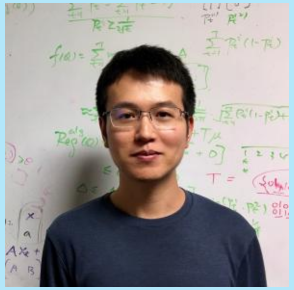
Kernelized Multiplicative Weights for 0/1-Polyhedral Games

Bridging the Gap Between
Learning in Extensive-Form and Normal-Form Games

Gabriele Farina
CMU

Chung-Wei Lee
USC



Haipeng Luo
USC

Christian Kroer
Columbia

No-regret learning in the context of normal-form games (NFGs) has been studied extensively

			
	0	-1	+1
	+1	0	-1
	-1	+1	0

Landmark result in theory of learning in games:

When all players learn using no-regret dynamics (e.g., MWU), the empirical frequency of play converges to the set of coarse correlated equilibria

Even more, in two-player zero-sum games, the average strategies converge to the set of Nash equilibria

As of today, learning is *by far* the most scalable way of computing game-theoretic solutions and equilibria in large games

1. *Linear time strategy updates*
2. *Each agent learns in parallel*
3. *Can often be implemented in a decentralized way*

Over the past decade, faster and faster no-regret dynamics have been developed for normal-form games

★ Most studied algorithm as of today: ***Optimistic Multiplicative Weights Update (OMWU)***

- Per-player regret bound:
  - ☑ Polylog dependence on the number of actions
  - ☑ Polylog(T) dependence on time

Implies $\tilde{O}\left(\frac{1}{T}\right)$ convergence to coarse correlated equilibrium in self-play

[Daskalakis et al. '21]

- Sum of players' regrets
  - ☑ Polylog dependence on #actions
  - ☑ Constant dependence on time

Implies $O\left(\frac{1}{T}\right)$ convergence to Nash eq. in two-player zero-sum games

[Syrgkanis et al. '15]

- ☑ Last-strategy convergence* (2pl 0sum)

[Hsieh et al. '21; Wei et al. '21]

However, normal-form games are a *rather limited* model of strategic interaction

All players act *once* and *simultaneously*

No sequential actions

No observations about other players' actions

Extensive-Form Games (EFGs)

Each player faces a tree-form decision problem

EFGs capture both sequential and simultaneous moves, as well as imperfect information and stochastic moves

Very expressive model of interaction
Examples of EFGs: chess, poker, bridge, security games, …

Online learning results for EFGs are harder to come by, due to their more intricate strategy sets

Normal-Form Games	***Extensive-Form Games***

• Per-player regret bound:
  • ☑ Polylog dependence on the number of actions
  • ☑ Polylog(T) dependence on time
	✗ Not known

• Sum of players' regrets
  • ☑ Polylog dependence on #actions
  • ☑ Constant dependence on time	

• ☑ Last-strategy convergence*	▨ Less is known

For many years, the EFG community has been "chasing" the NFG community, extending NFG breakthroughs to EFGs, when possible

For example, all these were all developed later for EFGs than NFGs (and sometimes only with weaker guarantees):

- Good distance measures [Hoda et al. '10; Kroer et al. '15; Farina et al. '21]
- Efficient optimistic algorithms [Farina et al. '19]
- Last-iterate convergence [Wei et al. '21, Lee et al. '21]

In fact, this paper was born from our desire to extend the polylog(T) regret bounds by [Daskalakis et al. '21] to EFGs.

For many years, the EFG community has been "chasing" the NFG community, extending NFG breakthroughs to EFGs, when possible

For example, all these were all developed later for EFGs than NFGs (and sometimes only with weaker guarantees):

- Good distance measures [Hoda et al. '10; Kroer et al. '15; Farina et al. '21]
- Efficient optimistic algorithms [Farina et al. '19]
- Last-iterate convergence [Wei et al. '21, Lee et al. '21]

Does it have to be like that? Or can we somehow bridge the gap and inherit the best properties of NFG algorithms also in EFGs?

Can we somehow bridge the gap?

Folklore result: any EFG can be converted into an equivalent NFG where each player's action set is the set of all deterministic policies in their tree-form decision problem. So, if we applied OMWU to that….

Catch: the number of such policies is exponential in each player's tree size

Common wisdom: because of the exponential blowup, the above approach is a *computational dead end*

⚡ Consequence: specialized techniques were developed for EFGs, and progress on EFGs and NFGs follows separate tracks for decades

The common wisdom is wrong

This paper: It is possible to simulate OMWU on the normal-form equivalent of an EFGs, in ***linear time per iteration*** in the tree size, via a ***kernel trick***

We call our algorithm **Kernelized OMWU (KOMWU)**

In fact, kernelized OMWU applies to any polyhedral domain with 0/1-coordinate vertices $\Omega \subseteq \mathbb{R}^d$

Main theorem: OMWU on the set of vertices of Ω can be simulated using $d + 1$ evaluations of the kernel at each iteration

So, if each kernel evaluation can be performed in poly(d) time, OMWU can be simulated in poly(d) time

KOMWU **closes part of the gap** between learning in NFGs and EFGs

- It achieves all the strong properties of OMWU there were so far only known to be achievable efficiently in NFGs (including polylog regret)
- ...as well as any future regret bounds that might get proven for OMWU!

As an unexpected byproduct, KOMWU obtains new state-of-the-art regret bounds among all online learning algorithms for extensive-form problems

Kernelized Multiplicative Weights for 0/1-Polyhedral Games

Algorithm		Per-player regret bound	Last-iter. conv.[†]
CFR (regret matching / regret matching$^+$)	(Zinkevich et al., 2007)	$\mathcal{O}(\sqrt{A}\,\|Q\|_1\,T^{1/2})$	no
CFR (MWU)	(Zinkevich et al., 2007)	$\mathcal{O}(\sqrt{\log A}\,\|Q\|_1\,T^{1/2})$	no
FTRL / OMD (dilated entropy)	(Kroer et al., 2020)	$\mathcal{O}(\sqrt{\log A}\,2^{D/2}\,\|Q\|_1\,T^{1/2})$	no
FTRL / OMD (dilatable global entropy)	(Farina et al., 2021a)	$\mathcal{O}(\sqrt{\log A}\,\|Q\|_1\,T^{1/2})$	no
Kernelized MWU	**(this paper)**	$\mathcal{O}(\sqrt{\log A}\,\sqrt{\|Q\|_1}\,T^{1/2})$	**no**
Optimistic FTRL / OMD (dilated entropy)	(Kroer et al., 2020)	$\mathcal{O}(\sqrt{m}\log(A)\,2^D\,\|Q\|_1^2\,T^{1/4})$	yes[*]
Optimistic FTRL / OMD (dilatable gl. ent.)	(Farina et al., 2021a)	$\mathcal{O}(\sqrt{m}\log(A)\,\|Q\|_1^2\,T^{1/4})$	no
Kernelized OMWU	**(this paper)**	$\mathcal{O}(m\log(A)\,\|Q\|_1\,\log^4(T))$	**yes**

Near-optimal O(polylog T) regret bound

Improved dependence on the ℓ_1 norm of the strategy space (half of the exponent)