Last-Iterate Convergence: Zero-Sum Games and Constrained Min-Max Optimization

Constantinos Daskalakis
EECS, CSAIL
MIT
USA
costis@csail.mit.edu

Ioannis Panageas
Information Systems, Technology and Design
SUTD
Singapore
ioannis@sutd.edu.sg

Abstract

Motivated by applications in Game Theory, Optimization, and Generative Adversarial Networks, recent work of Daskalakis et al [Daskalakis et al., ICLR, 2018] and follow-up work of Liang and Stokes [Liang and Stokes, 2018] have established that a variant of the widely used Gradient Descent/Ascent procedure, called "Optimistic Gradient Descent/Ascent (OGDA)", exhibits last-iterate convergence to saddle points in *unconstrained* convex-concave min-max optimization problems. We show that the same holds true in the more general problem of *constrained* min-max optimization under a variant of the no-regret Multiplicative-Weights-Update method called "Optimistic Multiplicative-Weights Update (OMWU)". This answers an open question of Syrgkanis et al [Syrgkanis et al., NIPS, 2015].

The proof of our result requires fundamentally different techniques from those that exist in no-regret learning literature and the aforementioned papers. We show that OMWU monotonically improves the Kullback-Leibler divergence of the current iterate to the (appropriately normalized) min-max solution until it enters a neighborhood of the solution. Inside that neighborhood we show that OMWU becomes a contracting map converging to the exact solution. We believe that our techniques will be useful in the analysis of the last iterate of other learning algorithms.[1]

1 Introduction

A central problem in Game Theory and Optimization is computing a pair of probability vectors $(\mathbf{x}, \mathbf{y})$, solving

$$\min_{\mathbf{y}\in\Delta_m} \max_{\mathbf{x}\in\Delta_n} \mathbf{x}^\top A\mathbf{y}, \tag{1}$$

where $\Delta_n \subset \mathbb{R}^n$ and $\Delta_m \subset \mathbb{R}^m$ are probability simplices, and A is a $n \times m$ matrix. Von Neumann's celebrated minimax theorem informs us that

$$\min_{\mathbf{y}\in\Delta_m} \max_{\mathbf{x}\in\Delta_n} \mathbf{x}^\top A\mathbf{y} = \max_{\mathbf{x}\in\Delta_n} \min_{\mathbf{y}\in\Delta_m} \mathbf{x}^\top A\mathbf{y}, \tag{2}$$

and that all solutions to the LHS are solutions to the RHS, and vice versa. This result was a founding stone in the development of Game Theory. Indeed, interpreting $\mathbf{x}^\top A\mathbf{y}$ as the payment of the "min player" to the "max player" when the former selects a distribution $\mathbf{y}$ over columns and the latter selects a distribution $\mathbf{x}$ over rows of matrix A, a solution to (1) constitutes an equilibrium of the

[1]The work has already been published in Innovations for Theoretical Computer Science (ITCS) 2019. The full version of this paper can be found in `https://arxiv.org/abs/1807.04252`

Abstract for: Smooth Games Optimization and Machine Learning Workshop (NeurIPS 2019), Vancouver, Canada.

game defined by matrix A, called a "minimax equilibrium", a pair of randomized strategies such that neither player can improve their payoff by unilaterally changing their distribution.

Besides their fundamental value for Game Theory, it is known that (1) and (2) are also intimately related to Linear Programming. It was shown by von Neumann that (2) follows from strong linear programming duality. Moreover, it was suggested by Dantzig [7] and recently proven by Adler [1] that any linear program can be solved by solving some min-max problem of the form (1). In particular, min-max problems of form (1) are exactly as expressive as min-max problems of the following form, which capture any linear program (by Lagrangifying the constraints):

$$\min_{\mathbf{y}\geq 0}\max_{\mathbf{x}\geq 0}\left(\mathbf{x}^\top A\mathbf{y}+b^\top\mathbf{x}+c^\top\mathbf{y}\right). \tag{3}$$

Soon after the minimax theorem was proven and its connection to linear programming was forged, researchers proposed dynamics for solving min-max optimization problems by having the min and max players of (1) run a simple learning procedure in tandem. An early method, proposed by Brown [4] and analyzed by Robinson [15], was fictitious play. Soon after, Blackwell's approachability theorem [3] propelled the field of online learning, which lead to the discovery of several learning algorithms converging to minimax equilibrium at faster rates, while also being robust to adversarial environments, situations where one of the players of the game deviates from the prescribed dynamics; see e.g. [5]. These learning methods, called "no-regret", include the celebrated multiplicative-weights-update method, follow-the-regularized-leader, and follow-the-perturbed-leader. Compared to centralized linear programming procedures the advantage of these methods is the simplicity of executing their steps, and their robustness to adversarial environments, as we just discussed.

Last vs Average Iterate Convergence. Despite the extensive literature on no-regret learning, an unsatisfactory feature of known results is that min-max equilibrium is shown to be attained only in an average sense. To be precise, if $(\mathbf{x}^t,\mathbf{y}^t)$ is the trajectory of a no-regret learning method, it is usually shown that the average $\frac{1}{t}\sum_{\tau\leq t}{\mathbf{x}^\tau}^\top A\mathbf{y}^\tau$ converges to the optimal value of (1), as $t\to\infty$. Moreover, if the solution to (1) is unique, then $\frac{1}{t}\sum_{\tau\leq t}(\mathbf{x}^\tau,\mathbf{y}^\tau)$ converges to the optimal solution. Unfortunately that does not mean that the last iterate $(\mathbf{x}^t,\mathbf{y}^t)$ converges to an optimal solution, and indeed it commonly diverges or enters a limit cycle. Furthermore, in the optimization literature, Nesterov [12] provides a method that can give pointwise convergence (i.e., convergence of the last iterate) to problem (1)[2], however his algorithm is not a no-regret learning algorithm. Recent work by Daskalakis et al [8] and Liang and Stokes [10] studies whether last iterate convergence can be established for no-regret learning methods in the simple unconstrained min-max problem of the form:

$$\min_{\mathbf{y}\in\mathbb{R}^m}\max_{\mathbf{x}\in\mathbb{R}^n}\left(\mathbf{x}^\top A\mathbf{y}+b^\top\mathbf{x}+c^\top\mathbf{y}\right). \tag{4}$$

For this problem, it is known that Gradient Descent/Ascent (GDA) is a no-regret learning procedure, corresponding to follow-the-regularized leader (FTRL) with ℓ_2^2-regularization. As such, the average trajectory traveled by GDA converges to a min-max solution, in the afore-described sense. On the other hand, it is also known that GDA may diverge from the min-max solution, even in trivial cases such as $A=\mathrm{I}, n=m=1, b=c=0$. Interestingly, [8, 10] show that a variant of GDA, called "Optimistic Gradient Descent/Ascent (OGDA),[3] exhibits last iterate convergence. Inspired by their theoretical result for the performance of OGDA in (4), Daskalakis et al. [8] even propose the use of OGDA for training Generative Adversarial Networks (GANs) [9]. Moreover, Syrgkanis et al. [16] provide numerical experiments which indicate that the trajectories of Optimistic Hedge (variant of Hedge in the same way OGDA is a variant of GDA) stabilize (i.e., converge pointwise) as opposed to (classic) Hedge and they *posed the question whether Optimistic Hedge actually converges pointwise.*

Motivated by the afore-described lines of work, and the importance of last iterate convergence for Game Theory and the modern applications of GDA-style methods in Optimization, our goal in this work is to generalize the results of [8, 10] to the general min-max problem (3), or equivalently (1); indeed, we will focus on the latter, but our algorithms are readily applicable to the former as the

[2]Nesterov showed that by optimizing $f_\mu(\mathbf{x}) := \mu\ln(\frac{1}{m}\sum_{j=1}^m e^{-\frac{1}{\mu}(A\mathbf{x})_j})$, $g_\nu(\mathbf{x}) := \nu\ln(\frac{1}{n}\sum_{j=1}^n e^{\frac{1}{\nu}(A^\top\mathbf{y})_j})$ for $\mu=\Theta(\frac{\epsilon}{\log m})$, $\nu=\Theta(\frac{\epsilon}{\log n})$ yields an $O(\epsilon)$ approximation to the problem problem (1).

[3]OGDA is tantamount to Optimistic FTRL with ℓ_2^2-regularization, in the same way that GDA is tantamount to FTRL with ℓ_2^2-regularization; see e.g. [14]. OGDA essentially boils down to GDA with negative momentum.

two problems are equivalent [1]. With the constraint that $(\mathbf{x}, \mathbf{y})$ should remain in $\Delta_n \times \Delta_m$, GDA and OGDA are not applicable. Indeed, the natural GDA-style method for min-max problems in this case is the celebrated Multiplicative-Weights-Update (MWU) method, which is tantamount to FTRL with entropy-regularization. Unsurprisingly, in the same way that GDA suffers in the unconstrained problem (4), MWU exhibits cycling in the constrained problem (1) (a recent work is [2] and was also shown empirically in [16]). So it is natural for us to study instead its optimistic variant, "Optimistic Multiplicative-Weights-Update (OMWU)," (called Optimistic Hedge in [16]) which corresponds to Optimistic FTRL with entropy-regularization, the equations of which are given in Section 2.2. Our main result is the following (restated as Theorem 2.2 after Section 2.2) and answers an open question asked in [16] as applicable to two player zero sum games:

Theorem 1.1 (**Last-Iterate Convergence of OMWU**, informal)**.** *Whenever* (1) *has a unique optimal solution* $(\mathbf{x}^*, \mathbf{y}^*)$, *OMWU with a small enough learning rate and initialized at the pair of uniform distributions* $(\frac{1}{n}\mathbf{1}, \frac{1}{m}\mathbf{1})$ *exhibits last-iterate convergence to the optimal solution. That is, if* $(\mathbf{x}^t, \mathbf{y}^t)$ *are the vectors maintained by OMWU at step t, then* $\lim_{t\to\infty}(\mathbf{x}^t, \mathbf{y}^t) = (\mathbf{x}^*, \mathbf{y}^*)$.

Remark 1.2. *We note that the assumption about uniqueness of the optimal solution for problem* (1) *is generic in the following sense: Within the set of all zero-sum games, the set of zero-sum games with non-unique equilibrium has Lebesgue measure zero [2, 6]. This implies that if A's entries are sampled independently from some continuous distribution, then with probability one the min-max problem* (1) *will have a unique solution.*

Our paper provides *two important messages*:

- It strengthens the intuition that optimism helps the trajectories of learning dynamics stabilize (e.g., Optimistic MWU vs MWU or Optimistic GDA vs GDA; as the papers of Syrgkanis et al [16] and Daskalakis et al [8] also do).
- The techniques we use (typically appear in dynamical systems literature) to prove convergence for the last iterate, are fundamentally different from those commonly used to prove convergence of the time average of a learning algorithm.

2 Preliminaries

2.1 Definitions and facts

Dynamical Systems. A recurrence relation of the form $\mathbf{x}^{t+1} = w(\mathbf{x}^t)$ is a discrete time dynamical system, with update rule $w : \mathcal{S} \to \mathcal{S}$ where $\mathcal{S} = \Delta_n \times \Delta_m \times \Delta_n \times \Delta_m$ for our purposes. The point $\mathbf{z}$ is called a *fixed point* or *equilibrium* of w if $w(\mathbf{z}) = \mathbf{z}$. We will be interested in the following well known fact that will be used in our proofs.

2.2 OMWU Method

Our main contribution is that the last iterate of OMWU converges to the optimal solution. The OMWU dynamics is defined as follows ($t \geq 1$):

$$
\begin{aligned}
x_i^{t+1} &= x_i^t \frac{e^{2\eta(A\mathbf{y}^t)_i - \eta(A\mathbf{y}^{t-1})_i}}{\sum_{j=1}^n x_j^t e^{2\eta(A\mathbf{y}^t)_j - \eta(A\mathbf{y}^{t-1})_j}} \text{ for all } i \in [n], \\
y_i^{t+1} &= y_i^t \frac{e^{-2\eta(A^\top\mathbf{x}^t)_i + \eta(A^\top\mathbf{x}^{t-1})_i}}{\sum_{j=1}^m y_j^t e^{-2\eta(A^\top\mathbf{x}^t)_j + \eta(A^\top\mathbf{x}^{t-1})_j}} \text{ for all } i \in [m].
\end{aligned} \tag{5}
$$

Points $(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^0, \mathbf{y}^0)$ are the initial conditions and are given as input. We call $0 < \eta < 1$ the *stepsize* of the dynamics. It is more convenient to interpret OMWU dynamics as mapping a quadruple to quadruple $((\mathbf{x}^t, \mathbf{y}^t, \mathbf{x}^{t-1}, \mathbf{y}^{t-1}) \to (\mathbf{x}^{t+1}, \mathbf{y}^{t+1}, \mathbf{x}^t, \mathbf{y}^t))$.

Remark 2.1. *Let* $(\mathbf{x}^*, \mathbf{y}^*)$ *be the optimal solution. We see that* $(\mathbf{x}^*, \mathbf{y}^*, \mathbf{x}^*, \mathbf{y}^*)$ *is a fixed point of the mapping. Furthermore,* $\Delta_n \times \Delta_m \times \Delta_n \times \Delta_m$ *is invariant under OMWU dynamics. For* $t \geq 1$*, if* $x_i^t = 0$ *then* x_i *remains zero for all times greater than t, and if it is positive, it remains positive (both numerator and denominator are positive)* [4]*. In words, at all times the OMWU satisfies the non-negativity constraints and the renormalization factor (denominator) makes both* $\mathbf{x}, \mathbf{y}$*'s coordinates*

[4] Same holds for vector $\mathbf{y}$.

sum up to one. A last observation is that every fixed point of OMWU dynamics (mapping a quadruple to quadruple) has the form $(\mathbf{x}, \mathbf{y}, \mathbf{x}, \mathbf{y})$ *(two same copies). One important task is to express OMWU dynamics as a dynamical system.*

Statement of our result. We finish by stating formally the main result.

Theorem 2.2 (OMWU converges)**.** *Let* A *be a* $n \times m$ *matrix and assume that*

$$\min_{\mathbf{y} \in \Delta_m} \max_{\mathbf{x} \in \Delta_n} \mathbf{x}^\top A \mathbf{y}$$

has a unique solution $(\mathbf{x}^*, \mathbf{y}^*)$*. It holds that for* η *sufficiently small (depends on* n, m, A*), starting from the uniform distribution, i.e.,* $(\mathbf{x}^1, \mathbf{y}^1) = (\mathbf{x}^0, \mathbf{y}^0) = (\frac{1}{n}\mathbf{1}, \frac{1}{m}\mathbf{1})$*, it holds*

$$\lim_{t \to \infty} (\mathbf{x}^t, \mathbf{y}^t) = (\mathbf{x}^*, \mathbf{y}^*),$$

under OMWU dynamics. The stepsize η *is constant, i.e., does not scale with time*[5].

We need to note that it is not clear from our theorem how small η is and its dependence on the size of A. Nevertheless, our convergence result holds for constant stepsizes as opposed to the classic no-regret learning literature where η scales like $\frac{1}{\sqrt{T}}$ after T iterations. Another result we know of this flavor is about MWU algorithm on congestion games [13].

Remarks about the contribution. We provide some remarks as asked from the organizers.

- OMWU is a no-regret online algorithm, so at every iteration, each player has information about the costs of the past before the update. Extra gradient methods [11] do not follow exactly at this framework because it asks information about costs twice in every update.
- This abstract does not include any proofs. The full version including the proofs can be found in `https://arxiv.org/abs/1807.04252`.
- We can prove rates of convergence, in particular we can show that after T iterations, the method reaches $O(1/T^{1/9})$ ℓ_1 from Nash equilibrium. The problem is that there might be exponential dependence on the size of the payoff matrix. The average regret for OMWU scales at $1/T$ and this is known from Rakhlin and Sridharan [14].

3 Acknowledgements

Costis Daskalakis acknowledges support of NSF Awards IIS-1741137, CCF-1617730 and CCF-1901292, a Simons Investigator Award, the DOE PhILMs project (No. DE-AC05-76RL01830), the DARPA award HR00111990021, a Google Faculty award, and the MIT Frank Quick Faculty Research and Innovation. Ioannis Panageas acknowledges SRG ISTD 2018 136 and NRF fellowship for AI.

[5]Our proof also works if the starting points $(\mathbf{x}^1, \mathbf{y}^1), (\mathbf{x}^0, \mathbf{y}^0)$ are both in the interior of $\Delta_n \times \Delta_m$ and not necessarily uniform, however the choice of η depends on the initial distributions as well and not only on n, m, A.

References

[1] Ilan Adler. The equivalence of linear programs and zero-sum games. In *International Journal of Game Theory*, pages 165–177, 2013.

[2] James P. Bailey and Georgios Piliouras. Multiplicative weights update in zero-sum games. In *Proceedings of the 2018 ACM Conference on Economics and Computation, Ithaca, NY, USA, June 18-22, 2018*, pages 321–338, 2018.

[3] David Blackwell. An analog of the minimax theorem for vector payoffs. In *Pacific J. Math.*, pages 1–8, 1956.

[4] G.W Brown. Iterative solutions of games by fictitious play. In *Activity Analysis of Production and Allocation*, 1951.

[5] Nikolo Cesa-Bianchi and Gabor Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[6] Eric Van Damme. *Stability and perfection of Nash equilibria*. Springer, 1991.

[7] George B Dantzig. A proof of the equivalence of the programming problem and the game problem. *Activity analysis of production and allocation*, (13):330–338, 1951.

[8] Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with Optimism. In *Proceedings of ICLR*, 2018.

[9] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[10] Tengyuan Liang and James Stokes. Interaction matters: A note on non-asymptotic local convergence of generative adversarial networks. *arXiv preprint:1802.06132*, 2018.

[11] Panayotis Mertikopoulos, Houssam Zenati, Bruno Lecouat, Chuan-Sheng Foo, Vijay Chandrasekhar, and Georgios Piliouras. Mirror descent in saddle-point problems: Going the extra (gradient) mile. *CoRR*, abs/1807.02629, 2018.

[12] Yurii Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1):127–152, 2005.

[13] Gerasimos Palaiopanos, Ioannis Panageas, and Georgios Piliouras. Multiplicative weights update with constant step-size in congestion games: Convergence, limit cycles and chaos. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5874–5884, 2017.

[14] Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *COLT 2013 - The 26th Annual Conference on Learning Theory, June 12-14, 2013, Princeton University, NJ, USA*, pages 993–1019, 2013.

[15] J. Robinson. An iterative method of solving a game. In *Annals of Mathematics*, pages 296–301, 1951.

[16] Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, and Robert E. Schapire. Fast convergence of regularized learning in games. In *Annual Conference on Neural Information Processing Systems 2015*, pages 2989–2997, 2015.