa-1}{2}T_\kappa - \frac{\kappa-2}{2}T_{\kappa-1} \leq -\frac{(\kappa-1)!}{2^{\kappa-2}}T_4 - \frac{(\kappa-2)!}{2^{\kappa-3}}T_3. \tag{13}$$

At the same time, Lemma C.9 implies that Player 2 is mixing between $a_2(\kappa)$ and $a_2(\kappa+1)$ for at most $3\|[\boldsymbol{r}_2^{(\overline{t_{\kappa-1}})}]^+\|_\infty + 2 \leq 5 \cdot 2^{(\kappa-1)/2} + 2$ rounds. To see this, we observe that it takes at most $\lceil\|[\boldsymbol{r}_2^{(\overline{t_{\kappa-1}})}]^+\|_\infty\rceil$ rounds for the action $a_2(\kappa+1)$ to be played with at least the same probability as $a_2(\kappa)$, which in turn holds because the regret of $a_2(\kappa+1)$ increases by $\boldsymbol{x}_2^{(t)}[a_2(\kappa)]$ while the regret of $a_2(\kappa)$ decreases by $1 - \boldsymbol{x}_2^{(t)}[a_2(\kappa)]$. From then on, the regret of $a_2(\kappa)$ decreases by at least $1/2$ in each round, so it takes at most $\lceil 2\|[\boldsymbol{r}_2^{(\overline{t_{\kappa-1}})}]^+\|_\infty\rceil$ rounds for it to be nonpositive. We now claim that, by (13), action $a_1(\kappa+2)$ is never played during those rounds. The reason is that since $T_3 \geq 5$ and $T_4 \geq 20$ (by our inductive basis),

$$\boldsymbol{r}_1^{(\overline{t_\kappa})}[a_1(\kappa+2)] \leq -20\frac{(\kappa-1)!}{2^{\kappa-2}} - 5\frac{(\kappa-2)!}{2^{\kappa-3}}.$$

and in each round the regret of $a_1(\kappa+2)$ can only decrease additively by 2. Since

$$\frac{1}{2}\left(20\frac{(\kappa-1)!}{2^{\kappa-2}}+5\frac{(\kappa-2)!}{2^{\kappa-3}}\right) > 5\cdot 2^{(\kappa-1)/2}+2 \quad \forall \kappa \geq 4,$$

the inductive step follows. □

The next lemma shows that, under the invariance of Property 4.1, the only way to reach an approximate Nash equilibrium is to start playing the actions corresponding to $2m-1$, which is the maximum payoff in the matrix.

Lemma C.10. *Consider any strategy profile $(\boldsymbol{x}_1, \boldsymbol{x}_2)$ such that Player 1 only assigns positive probability to actions in $\{a_1(k), a_1(k+1)\}$ and Player 2 only assigns positive probability to actions in $\{a_2(k), a_2(k+1)\}$, where $k+1 < 2m-1$. Then either Player 1 or Player 2 has a deviation benefit of at least $1/k+2 - \gamma$ for any $\gamma > 0$.*

Proof. By construction of the game, either $a_1(k) = a_1(k+1)$ or $a_2(k) = a_2(k+1)$. We can assume that $a_1(k) = a_1(k+1)$; the argument when $a_2(k) = a_2(k+1)$ is symmetric. Let p be the probability Player 2 places at $a_2(k+1)$ and $1-p$ at $a_2(k)$. Suppose that the deviation benefit of each player is at most ϵ. The utility of Player 2 under the current strategy profile is $k(1-p)+(k+1)p = k+p$, while deviating to $a_2(k+1)$ gives $k+1$. So, $p > 1-\epsilon$. Given that $k+1 < 2m-1$, Player 1 can deviate to $a_1(k+2)$ to obtain a utility of $p(k+2) \leq k+p+\epsilon$. Combining with the fact that $p \geq 1-\epsilon$, this implies $\epsilon \geq 1/k+2$. □