The Thirty-Ninth AAAI Conference on Artificial Intelligence (AAAI-25)

Rapid Learning in Constrained Minimax Games with Negative Momentum

Zijian Fang[1*], Zongkai Liu[1, 3*], Chao Yu[1, 2†], Chaohao Hu[1]

[1]School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, China
[2]Pengcheng Laboratory, Shenzhen, China
[3]Shanghai Innovation Institute, Shanghai, China
{fangzj, liuzk}@mail2.sysu.edu.cn, yuchao3@mail.sysu.edu.cn, huchh9@mail3.sysu.edu.cn

Abstract

In this paper, we delve into the utilization of the negative momentum technique in constrained minimax games. From an intuitive mechanical standpoint, we introduce a novel framework for momentum buffer updating, which extends the findings of negative momentum from the unconstrained setting to the constrained setting and provides a universal enhancement to the classic game-solver algorithms. Additionally, we provide theoretical guarantee of convergence for our momentum-augmented algorithms with entropy regularizer. We then extend these algorithms to their extensive-form counterparts. Experimental results on both Normal Form Games (NFGs) and Extensive Form Games (EFGs) demonstrate that our momentum techniques can significantly improve algorithm performance, surpassing both their original versions and the SOTA baselines by a large margin.

Code — https://github.com/kkkaiaiai/NM-Method

Introduction

In recent years, a broad spectrum of applications in machine learning and robust optimization have been cast as a minimax optimization problem in the form of $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} f(\boldsymbol{x}, \boldsymbol{y})$. Examples formulated under this framework include generative adversarial networks (GANs) (Goodfellow et al. 2020), adversarial training (2018), fair statistical inference (Madras et al. 2018), market equilibrium (Kroer et al. 2019), primal-dual reinforcement learning (Du et al. 2017) and numerous others. This optimization problem can be conceptualized as a zero-sum game involving two players: the first player minimizes $f(\boldsymbol{x}, \boldsymbol{y})$ by tuning $\boldsymbol{x}$, while the other player maximizes $f(\boldsymbol{x}, \boldsymbol{y})$ by tuning $\boldsymbol{y}$.

With the inextricably intertwined advancement of online learning and game theory, several algorithms have become foundational solvers for minimax games, including the Online Mirror Descent (OMD) (Warmuth, Jagota et al. 1997), Follow-The-Regularized-Leader (FTRL) (Abernethy, Hazan, and Rakhlin 2008), and particularly Regret Matching (RM) (Hart and Mas-Colell 2000), which is an algorithm more closely aligned with game theory and becomes the building block of solving imperfect-information games (Moravčík et al. 2017; Brown and Sandholm 2018). However, these algorithms are known to exhibit rotation behaviour and fail to converge pointwise even in simple bilinear cases (Vlatakis-Gkaragkounis, Flokas, and Piliouras 2019). A corpus of studies aim to address this divergence issue through plain modifications of standard algorithms, with a specific focus on achieving an enhanced convergence rate and/or securing last-iterate convergence guarantees (Golowich, Pattathil, and Daskalakis 2020).

Among these techniques, regularization and optimistic gradient are two widely used methods, with Magnet Mirror Descent (MMD) (Sokota et al. 2023) and Optimistic Gradient Descent Ascent (OGDA) (Mertikopoulos et al. 2019) as their respective representative algorithms. MMD provides a linear convergence rate to the regularized equilibrium by utilizing the influences of regularization on last-iterate convergence. OGDA, on the other hand, introduces an optimistic gradient estimate to guide the convergence process more effectively by predicting future gradients. In addition, Negative Momentum (NM) is introduced as an enhancement technique in recent research (Gidel et al. 2019), achieving a linear convergence rate comparable to Extragradient (Korpelevich 1976) and OGDA in unconstrained bilinear games. Subsequent works have further delved into the convergence properties of NM (Zhang and Wang 2021; Lorraine et al. 2022). However, recent analysis of NM predominantly focus on the impact of its integration with GDA and the learning dynamics over the unconstrained setting, leaving a gap in discussions concerning its interaction with game-solver algorithms and its performance in constrained settings. Hence, this paper places particular emphasis on the following two questions:

- Can NM be extended from the unconstrained setting to the constrained setting?
- Can NM provide a significant empirical improvement over existing methods like regularization and optimistic?

To provide affirmative answers to these questions, we make the following contributions in this paper:

- We introduce a negative momentum updating framework tailored for the constrained setting, coupled with an intuitive paradigm for updating the momentum buffer,

[*]These authors contributed equally.
[†]Corresponding author.

which can be seamlessly integrated with classic algorithms. Furthermore, by using the dilated distance generated function (Hoda et al. 2010) and regret decomposition framework (Farina, Kroer, and Sandholm 2019a), we propose momentum-augmented versions of their extensive-form counterparts.

- We theoretically prove that our momentum-augmented variant with negative entropy regularizer achieves an exponential convergence rate to an approximate equilibrium with an infinitely large buffer or converges to the set of Nash equilibria with a sufficiently large buffer.
- We conduct comprehensive experiments over randomly generated NFGs and four standard EFGs, including Kuhn Poker, Leduc Poker, Goofspiel and Liar's dice. The experimental results demonstrate that the momentum-augmented algorithms exhibit significant improvements over both their original versions and other existing strong last-iterate convergent baselines. It is noteworthy that our proposed algorithms $\text{MoRM}^{+}(\text{MoCFR}^{+})$ consistently obtain 10^{9} times lower exploitability than $\text{RM}^{+}(\text{CFR}^{+})$, and outperform another SOTA variant PCFR^{+}. To our knowledge, this marks the first instance where an algorithm surpasses CFR^{+} across various types of games.

Related Work

The related work is organized to encompass existing general techniques for facilitating minimax training and addressing the convergence problem over both the unconstrained and constrained setting.

Timescale separation. Timescale separation involves solving the inside maximization problem initially to get an approximation of $\boldsymbol{y}^{*}$ and compute the gradient of $\boldsymbol{x}$ as if $\boldsymbol{y}^{*}$ is fixed, serving as a potential good descent direction. In training GANs, Heusel et al. (2017) utilize a larger learning rate for the discriminator to ensure convergence to a local Nash Equilibrium (NE). Fiez and Ratliff (2021) explore more general non-convex non-concave zero-sum games and elucidate the local convergence to strict local minimax equilibrium with finite timescale separation. The two-timescale update rule resembles a softened "learning vs. best response" scheme (Daskalakis, Foster, and Golowich 2020), which has also been investigated in the literature of the constrained setting like game solving, guaranteeing the convergence to the NE (Lockhart et al. 2019). Nevertheless, the faster-updating player requires training a costly best response oracle at each iteration, and asymmetric updates typically lead to less desirable unilateral convergence.

Predictive updates. Predictive updates come from the intuitions that players could utilize heuristics to predict each other's next move (Foerster et al. 2017; Chavdarova et al. 2021). Algorithmically, these methods can be considered variations of the optimistic/extra-gradient methods, where the gradient dynamics are modified by incorporating approximate second-order information (Schäfer and Anandkumar 2019). Previous studies have investigated the last-iterate convergence in the unconstrained setting such as training GANs (Liang and Stokes 2019). In cases where a unique NE is assumed, Daskalakis and Panageas (2019) and Wei et al. (2021) have extended the scope of research for Optimistic Multiplicative Weight Update (OMWU) in NFGs. In the context of EFGs, Farina, Kroer, and Sandholm (2019b) empirically demonstrate the last-iterate convergence of OMWU, while Lee, Kroer, and Luo (2021) subsequently establish theoretical proofs with the uniqueness assumption of NE. Recent work by (Farina et al. 2023; Cai et al. 2023) has extended the analysis to include tighter ergodic convergence rate and last-iterate convergence. While the predictive updates algorithms are normally accompanied by theoretical properties, their practical implementation often necessitates the computation of multiple strategies at each iteration. Furthermore, these algorithms may not consistently yield a significant acceleration, especially in games with more intricate structures (Farina, Kroer, and Sandholm 2021) or with a larger scale (Lee, Kroer, and Luo 2021).

Regularization. Regularization has emerged as a pivotal tool for accelerating convergence. Pérolat et al. (2021) conduct a comprehensive analysis of the impact of entropy regularization on continuous-time dynamics, and propose a reward transformation method to achieve linear convergence in EFGs using counterfactual values. However, their theoretical findings cannot be inherently extended to desired discrete-time results. Abe et al. (2023) propose a variant of MWU by incorporating an additional regularization term serving as the mutation dynamic, while Liu et al. (2023) achieve improved convergence results by regularizing the payoff functions of the games. Magnet Mirror Descent (MMD) (Sokota et al. 2023) investigate the influences of general-case regularization on last-iterate convergence and provide a linear convergence rate to the regularized equilibrium. In a parallel study, Abe et al. (2024) yield comparable results but aim for an exact Nash equilibrium while imposing a more stringent constraint on the learning rate.

Other techniques. Other methods modify algorithms with ad-hoc adjustments to game dynamics. Consensus optimization (CO) (Mescheder, Nowozin, and Geiger 2017) and gradient penalty (Gulrajani et al. 2017) improve convergence by minimizing the players' gradient magnitude. Balduzzi et al. (2018) improve convergence by disentangling convergent potential components from rotation Hamiltonian components of vector field. However, these approaches require estimating coupled gradient and Hessian information, which is computationally expensive, prone to high variance, and less practical even in centralized settings.

Preliminaries

Problem Formulation

In this paper, we consider the problem of solving a constrained bilinear *saddle-point problem* (SPP):

$$\min_{\boldsymbol{x}\in\mathcal{X}}\max_{\boldsymbol{y}\in\mathcal{Y}} \boldsymbol{b}^{\top}\boldsymbol{x}+\boldsymbol{x}^{\top}\boldsymbol{G}\boldsymbol{y}+\boldsymbol{c}^{\top}\boldsymbol{y}, \tag{1}$$

where $G \in [-1,+1]^{M\times N}$ is a known loss function matrix, and $\mathcal{X} \subset \mathbb{R}^{M}$, $\mathcal{Y} \subset \mathbb{R}^{N}$ are the convex and compact decision sets (i.e. strategies) for min/max players. The linear terms are not essential in our analysis and thus we take $\boldsymbol{b} = \boldsymbol{c} = 0$ throughout the paper[1]. This problem formulation captures several game-theoretical applications such as finding the NE in norm-form/extensive-form zero-sum games when $\mathcal{X}$ and $\mathcal{Y}$ represent simplex Δ and treeplex Δ_T, respectively. We denote the simplex/treeplex of dimension $d-1$ as Δ^d/Δ_T^d. By the celebrated minimax theorem (v. Neumann 1928), we have $\min_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y} = \max_{\boldsymbol{y}\in\mathcal{Y}} \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y}$. The set of Nash equilibria is defined as $\mathcal{Z}^* = \mathcal{X}^* \times \mathcal{Y}^*$ where $\mathcal{X}^* = \operatorname{argmin}_{\boldsymbol{x}\in\mathcal{X}} \max_{\boldsymbol{y}\in\mathcal{Y}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y}$ and $\mathcal{Y}^* = \operatorname{argmax}_{\boldsymbol{y}\in\mathcal{Y}} \min_{\boldsymbol{x}\in\mathcal{X}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y}$, which is always convex for two-player zero-sum games. The duality gap (i.e. exploitability) of a pair of feasible strategies $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ is defined as:

$$\textit{DualityGap}(\boldsymbol{x}, \boldsymbol{y}) = \max_{\boldsymbol{y}'\in\mathcal{Y}} \boldsymbol{x}^\top \boldsymbol{G}\boldsymbol{y}' - \min_{\boldsymbol{x}'\in\mathcal{X}} \boldsymbol{x}'^\top \boldsymbol{G}\boldsymbol{y}. \quad (2)$$

Note that $\textit{DualityGap}(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ holds and $\textit{DualityGap}(\boldsymbol{x}, \boldsymbol{y}) \leq \epsilon$ implies that the strategy profile $\boldsymbol{z} \in \mathcal{Z}$ is an ϵ-Nash equilibrium of the bilinear game.

For notation convenience, we let $P = M + N$ and denote the loss vector of the bilinear form in Equation (1) as $F(\boldsymbol{z}_t) = (F(\boldsymbol{x}_t), F(\boldsymbol{y}_t)) = (\boldsymbol{G}\boldsymbol{y}_t, -\boldsymbol{G}^\top \boldsymbol{x}_t) = (\boldsymbol{f}_t, -\boldsymbol{g}_t)$ for any $\boldsymbol{z}_t = (\boldsymbol{x}_t, \boldsymbol{y}_t) \in \mathcal{Z} = \mathcal{X} \times \mathcal{Y} \subset \mathbb{R}^P$, where $\boldsymbol{f}_t$ and $\boldsymbol{g}_t$ represent the gradients of the current strategy profile. We assume $\|F(\boldsymbol{z})\|_\infty \leq 1$ for all $\boldsymbol{z} \in \mathcal{Z}$, which can be always satisfied by normalizing the entries of G.

One way to solve bilinear SPPs is by viewing SPP as a repeated game between two players: at iteration t, players choose $\boldsymbol{z}_t \in \mathcal{Z}$ and then observe their loss $l_t^{\mathcal{Z}}(\boldsymbol{z}_t) = \left(\boldsymbol{x}_t^\top \boldsymbol{G}\boldsymbol{y}_t, -\boldsymbol{x}_t^\top \boldsymbol{G}\boldsymbol{y}_t\right) = \langle \boldsymbol{z}_t, F(\boldsymbol{z}_t)\rangle$. The goal of each player is to minimize their regrets $R_{T,\boldsymbol{x}}$, $R_{T,\boldsymbol{y}}$ across T iterations:

$$\begin{aligned} R_{T,\boldsymbol{x}} &= \sum_{t=1}^{T} \langle \boldsymbol{f}_t, \boldsymbol{x}_t\rangle - \min_{\boldsymbol{x}\in\mathcal{X}} \sum_{t=1}^{T} \langle \boldsymbol{f}_t, \boldsymbol{x}\rangle, \\ R_{T,\boldsymbol{y}} &= \max_{\boldsymbol{y}\in\mathcal{Y}} \sum_{t=1}^{T} \langle \boldsymbol{g}_t, \boldsymbol{y}\rangle - \sum_{t=1}^{T} \langle \boldsymbol{g}_t, \boldsymbol{y}_t\rangle, \end{aligned} \quad (3)$$

which measure the difference between the loss accumulated by the sequence of $(\boldsymbol{z}_1, \ldots \boldsymbol{z}_T)$ and the loss that would have been accumulated by employing the best time-independent strategies $\boldsymbol{z}$ in hindsight. An algorithm is called *regret minimizer* if the regret grows sublinearly in T. We denote $R_{T,\hat{\boldsymbol{x}}}$ as the regret of arbitrary $\hat{\boldsymbol{x}} \in \mathcal{X}$.

It is well known that if $\boldsymbol{z}_t$ follows the trajectory of a *regret minimizer* learning algorithm, $\frac{1}{t}\sum_{\tau\leq t} \boldsymbol{x}_\tau^\top \boldsymbol{G}\boldsymbol{y}_\tau$ converges to the optimal value of the bilinear SPP (1) as $t \to \infty$, and the average strategies $\frac{1}{t}\sum_{\tau\leq t} \boldsymbol{z}_\tau$ converges to the optimal solution to the SPP (i.e. NE) if the solution is unique.

Online Learning and Regret Matching

Online Linear Optimization Oracles. The online optimization oracles all follow a reminiscent procedure that within the operation loop of observing the loss vector $F(\boldsymbol{z}_t)$ and updating the next strategies $\boldsymbol{z}_{t+1}$. For solving bilinear SPPs over constrained sets, Online Mirror Descent (OMD) (Warmuth, Jagota et al. 1997) and Follow-The-Regularized-Leader (FTRL) (Abernethy, Hazan, and Rakhlin 2008) stand out as two most classical online linear optimization algorithms. With arbitrary $\boldsymbol{z}_0 \in \mathcal{Z}$, the OMD algorithm proceeds iterations following the rule:

$$\boldsymbol{z}_{t+1} = \operatorname*{argmin}_{\boldsymbol{z}\in\mathcal{Z}} \left\{\eta \langle \boldsymbol{z}, F(\boldsymbol{z}_t)\rangle + D_\psi(\boldsymbol{z}, \boldsymbol{z}_t)\right\}, \quad (4)$$

where ψ is a convex function called *regularizer*, $\psi(\boldsymbol{z}) = \psi(\boldsymbol{x}) + \psi(\boldsymbol{y})$ and $D_\psi(p, q) = \psi(p) - \psi(q) - \langle \nabla\psi(q), p - q\rangle$ is the *Bregman divergence*. The FTRL produces iterations following the rule:

$$\boldsymbol{z}_{t+1} = \operatorname*{argmin}_{\boldsymbol{z}\in\mathcal{Z}} \left\{\sum_{k=1}^{t} \eta \langle \boldsymbol{z}, F(\boldsymbol{z}_k)\rangle + \psi(\boldsymbol{z})\right\}. \quad (5)$$

FTRL can be equivalent to OMD with linearized loss (Orabona 2023). OMD/FTRL instantiate Gradient Descent Ascent (GDA) and Multiplicative Weights Update (MWU) when regularizer $\psi(\boldsymbol{z})$ is specified as the negative entropy $\psi(\boldsymbol{z}) = \boldsymbol{z}\log\boldsymbol{z}$ or the L_2-norm $\psi(\boldsymbol{z}) = \frac{1}{2}\|\boldsymbol{z}\|^2$, respectively.

Regret Matching. Regret Matching (RM) (Hart and Mas-Colell 2000) stands as one of the preeminent *regret minimizer* learning algorithms, extensively utilized in the domain of game-solving applications. An instantaneous regret vector is defined as $\boldsymbol{r}(\boldsymbol{x}_t) = \langle \boldsymbol{x}_t, F(\boldsymbol{x}_t)\rangle \cdot \mathbf{1}_L - F(\boldsymbol{x}_t)$,[2] which measures the change in regret incurred at iteration t relative to each dimension of the decision sets. RM keeps an accumulative regret $\boldsymbol{R}_t^x$ and updates strategy $\boldsymbol{x}_{t+1}$ by normalizing the thresholded accumulative regret:

$$\boldsymbol{R}_{t+1}^x = \sum_{\tau=0}^{t} \boldsymbol{r}(\boldsymbol{x}_\tau), \quad \boldsymbol{x}_{t+1} = \left[\boldsymbol{R}_{t+1}^x\right]^+ / \| \left[\boldsymbol{R}_{t+1}^x\right]^+ \|_1, \quad (6)$$

where $[\cdot]^+$ denotes thresholding at zero. Regret Matching$^+$ (RM$^+$) (Tammelin 2014) is a variant of RM that further thresholds the accumulative regret at zero at every iteration: $\boldsymbol{R}_{t+1}^x = [\boldsymbol{R}_t^x + \boldsymbol{r}(\boldsymbol{x}_t)]^+$. Burch, Moravcik, and Schmid (2019) show that combining RM$^+$ with the alternation trick, wherein the strategies of two players are updated asynchronously, yields faster empirical performance and is proven to exhibit strict improvement for game solving (Grand-Clément and Kroer 2024). In contrast to FTRL/OMD, the update rules of RM stand out for being parameter-free and exclusively involving closed-form operations, specifically, thresholding and normalizing.

Methods

The Negative Momentum Mechanism

We recall that in the unconstrained case, the iterations of the Gradient Descent Ascent with augmented (Polyak) momen-

[1] If they are not zero, one can translate $\boldsymbol{x}$ and $\boldsymbol{y}$ to cancel the linear terms by linear transform, see e.g. (Gidel et al. 2019).

[2] L can be either M or N, that is we overload the notation r so its domain depends on the input.

tum (Polyak 1964) term (GDAm) proceed as follows:

$$\boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \eta F(\boldsymbol{z}_t) + \beta(\boldsymbol{z}_t - \boldsymbol{z}_{t-1}), \tag{7}$$

where η is a positive step size. Alternatively, with $\boldsymbol{\mu}_0 = 0$, the procedure can be written in an equivalent form with momentum buffer $\boldsymbol{\mu}_t = (\boldsymbol{z}_t - \boldsymbol{z}_{t-1})/\eta$:

$$\boldsymbol{\mu}_t = \beta\boldsymbol{\mu}_{t-1} - F(\boldsymbol{z}_t), \quad \boldsymbol{z}_t = \boldsymbol{z}_{t-1} + \eta\boldsymbol{\mu}_t. \tag{8}$$

The convergence process of GDAm can be illustrated by the Newton's 2^{nd} law $m\ddot{X} = \boldsymbol{F}$ of a particle of mass m. Without loss of generality, we henceforth assume the mass of our object to be unity. By viewing the $\boldsymbol{F}_{\text{curl}} = -F(\boldsymbol{z})$ as the curl force of the 2-dimensional $X - Y$ surface (Berry and Shukla 2016), the discretization of the continuous-time dynamic $\ddot{Z} = -F(Z)$ with the discretization step size $\delta = \sqrt{\eta}$ can be written as follows:

$$\boldsymbol{z}_{t+1} = 2\boldsymbol{z}_t - \boldsymbol{z}_{t-1} - \eta F(\boldsymbol{z}_t), \tag{9}$$

which is corresponding to setting $\beta = 1$ in Equation (7).

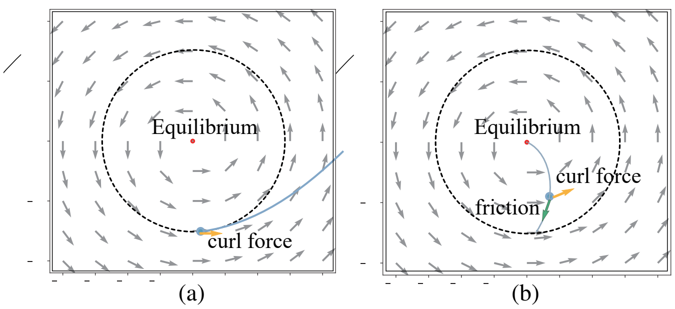

Figure 1: The mechanic dynamic of a particle with different force. In (a), the particle situated within a curl force diverges from the equilibrium. In (b), the augmented friction results in a reduction of the particle's velocity, thereby dampening oscillations and facilitating eventual convergence.

Nevertheless, as shown in Figure 1(a), the mode of GDAm with non-negative parameters is divergent in the min-max objective since the curl force increases the particle's speed over time and thus prevents convergence. Therefore, to decrease the velocity of particle for convergence, a straightforward approach is augmenting the dynamic with an extra linear friction $\boldsymbol{F}_{\text{fric}} = -\mu\dot{Z}$ as follows:

$$\ddot{Z} = -F(Z) - \mu\dot{Z}, \tag{10}$$

where μ can be considered as the coefficient of friction. Figure 1(b) illustrates that the friction in Equation (10) effectively dampens oscillation and leads to particle's convergence. Furthermore, through a combination of implicit and explicit update steps as Shi et al. (2019), Proposition 1 illustrates that the discretization of Equation (10) corresponds to the GDAm with negative momentum.

Proposition 1. *With the discretization step-size $\delta = \sqrt{\eta}$ and a sufficiently large $\mu > \frac{1}{\delta}$, Equation (10) can be discretized in the form of GDAm with negative momentum:*

$$\boldsymbol{z}_{t+1} = \boldsymbol{z}_t - \eta F(\boldsymbol{z}_t) + \beta(\boldsymbol{z}_t - \boldsymbol{z}_{t-1}), \tag{11}$$

where $\beta = 1 - \mu\delta < 0$.

Algorithm 1: Restarting Aggregated Momentum (RAM)

Input: $\mathcal{L}$, μ, β, Restarting interval k
$\mathcal{L} \leftarrow \emptyset, \boldsymbol{\mu}_0 = 0, \boldsymbol{z}_0 \in \mathcal{Z}$
for $t = 0$ **to** T **do**
  $\boldsymbol{\mu}_t = \beta \sum_{\boldsymbol{\mu} \in \mathcal{L}} \boldsymbol{\mu} - F(\boldsymbol{z}_t)$
  $\boldsymbol{z}_{t+1} = \mathit{Oracle}(\boldsymbol{z}_t, \boldsymbol{\mu}_t)$ following (12) or (13)
  if $L = k$ **then**
    Reset the snapshot profile $\mathcal{L} \rightarrow \emptyset$
  end if
  Save current momentum $\{\boldsymbol{\mu}_t\} \cup \mathcal{L} \rightarrow \mathcal{L}$
end for
Output: $\boldsymbol{z}_{t+1}$

Generalization to The Constrained Setting

Motivated by the effect of negative momentum terms, a key objective of our work is to extend these findings to the constrained setting. Inspired by Eq. 8, we introduce a momentum buffer defined by $\boldsymbol{\mu}_t = \beta\boldsymbol{\mu}_{t-1} - F(\boldsymbol{z}_t)$ with $\boldsymbol{\mu}_0 = 0$. This allows us to replace $F(\boldsymbol{z}_t)$ in Eq. 4 with the negative momentum term $\boldsymbol{\mu}_t$ and propose the Mirror Descent with Momentum (MoMD) updating rules:

$$\boldsymbol{z}_{t+1} = \underset{\boldsymbol{z} \in \mathcal{Z}}{\operatorname{argmin}} \left\{ \eta \left\langle \boldsymbol{z}, -\boldsymbol{\mu}_t \right\rangle + D_\psi\left(\boldsymbol{z}, \boldsymbol{z}_t\right) \right\}. \tag{12}$$

Following the equivalence between OMD and FTRL, the updating rules of the FTRL with Momentum (MoFTRL) can be written as:

$$\begin{aligned} \boldsymbol{L}_t = \boldsymbol{L}_{t-1} - \boldsymbol{\mu}_t = \boldsymbol{L}_{t-1} + F(\boldsymbol{z}_t) - \beta(\boldsymbol{L}_{t-2} - \boldsymbol{L}_{t-1}), \\ \boldsymbol{z}_{t+1} = \underset{\boldsymbol{z} \in \mathcal{Z}}{\operatorname{argmin}} \left\{ \eta \left\langle \boldsymbol{z}, \boldsymbol{L}_t \right\rangle + \psi(\boldsymbol{z}) \right\}. \end{aligned} \tag{13}$$

It is noteworthy that the negative momentum term is kept in the dual space rather than the prime space. Taking an alternative perspective, negative momentum involves leveraging the buffer to store historical gradients and alternates between adding and subtracting the gradient at each iteration.

To intensify the focus on (i.e., allocate greater weights to) gradients from recent iterations, inspired by prior research on momentum buffer variants (Lorraine et al. 2022) , we additionally propose to cache a profile $\mathcal{L}$ saving buffer snapshots within a fixed time duration k and aggregate them for the final updates. That is, employing a degree of notation flexibility, we transform the formula for updating the momentum from $\boldsymbol{u}_t = \beta\boldsymbol{u}_{t-1} - F(\boldsymbol{z}_t)$ to $\boldsymbol{u}_t = \beta\sum_{\boldsymbol{u} \in \mathcal{L}} \boldsymbol{u} - F(\boldsymbol{z}_t)$ in Equation (12) or (13), and the profile is reset when the length L of $\mathcal{L}$ reaches specified k. This updating paradigm, denoted as Restarting Aggregated Momentum (RAM), is illustrated in Algorithm 1.

Taking MoFTRL as an example, if we substitute $\boldsymbol{\mu}_t = \boldsymbol{L}_{t-1} - \boldsymbol{L}_t$ into the aggregation of snapshots profile and cancel out the interleave terms, the updating rules of the cumulative past loss in Equation (13) become:

$$\begin{aligned} \boldsymbol{L}_t = &\boldsymbol{L}_{t-1} + F(\boldsymbol{z}_t) - \beta(\boldsymbol{L}_{att} - \boldsymbol{L}_{t-1}), \\ &\boldsymbol{L}_{att} \leftarrow \boldsymbol{L}_{t-1} \quad \text{if } t \,\%\, k = 0. \end{aligned} \tag{14}$$

Intuitively, the momentum updating paradigm within Algorithm 1 resembles mounting a spring between a periodically

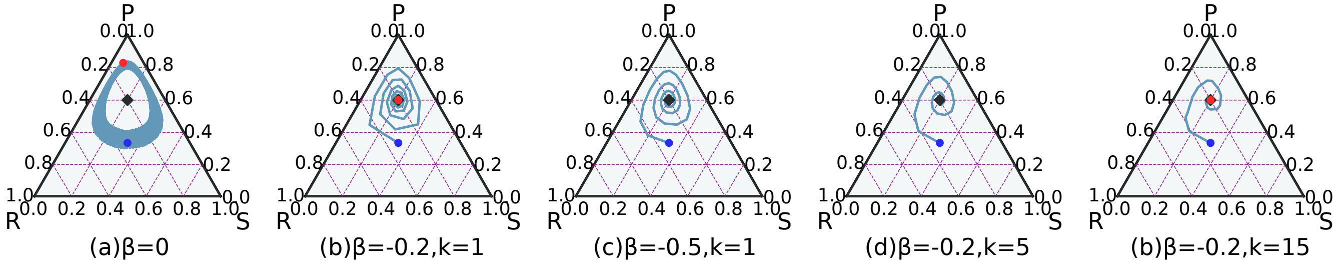

Figure 2: The trajectory plots depict the MoMWU algorithm under varying negative momentum coefficients β and intervals k in Bias RPS. The initial strategy is set to $(\boldsymbol{x}, \boldsymbol{y}) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. The equilibrium strategy is denoted by a black point, whereas the blue and red points symbolize the starting and ending points of the trajectory. The learning rate η is fixed at 2 in this context.

updating attachment point L_{att} and the current point L_t, transforming the dynamic of friction into an effect reminiscent of restoring force of the spring. Algorithm 1 reinstates the conventional momentum update rules when $k = 1$.

In order to illustrate the effects of RAM on the learning dynamics of algorithms, Figure 2 demonstrates the trajectory of MoMD with negative entropy variant (MoMWU) in the context of Bias RPS, where the game matrix is defined as $G = [[0, -1, 3], [1, 0, -1], [-3, 1, 0]]$. When contrasted with the original MWU (equivalent to setting $\beta = 0$ in MoMWU) depicted in Figure 2(a), the incorporation of negative momentum contributes to the damping of oscillations. Furthermore, higher values of β exhibit improved effectiveness in mitigating oscillations within a reasonable range as shown in Figures 2(b) and 2(c). Extending the buffer length introduces a distinct periodic updating behavior in the trajectory. The additional attachment point adds a supplementary "restoring force", amplifying resistance in each iteration compared to the standard momentum updating approach with $k = 1$. This heightened resistance significantly dampens oscillations, as observed in Figures 2(d) and 2(e). The ablation experiments in Appendix C further discuss the effect of momentum related to k on the algorithms.

The following theoretical analysis supports the experimental results above. Assume that $\boldsymbol{z}_*$ is the Nash equilibrium (NE) of the following modified game:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \max_{\boldsymbol{y} \in \mathcal{Y}} \boldsymbol{x}^\top \boldsymbol{G} \boldsymbol{y} - \frac{\beta}{\eta} D_\psi(\boldsymbol{x}, \boldsymbol{x}_{att}) + \frac{\beta}{\eta} D_\psi(\boldsymbol{y}, \boldsymbol{y}_{att}), \quad (15)$$

where $(\boldsymbol{x}_{att}, \boldsymbol{y}_{att}) = \underset{\boldsymbol{z} \in \mathcal{Z}}{\operatorname{argmin}} \{\eta \langle \boldsymbol{z}, \boldsymbol{L}_{att} \rangle + \psi(\boldsymbol{z})\}$. Theorem 2 establishes that Algorithm 1, with a constant learning rate and a profile $\mathcal{L}$ that stores all past snapshots $\boldsymbol{\mu}_t$, converges at an exponentially fast rate to the modified equilibrium $\boldsymbol{z}_*$. Theorem 3 indicates that the modified equilibrium serves as an $\mathcal{O}(\frac{-\beta}{\eta})$-NE of the original game. This implies that a lower $-\beta$ reduces the duality gap of the modified equilibrium, while a higher $-\beta$ accelerates the convergence rate, introducing a trade-off between convergence speed and the bias of NE attributed to β. Moreover, with an updating attachment point $\boldsymbol{L}_{att}$, Theorem 4 proves that Algorithm 1 converges to NE with a sufficiently large k. Complete proofs of our theoretical results can be found in Appendix D.

Theorem 2. *Let $k = \infty$, i.e., $\boldsymbol{L}_{att} = 0$. Then, $\boldsymbol{z}_t$ derived by Algorithm 1 satisfies:*

$$D_\psi(\boldsymbol{z}_*, \boldsymbol{z}_t) \leq D_\psi(\boldsymbol{z}_*, \boldsymbol{z}_0) \cdot (1 + \frac{\beta}{2})^t,$$

if $\psi(p) = \langle p, \ln p \rangle$, $-\frac{2}{3} < \beta < 0$ and $0 < \eta \leq \frac{\sqrt{-(1+\frac{3}{2}\beta)\beta}}{2}$.

Theorem 3. *In the same setup of Theorem 2, the duality gap for the updated strategy $\boldsymbol{z}_t$ of Algorithm 1 can be bounded as:*

$$\begin{aligned} & DualityGap(\boldsymbol{z}_t) \\ \leq & \frac{-\beta}{\eta} \cdot diam(\boldsymbol{Z}) \cdot \|\log \frac{\boldsymbol{z}_*}{\boldsymbol{z}_{att}}\| + \mathcal{O}((1 + \frac{\beta}{2})^{\frac{t}{2}}), \end{aligned}$$

where $diam(\boldsymbol{Z}) = \sup_{\boldsymbol{z}, \boldsymbol{z}' \in \boldsymbol{Z}} \|\boldsymbol{z} - \boldsymbol{z}'\|$, $\boldsymbol{z}_{att}$ equals $(\boldsymbol{x}_{att}, \boldsymbol{y}_{att}) = \underset{\boldsymbol{z} \in \mathcal{Z}}{\operatorname{argmin}} \{\eta \langle \boldsymbol{z}, \boldsymbol{L}_{att} \rangle + \psi(\boldsymbol{z})\}$.

Theorem 4. *Algorithm 1 with $\psi(p) = \langle p, \ln p \rangle$ converges to the set of Nash equilibria.*

We next further extend the concept of negative momentum to RM/RM$^+$. According to recent work on the interesting connection between FTRL/OMD and RM/RM$^+$ (Farina, Kroer, and Sandholm 2021), the regret update $\boldsymbol{R}_{t+1}^x = [\boldsymbol{R}_t^x + \boldsymbol{r}(\boldsymbol{x}_t)]^+$ of RM$^+$ can be reformulated as:

$$\boldsymbol{R}_{t+1} = \underset{\hat{\boldsymbol{R}} \in \mathbb{R}_+^M}{\operatorname{argmin}} \left\{ \eta \langle \hat{\boldsymbol{R}}, \boldsymbol{r}(\boldsymbol{x}_t) \rangle + D_\psi(\hat{\boldsymbol{R}}, \boldsymbol{R}_t) \right\}, \quad (16)$$

where $\psi = \frac{1}{2}\| \cdot \|^2$ and $\eta = -1$. Therefore, RM$^+$ is intricately linked to OMD instantiated with the non-negative orthant (after thresholding) as the decision set and facing a sequence of loss $(\boldsymbol{r}(\boldsymbol{x}_1))_{t \geq 1}$. Formally, Lemma 5 draws the relation for the regret in the strategy sequence $\boldsymbol{x}_1, ...\boldsymbol{x}_T$ and the regret $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$.

Lemma 5. *(Farina et al. 2023) Let $\boldsymbol{x}_1, ...\boldsymbol{x}_T \in \Delta^M$ be generated as $\boldsymbol{x}_t = [\boldsymbol{R}_t^x]^+ / \| [\boldsymbol{R}_t^x]^+ \|_1$ for some sequence $R_1^x, ...R_T^x \in \mathbb{R}_+^M$. The regret $R_{T,\hat{\boldsymbol{x}}}$ of $\boldsymbol{x}_1, ...\boldsymbol{x}_T$ facing a sequence of loss $F(\boldsymbol{x}_1), ...F(\boldsymbol{x}_T)$ is equal to $R_{T,\hat{\boldsymbol{R}}}$, i.e., the regret of $\boldsymbol{R}_1, ..., \boldsymbol{R}_T$ facing the sequence of loss $\boldsymbol{r}(\boldsymbol{x}_1), ...\boldsymbol{r}(\boldsymbol{x}_T)$, compared against $\hat{\boldsymbol{R}} = \hat{\boldsymbol{x}}$: $R_{T,\hat{\boldsymbol{R}}} = \sum_{t=1}^T \langle r(\boldsymbol{x}_t), \boldsymbol{R}_t - \hat{\boldsymbol{R}} \rangle$.*

Algorithm 2: Momentum RM$^+$ (MoRM$^+$)

$(\boldsymbol{R}_t^x, \boldsymbol{R}_t^y) = 0, (\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathcal{Z}$
for $t = 0$ **to** T **do**
  $\boldsymbol{R}_{t+1}^x = [\boldsymbol{R}_t^x + \boldsymbol{r}(x_t) - \beta(\boldsymbol{R}_{att}^x - \boldsymbol{R}_t^x)]^+$
  $\boldsymbol{x}_{t+1} = [\boldsymbol{R}_{t+1}^x]^+ / \| [\boldsymbol{R}_{t+1}^x]^+ \|_1$
  $\boldsymbol{R}_{t+1}^y = [\boldsymbol{R}_t^y + \boldsymbol{r}(y_t) - \beta(\boldsymbol{R}_{att}^y - \boldsymbol{R}_t^y)]^+$
  $\boldsymbol{y}_{t+1} = [\boldsymbol{R}_{t+1}^y]^+ / \| [\boldsymbol{R}_{t+1}^y]^+ \|_1$
  if $t \,\%\, k = 0$ **then**
    Update the attachment regret vector $\boldsymbol{R}_{att}^x \leftarrow \boldsymbol{R}_{t-1}^x, \boldsymbol{R}_{att}^y \leftarrow \boldsymbol{R}_{t-1}^y$
  end if
end for

In light of this correspondence, we integrate the updating rules within RAM into the regret updating framework of RM$^+$, leading to the introduction of variant referred to as **Momentum RM$^+$** (Algorithm 2).

Implementation for Extensive-Form Game (EFG). Here we briefly elucidate the process of extending our algorithm to accommodate EFG settings. Generally, there are two approaches to render the resolution of EFG problems computationally feasible. One avenue of research employs the regret decomposition framework (Farina, Kroer, and Sandholm 2019a) adopted by the Counterfactual Regret Minimization (CFR) (Zinkevich et al. 2007) family. This framework is founded on the concept that the global regret of the entire game can be decomposed into the summation of local regrets associated with each simplex corresponding to a decision node in EFG. We employ MoRM$^+$ as the local regret minimizer, leading to the algorithm named **MoCFR$^+$**.

Another line of research formulates EFG as a bilinear SPP over the *sequence-form* strategy, where $\mathcal{X}$ and $\mathcal{Y}$ represent sequence-form polytopes, which equivalently can be viewed as treeplexes (Hoda et al. 2010). When employing classical first-order methods such as OMD/FTRL, the class of *dilated distance generating function* (Farina, Kroer, and Sandholm 2019b) becomes crucial for efficient computation and convexity guarantees. We create the DMoMD algorithm by applying the dilated mapping to MoMD. Details on EFGs and our extensive-form algorithms are in Appendix A.

Experiments

In this section, we validate our methods utilizing the exploitability metric under two experimental settings of NFGs and EFGs. Detailed game description and hyper-parameters can be found in the Appendix B.

Normal-Form Games

For tabular normal-form games, we conduct experiments on randomly generated NFGs with action dimensions of 25, 50, and 75. The payoff matrix is drawn from a standard Gaussian distribution with i.i.d realizations using different seed ranging from 0 to 10, and constrains its elements to lie within the range $[-1, 1]$ after normalization. We also verify our algorithms on a 3×3 game matrix $G = [[3, 0, -3], [0, 3, -4], [0, 0, 1]]$ that has the a unique Nash equilibrium $(\boldsymbol{x}^*, \boldsymbol{y}^*) = \left(\left[\frac{1}{12}, \frac{1}{12}, \frac{5}{6}\right], \left[\frac{1}{3}, \frac{5}{12}, \frac{1}{4}\right]\right)$, which has also been used in Farina et al. (2023) to illustrate the slow ergodic convergence of RM$^+$.

We compare MoMWU and MoRM$^+$ to average-iterate convergent algorithms (RM, RM$^+$) and last-iterate convergent algorithms (OMWU, OGDA). We also include the comparison to the Magnet Mirror Descent (MMD) (Sokota et al. 2023). MMD can be adapted into an NE solver by either annealing the amount of regularization over time (MMD-A) or by employing a moving magnet strategy that trails behind the current iteration (MMD-M).

Note that for RM$^+$ and MoRM$^+$, we employ alternating updates, while other algorithms are kept with their default settings, i.e., simultaneous updates. We plot the average strategy for RM, linear average strategy for RM$^+$ and current strategy for other algorithms. We set the uniform strategy as the initial magnet strategy for MMD-M.

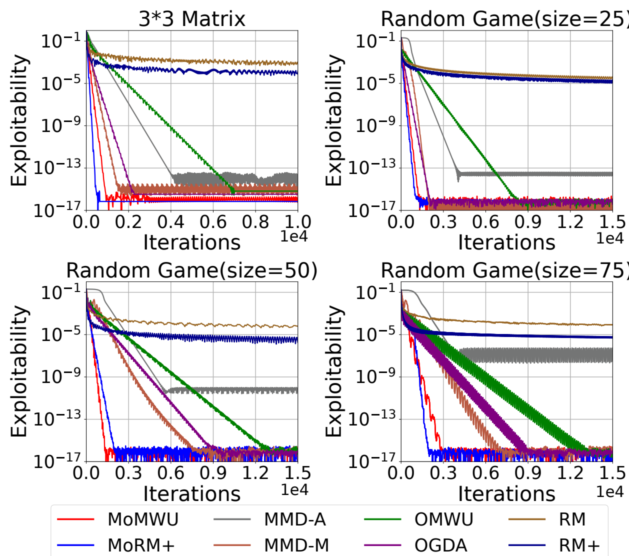

Figure 3: The performance evaluation of the momentum variants and other baseline algorithms in NFGs. In all plots, the x-axis represents the number of iterations for each algorithm, while the y-axis, presented on a logarithmic scale, illustrates the exploitability.

The results of NFGs are illustrated in Figure 3. Specifically, in the specific 3×3 matrix game, MoRM$^+$ effectively overcomes the slow ergodic convergence of RM$^+$ as observed in (Cai et al. 2023), showcasing a notable improvement in convergence speed. This observation underscores the beneficial impact of the momentum technique on enhancing convergence properties. The results of the random matrix game experiments indicate that our momentum variants (MoRM$^+$, MoMWU) consistently demonstrate the fastest convergence rates across all games when compared to their original versions and other rapidly convergent algorithms. Moreover, our algorithms exhibit stability and robustness as the game size increases, whereas optimistic methods (OMWU, OGDA) display a slight oscillation and slower convergence. This observation is further substantiated in subsequent experiments on EFGs.

Extensive-Form Games

In the tabular setting of EFGs, we utilize games implemented in OpenSpiel (Lanctot et al. 2020), encompassing Kuhn Poker, Goofspiel (4/5 cards), Liar's dice (4/5 sides), and Leduc Poker. We evaluate DMoMD with L2-norm (DMoGDA) and MoCFR$^+$ against average-iterate convergent algorithms (CFR, CFR$^+$). Additionally, predictive update algorithms (OGDA, PCFR$^+$ (Farina, Kroer, and Sandholm 2021)) along with the regularization algorithms (MMD-M, Reg-CFR, and Reg-DOGDA (Liu et al. 2023)) are included in the comparison. Note that for algorithms incorporating RM$^+$ updating rules (CFR$^+$, PCFR$^+$, and MoCFR$^+$), we employ alternating updates, while other algorithms use simultaneous updates. We present the average strategy for CFR, linear average strategy for CFR$^+$ and the quadratic average strategy for PCFR$^+$, as recommended in their default setting. Other algorithms are evaluated based on their current strategy. The uniform strategy is set as the initial magnet strategy for MMD-M.

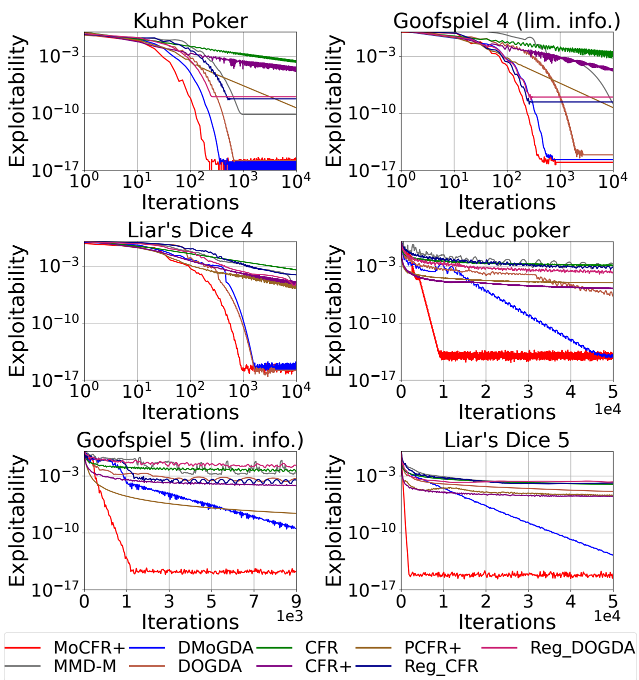

Figure 4: Evaluating the momentum variants and baselines in EFGs. Results are arranged by game sizes. The first three games use a logarithmic x-axis for clearer presentation.

Figure 4 illustrates the results in EFGs. Our momentum-augmented algorithms significantly enhance the performance across all games, demonstrating superior empirical convergence rates and substantially lower final exploitability. Remarkably, MoCFR$^+$ consistently attains significantly lower exploitability compared to CFR$^+$ and PCFR$^+$. To the best of our knowledge, this marks the first instance where an algorithm surpasses CFR$^+$ performance across various types of games. Additionally, DMoGDA also demonstrates rapid convergence, outpacing all the baselines by a considerable margin. This represents the first occurrence where GDA-type algorithms demonstrate remarkable performance, outperforming SOTA variants of CFR in larger games.

Similar to the observations in NFGs, predictive updates suffer substantial performance degradation with the increasing game size. In contrast, our algorithms can maintain robustness and consistently demonstrate favorable convergence results. An intuitive conjecture posits that, in contract to the friction-like dynamic induced by the negative momentum, the incorporation of predictive updates introduces supplementary second-order information, which can be likened to the application of centripetal force in the learning dynamics, serving to effectively alleviate oscillations (Peng et al. 2020). However, as the game expands, the escalating non-transitivity poses a significant challenge (Czarnecki et al. 2020).Methods that utilizes the variations in two consecutive gradient, such as optimistic/extra-gradient, may inadequately furnish the requisite energy to proficiently guide the gradient and disentangle the circular behavior in the learning dynamics. In contrast, our momentum updating framework utilizes a history of past gradients within a period, providing increased energy to mitigate oscillations. This potentially elucidates the robust performance of momentum-augmented algorithms in larger games.

Conclusion

In this paper, we extend the negative momentum updating paradigm from the unconstrained setting to the constrained setting, seamlessly integrating it with classical algorithms. We formulate the momentum-augmented versions of FTRL, OMD, and RM+. Leveraging regret decomposition and the dilated distance generate function, we introduce MoCFR+ and DMoMD for solving EFGs. Experiments conducted across numerous benchmark games, demonstrate that momentum-augmented algorithms significantly outperform SOTA algorithms on all tested games.

This work paves the way for some potential research directions. For example, enhancing performance through the combination of predictive and momentum-augmented approaches for improvements is a valuable consideration (Huang and Zhang 2022). Additionally, the integration of other momentum updating paradigms from the literature, such as complex momentum (Lorraine et al. 2022), holds promise for achieving faster convergence rates or convergence guarantees over more general settings.

Acknowledgments

We gratefully acknowledge the support from the National Natural Science Foundation of China (No. 62076259, 62402252), the Fundamental and Applicational Research Funds of Guangdong Province (No. 2023A1515012946), the Fundamental Research Funds for the Central Universities Sun Yat-sen University, and the Pengcheng Laboratory Project (PCL2023A08, PCL2024Y02).

References

Abe, K.; Ariu, K.; Sakamoto, M.; and Iwasaki, A. 2024. Adaptively Perturbed Mirror Descent for Learning in Games. arXiv:2305.16610.

Abe, K.; Ariu, K.; Sakamoto, M.; Toyoshima, K.; and Iwasaki, A. 2023. Last-Iterate Convergence with Full and Noisy Feedback in Two-Player Zero-Sum Games. arXiv:2208.09855.

Abernethy, J. D.; Hazan, E.; and Rakhlin, A. 2008. Competing in the Dark: An Efficient Algorithm for Bandit Linear Optimization. In *Annual Conference on Learning Theory*, 263–274. Omnipress.

Balduzzi, D.; Racaniere, S.; Martens, J.; Foerster, J.; Tuyls, K.; and Graepel, T. 2018. The mechanics of n-player differentiable games. In *International Conference on Machine Learning*, 354–363. PMLR.

Berry, M.; and Shukla, P. 2016. Curl force dynamics: symmetries, chaos and constants of motion. *New Journal of Physics*, 18(6): 063018.

Brown, N.; and Sandholm, T. 2018. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals. *Science*, 359(6374): 418–424.

Burch, N.; Moravcik, M.; and Schmid, M. 2019. Revisiting CFR+ and alternating updates. *Journal of Artificial Intelligence Research*, 64: 429–443.

Cai, Y.; Farina, G.; Grand-Clément, J.; Kroer, C.; Lee, C.-W.; Luo, H.; and Zheng, W. 2023. Last-Iterate Convergence Properties of Regret-Matching Algorithms in Games. arXiv:2311.00676.

Chavdarova, T.; Pagliardini, M.; Stich, S. U.; Fleuret, F.; and Jaggi, M. 2021. Taming GANs with Lookahead-Minmax. arXiv:2006.14567.

Czarnecki, W. M.; Gidel, G.; Tracey, B.; Tuyls, K.; Omidshafiei, S.; Balduzzi, D.; and Jaderberg, M. 2020. Real world games look like spinning tops. *Advances in Neural Information Processing Systems*, 33: 17443–17454.

Daskalakis, C.; Foster, D. J.; and Golowich, N. 2020. Independent Policy Gradient Methods for Competitive Reinforcement Learning. In *Advances in Neural Information Processing Systems*.

Daskalakis, C.; and Panageas, I. 2019. Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization. In *Innovations in Theoretical Computer Science Conference*, volume 124 of *LIPIcs*, 27:1–27:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Du, S. S.; Chen, J.; Li, L.; Xiao, L.; and Zhou, D. 2017. Stochastic variance reduction methods for policy evaluation. In *International Conference on Machine Learning*, 1049–1058. PMLR.

Farina, G.; Grand-Clément, J.; Kroer, C.; Lee, C.-W.; and Luo, H. 2023. Regret Matching+: (In)Stability and Fast Convergence in Games. arXiv:2305.14709.

Farina, G.; Kroer, C.; and Sandholm, T. 2019a. Online convex optimization for sequential decision processes and extensive-form games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 1917–1925.

Farina, G.; Kroer, C.; and Sandholm, T. 2019b. Optimistic Regret Minimization for Extensive-Form Games via Dilated Distance-Generating Functions. In *Advances in Neural Information Processing Systems*, 5222–5232.

Farina, G.; Kroer, C.; and Sandholm, T. 2021. Faster game solving via predictive blackwell approachability: Connecting regret matching and mirror descent. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 5363–5371.

Fiez, T.; and Ratliff, L. J. 2021. Local convergence analysis of gradient descent ascent with finite timescale separation. In *International Conference on Learning Representation*.

Foerster, J. N.; Chen, R. Y.; Al-Shedivat, M.; Whiteson, S.; Abbeel, P.; and Mordatch, I. 2017. Learning with Opponent-Learning Awareness. In *Adaptive Agents and Multi-Agent Systems*.

Gidel, G.; Hemmat, R. A.; Pezeshki, M.; Le Priol, R.; Huang, G.; Lacoste-Julien, S.; and Mitliagkas, I. 2019. Negative momentum for improved game dynamics. In *The 22nd International Conference on Artificial Intelligence and Statistics*, 1802–1811. PMLR.

Golowich, N.; Pattathil, S.; and Daskalakis, C. 2020. Tight last-iterate convergence rates for no-regret learning in multiplayer games. *Advances in Neural Information Processing Systems*, 33: 20766–20778.

Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A. C.; and Bengio, Y. 2020. Generative adversarial networks. *Communications of the ACM*, 63(11): 139–144.

Grand-Clément, J.; and Kroer, C. 2024. Solving optimization problems with Blackwell approachability. *Mathematics of Operations Research*, 49(2): 697–728.

Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; and Courville, A. C. 2017. Improved Training of Wasserstein GANs. In *Advances in Neural Information Processing Systems*, 5767–5777.

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

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. In *Advances in Neural Information Processing Systems*, 6626–6637.

Hoda, S.; Gilpin, A.; Pena, J.; and Sandholm, T. 2010. Smoothing techniques for computing Nash equilibria of sequential games. *Mathematics of Operations Research*, 35(2): 494–512.

Huang, K.; and Zhang, S. 2022. New first-order algorithms for stochastic variational inequalities. *SIAM Journal on Optimization*, 32(4): 2745–2772.

Korpelevich, G. M. 1976. The extragradient method for finding saddle points and other problems. *Matecon*, 12: 747–756.

Kroer, C.; Peysakhovich, A.; Sodomka, E.; and Stier-Moses, N. E. 2019. Computing large market equilibria using abstractions. In *ACM Conference on Economics and Computation*, 745–746.

Lanctot, M.; Lockhart, E.; Lespiau, J.-B.; Zambaldi, V.; Upadhyay, S.; Pérolat, J.; Srinivasan, S.; Timbers, F.; Tuyls, K.; Omidshafiei, S.; Hennes, D.; Morrill, D.; Muller, P.; Ewalds, T.; Faulkner, R.; Kramár, J.; Vylder, B. D.; Saeta, B.; Bradbury, J.; Ding, D.; Borgeaud, S.; Lai, M.; Schrittwieser, J.; Anthony, T.; Hughes, E.; Danihelka, I.; and Ryan-Davis, J. 2020. OpenSpiel: A Framework for Reinforcement Learning in Games. arXiv:1908.09453.

Lee, C.; Kroer, C.; and Luo, H. 2021. Last-iterate Convergence in Extensive-Form Games. In *Advances in Neural Information Processing Systems*, 14293–14305.

Liang, T.; and Stokes, J. 2019. Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks. In *International Conference on Artificial Intelligence and Statistics*, 907–915. PMLR.

Liu, M.; Ozdaglar, A.; Yu, T.; and Zhang, K. 2023. The Power of Regularization in Solving Extensive-Form Games. arXiv:2206.09495.

Lockhart, E.; Lanctot, M.; Pérolat, J.; Lespiau, J.; Morrill, D.; Timbers, F.; and Tuyls, K. 2019. Computing Approximate Equilibria in Sequential Adversarial Games by Exploitability Descent. In *International Joint Conference on Artificial Intelligence*, 464–470.

Lorraine, J. P.; Acuna, D.; Vicol, P.; and Duvenaud, D. 2022. Complex momentum for optimization in games. In *International Conference on Artificial Intelligence and Statistics*, 7742–7765. PMLR.

Madras, D.; Creager, E.; Pitassi, T.; and Zemel, R. S. 2018. Learning Adversarially Fair and Transferable Representations. In *International Conference on Machine Learning*, 3381–3390. PMLR.

Mertikopoulos, P.; Lecouat, B.; Zenati, H.; Foo, C.; Chandrasekhar, V.; and Piliouras, G. 2019. Optimistic mirror descent in saddle-point problems: Going the extra (gradient) mile. In *International Conference on Learning Representations*.

Mescheder, L. M.; Nowozin, S.; and Geiger, A. 2017. The Numerics of GANs. In *Advances in Neural Information Processing Systems*, 1825–1835.

Moravčík, M.; Schmid, M.; Burch, N.; Lisỳ, V.; Morrill, D.; Bard, N.; Davis, T.; Waugh, K.; Johanson, M.; and Bowling, M. 2017. Deepstack: Expert-level artificial intelligence in heads-up no-limit poker. *Science*, 356(6337): 508–513.

Orabona, F. 2023. A Modern Introduction to Online Learning. arXiv:1912.13213.

Peng, W.; Dai, Y.-H.; Zhang, H.; and Cheng, L. 2020. Training GANs with centripetal acceleration. *Optimization Methods and Software*, 35(5): 955–973.

Pérolat, J.; Munos, R.; Lespiau, J.-B.; Omidshafiei, S.; Rowland, M.; Ortega, P.; Burch, N.; Anthony, T.; Balduzzi, D.; De Vylder, B.; et al. 2021. From poincaré recurrence to convergence in imperfect information games: Finding equilibrium via regularization. In *International Conference on Machine Learning*, 8525–8535. PMLR.

Polyak, B. T. 1964. Some methods of speeding up the convergence of iteration methods. *Ussr computational mathematics and mathematical physics*, 4(5): 1–17.

Schäfer, F.; and Anandkumar, A. 2019. Competitive Gradient Descent. In *Advances in Neural Information Processing Systems*, 7623–7633.

Shi, B.; Du, S. S.; Su, W. J.; and Jordan, M. I. 2019. Acceleration via Symplectic Discretization of High-Resolution Differential Equations. In *Advances in Neural Information Processing Systems*.

Sinha, A.; Namkoong, H.; and Duchi, J. C. 2018. Certifying Some Distributional Robustness with Principled Adversarial Training. In *International Conference on Learning Representations*.

Sokota, S.; D'Orazio, R.; Kolter, J. Z.; Loizou, N.; Lanctot, M.; Mitliagkas, I.; Brown, N.; and Kroer, C. 2023. A Unified Approach to Reinforcement Learning, Quantal Response Equilibria, and Two-Player Zero-Sum Games. In *International Conference on Learning Representations,*.

Tammelin, O. 2014. Solving Large Imperfect Information Games Using CFR+. arXiv:1407.5042.

v. Neumann, J. 1928. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1): 295–320.

Vlatakis-Gkaragkounis, E.; Flokas, L.; and Piliouras, G. 2019. Poincaré Recurrence, Cycles and Spurious Equilibria in Gradient-Descent-Ascent for Non-Convex Non-Concave Zero-Sum Games. In *Advances in Neural Information Processing Systems*, 10450–10461.

Warmuth, M. K.; Jagota, A. K.; et al. 1997. Continuous and discrete-time nonlinear gradient descent: Relative loss bounds and convergence. In *International Symposium on Artificial Intelligence and Mathematics*, volume 326. Citeseer.

Wei, C.; Lee, C.; Zhang, M.; and Luo, H. 2021. Linear Last-iterate Convergence in Constrained Saddle-point Optimization. In *International Conference on Learning Representations*.

Zhang, G.; and Wang, Y. 2021. On the suboptimality of negative momentum for minimax optimization. In *International Conference on Artificial Intelligence and Statistics*, 2098–2106. PMLR.

Zinkevich, M.; Johanson, M.; Bowling, M.; and Piccione, C. 2007. Regret minimization in games with incomplete information. *Advances in Neural Information Processing Systems*, 20.