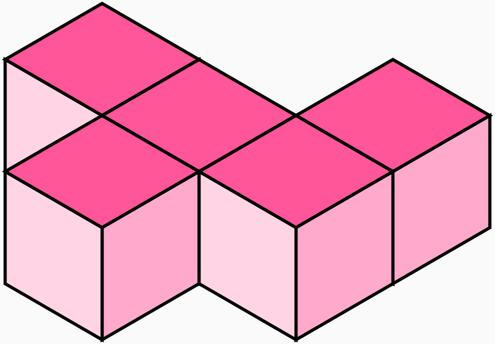
On the Computation of Saddle Points arising in Bandit Problems



Wouter M. Koolen

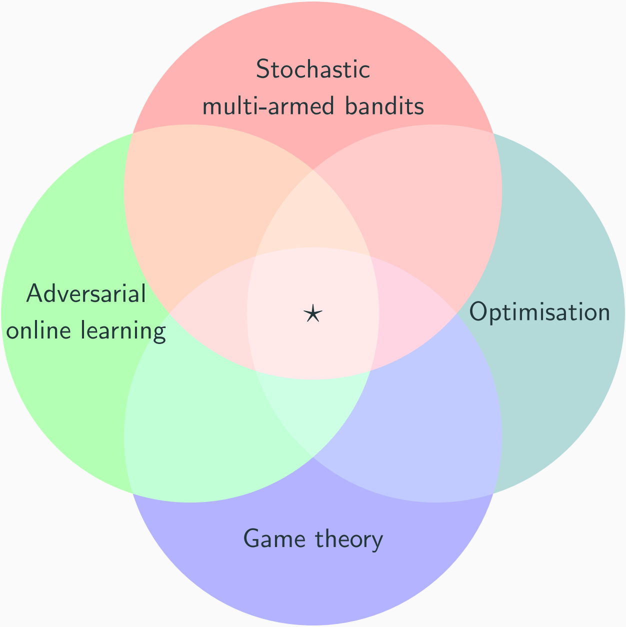
CWI
Centrum Wiskunde & Informatica

UNIVERSITY OF TWENTE.

KICK-OFF international research collaboration Inria-CWI

Thursday 14th September, 2023

Project


Stochastic
multi-armed bandits
Adversarial
online learning
⋆
Optimisation
Game theory

Sneak preview of Question

What is the computational complexity (# linear optimisation oracle calls) of approximately solving particular bilinear **saddle point** problems?

Motivation from Multi-Armed Bandits

Seminal Insight (Garivier and Kaufmann, 2016)

A **good** algorithm for Best Arm Identification must be solving

$$\boldsymbol{w}^* = \operatorname*{arg\,max}_{\boldsymbol{w} \in \triangle_K} \min_{\boldsymbol{\lambda} \in \Lambda} \sum_{k=1}^{K} w_k d(\mu_k, \lambda_k)$$

Motivation from Multi-Armed Bandits

Seminal Insight (Garivier and Kaufmann, 2016)

A **good** algorithm for Best Arm Identification must be solving

$$w^* \quad = \quad \operatorname*{arg\,max}_{\boldsymbol{w} \in \triangle_K} \min_{\boldsymbol{\lambda} \in \Lambda} \sum_{k=1}^{K} w_k d(\mu_k, \lambda_k)$$

Explosion of bandit papers with analogous problems:

- Top-m 2017; 2017
- Spectral 2021
- Stratified 2021
- Lipschitz 2019
- Linear 2020; 2020
- Threshold 2017
- MaxGap 2019
- Duelling 2021
- Contextual 2020; 2020
- Pareto 2023
- Minimum 2018
- MCTS 2016
- Markov 2019
- Tail-Risk 2021
- MDP 2021

The core abstract problem

Given $\mathcal{X} \subseteq \mathbb{R}^K$, consider the min-max problem

$$V^* \; := \; \min_{\boldsymbol{w} \in \triangle_K} \; \max_{\boldsymbol{x} \in \mathcal{X}} \; \boldsymbol{w}^\intercal \boldsymbol{x}$$

The core abstract problem

Given $\mathcal{X} \subseteq \mathbb{R}^K$, consider the min-max problem

$$V^* \;\coloneqq\; \min_{\boldsymbol{w} \in \triangle_K} \; \max_{\boldsymbol{x} \in \mathcal{X}} \; \boldsymbol{w}^\intercal \boldsymbol{x}$$

Definition

A point $\tilde{\boldsymbol{w}} \in \triangle_K$ is an **ϵ-optimal** solution if

$$\max_{\boldsymbol{x} \in \mathcal{X}} \; \tilde{\boldsymbol{w}}^\intercal \boldsymbol{x} \;\le\; V^* + \epsilon$$

The core abstract problem

Given $\mathcal{X} \subseteq \mathbb{R}^K$, consider the min-max problem

$$V^* \coloneqq \min_{\boldsymbol{w} \in \triangle_K} \max_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{w}^\intercal \boldsymbol{x}$$

Definition

A point $\tilde{\boldsymbol{w}} \in \triangle_K$ is an **ϵ-optimal** solution if

$$\max_{\boldsymbol{x} \in \mathcal{X}} \tilde{\boldsymbol{w}}^\intercal \boldsymbol{x} \leq V^* + \epsilon$$

Min-max theorem

There exists a saddle point $\boldsymbol{w}^* \in \triangle_K$, $\boldsymbol{x}^* \in \overline{\text{conv}}(\mathcal{X})$

The core abstract problem

Given $\mathcal{X} \subseteq \mathbb{R}^K$, consider the min-max problem

$$V^* \coloneqq \min_{\boldsymbol{w} \in \triangle_K} \max_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{w}^\intercal \boldsymbol{x}$$

Definition

A point $\tilde{\boldsymbol{w}} \in \triangle_K$ is an **ϵ-optimal** solution if

$$\max_{\boldsymbol{x} \in \mathcal{X}} \tilde{\boldsymbol{w}}^\intercal \boldsymbol{x} \le V^* + \epsilon$$

Min-max theorem

There exists a saddle point $\boldsymbol{w}^* \in \triangle_K$, $\boldsymbol{x}^* \in \overline{\text{conv}}(\mathcal{X})$

Problem

How to compute an ϵ-optimal point?

A beautiful line of work

$\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots \in \triangle_K$

$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in \overline{\mathsf{conv}}(\mathcal{X})$

Idea: have one adversarial online learning algorithm choose $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots \in \triangle_K$, and another $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in \overline{\mathsf{conv}}(\mathcal{X})$.

A beautiful line of work

$w_1, w_2, \ldots \in \triangle_K$

$x_1, x_2, \ldots \in \overline{\text{conv}}(\mathcal{X})$

Idea: have one adversarial online learning algorithm choose $w_1, w_2, \ldots \in \triangle_K$, and another $x_1, x_2, \ldots \in \overline{\text{conv}}(\mathcal{X})$.

Standard regret bounds for learners imply

Theorem (Freund and Schapire, 1999)

Time-average of T iterates is an $O(1/\sqrt{T})$-optimal saddle point.

A beautiful line of work

$\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots \in \triangle_K$

$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in \overline{\mathsf{conv}}(\mathcal{X})$

Idea: have one adversarial online learning algorithm choose $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots \in \triangle_K$, and another $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \in \overline{\mathsf{conv}}(\mathcal{X})$.

Standard regret bounds for learners imply

Theorem (Freund and Schapire, 1999)

Time-average of T iterates is an $O(1/\sqrt{T})$-optimal saddle point.

Acceleration with optimistic learning (exploiting **stability**)

Theorem (Rakhlin and Sridharan, 2013; Abernethy et al., 2018)

Time-average of T iterates is an $O(1/T)$-optimal saddle point.

Let's Implement!

Online learning for $\boldsymbol{w}_t \in \triangle_K$: Optimistic Hedge algorithm

Online learning for $\boldsymbol{x}_t \in \overline{\mathsf{conv}}(\mathcal{X})$: Online Gradient Descent

Let's Implement!

Online learning for $\boldsymbol{w}_t \in \triangle_K$: Optimistic Hedge algorithm

Online learning for $\boldsymbol{x}_t \in \overline{\text{conv}}(\mathcal{X})$: Online Gradient Descent

Snag

OGD and friends require Euclidean projection onto $\overline{\text{conv}}(\mathcal{X})$.

Let's Implement!

Online learning for $\boldsymbol{w}_t \in \triangle_K$: Optimistic Hedge algorithm

Online learning for $\boldsymbol{x}_t \in \overline{\text{conv}}(\mathcal{X})$: Online Gradient Descent

Snag

OGD and friends require Euclidean projection onto $\overline{\text{conv}}(\mathcal{X})$.

What access can we assume to $\mathcal{X}$ (or $\overline{\text{conv}}(\mathcal{X})$)?

Let's Implement!

Online learning for $\boldsymbol{w}_t \in \triangle_K$: Optimistic Hedge algorithm

Online learning for $\boldsymbol{x}_t \in \overline{\mathsf{conv}}(\mathcal{X})$: Online Gradient Descent

Snag

OGD and friends require Euclidean projection onto $\overline{\mathsf{conv}}(\mathcal{X})$.

What access can we assume to $\mathcal{X}$ (or $\overline{\mathsf{conv}}(\mathcal{X})$)?

Common assumptions

- Small finite set / polytope / polyhedron
- Linear optimisation oracle ← **this talk**

$$\boldsymbol{c} \quad \mapsto \quad \underset{\boldsymbol{x} \in \overline{\mathsf{conv}}(\mathcal{X})}{\arg\max} \ \boldsymbol{x}^\intercal \boldsymbol{c} \quad = \quad \underset{\boldsymbol{x} \in \overline{\mathcal{X}}}{\arg\max} \, \boldsymbol{x}^\intercal \boldsymbol{c}$$

- Membership oracle for $\overline{\mathsf{conv}}(\mathcal{X})$
- Euclidean projection onto $\overline{\mathsf{conv}}(\mathcal{X})$
- ...

The Question

Given $\mathcal{X} \subseteq \mathbb{R}^K$, consider the min-max problem

$$V^* := \min_{\boldsymbol{w} \in \triangle_K} \max_{\boldsymbol{x} \in \mathcal{X}} \boldsymbol{w}^\intercal \boldsymbol{x}$$

Question

What is the oracle complexity of computing an ϵ-optimal point using only linear optimisation access to $\mathcal{X}$. In particular, is $\epsilon = O(1/\sqrt{T})$, $\epsilon = O(1/T)$ or else?

What about Frank-Wolfe?

Frank-Wolfe family of algorithms: ?

Learning/optimisation algorithms using only **linear optimisation**

- Online Frank Wolfe (Hazan and Kale, 2012) has slow adversarial regret $O(T^{3/4})$.
- Faster rates for offline optimisation with smooth objective and/or strongly convex domain.
- Need to study versions with optimism or clairvoyance
- Saddle point interaction is far from worst-case data

What about MetaGrad?

Best Response over $\overline{\mathcal{X}}$ is a single linear optimisation call.

Partial Positive Result

MetaGrad vs Best Response gives a $\tilde{O}(\frac{K}{T})$ rate.

- Fast rate without stability. Fun!
- Expensive to compute: $O(K^2)$ per iteration.
- Factor dimension (K) should not be there.

What about MetaGrad?

Best Response over $\overline{\mathcal{X}}$ is a single linear optimisation call.

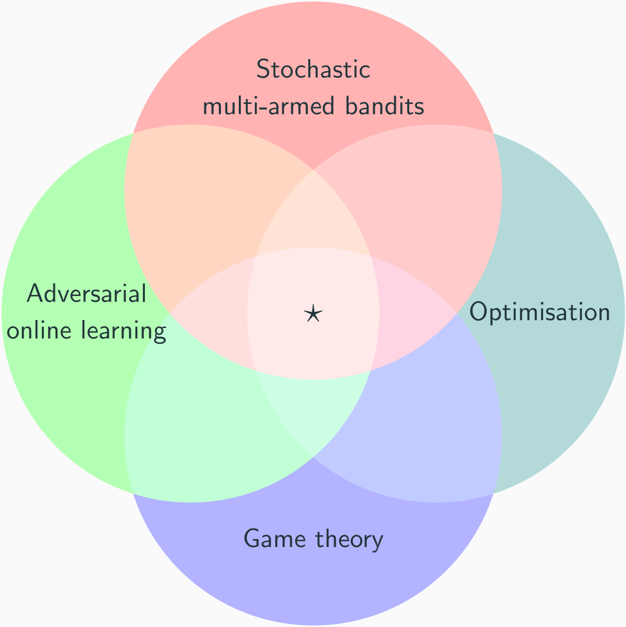
Partial Positive Result

MetaGrad vs Best Response gives a $\tilde{O}(\frac{K}{T})$ rate.

- Fast rate without stability. Fun!
- Expensive to compute: $O(K^2)$ per iteration.
- Factor dimension (K) should not be there.

$O(K^2)$ computation between oracle calls also arises for cutting plane methods (Ellipsoid, ...)

You are welcome to join



Thanks!

References i

- Abernethy, J. D., K. A. Lai, K. Y. Levy, and J. Wang (2018). **“Faster Rates for Convex-Concave Games”.** In: **COLT**. Vol. 75. Proceedings of Machine Learning Research.
- Agrawal, S., W. M. Koolen, and S. Juneja (Dec. 2021). **“Optimal Best-Arm Identification Methods for Tail-Risk Measures”.** In: **Advances in Neural Information Processing Systems (NeurIPS) 34**.
- Al Marjani, A. and A. Proutiere (2021). **“Adaptive sampling for best policy identification in markov decision processes”.** In: **International Conference on Machine Learning**. PMLR.
- Chen, J., X. Chen, Q. Zhang, and Y. Zhou (2017). **“Adaptive multiple-arm identification”.** In: **International Conference on Machine Learning**. PMLR.
- Chen, L., J. Li, and M. Qiao (2017). **“Nearly instance optimal sample complexity bounds for top-k arm selection”.** In: **Artificial Intelligence and Statistics**. PMLR.

Degenne, R., W. M. Koolen, and P. Ménard (Dec. 2019). **"Non-Asymptotic Pure Exploration by Solving Games".** In: **Advances in Neural Information Processing Systems (NeurIPS) 32**.

Degenne, R., P. Ménard, X. Shang, and M. Valko (2020). **"Gamification of pure exploration for linear bandits".** In: **International Conference on Machine Learning**. PMLR.

Freund, Y. and R. Schapire (1999). **"Adaptive game playing using multiplicative weights".** In: **Games and Economic Behavior** 29.

Garivier, A. and E. Kaufmann (2016). **"Optimal Best arm Identification with Fixed Confidence".** In: **Proceedings of the 29th Conference On Learning Theory (COLT)**.

Garivier, A., E. Kaufmann, and W. M. Koolen (June 2016). **"Maximin Action Identification: A New Bandit Framework for Games".** In: **Proceedings of the 29th Annual Conference on Learning Theory (COLT)**.

- Garivier, A., P. Ménard, and L. Rossi (2017). **"Thresholding bandit for dose-ranging: The impact of monotonicity".** In: **arXiv preprint arXiv:1711.04454**.
- Haddenhorst, B., V. Bengs, and E. Hüllermeier (2021). **"Identification of the generalized Condorcet winner in multi-dueling bandits".** In: **Advances in Neural Information Processing Systems** 34.
- Hao, B., T. Lattimore, and C. Szepesvari (2020). **"Adaptive exploration in linear contextual bandit".** In: **International Conference on Artificial Intelligence and Statistics**. PMLR.
- Hazan, E. and S. Kale (2012). **"Projection-free online learning".** In: **29th International Conference on Machine Learning, ICML 2012**.
- Jedra, Y. and A. Proutiere (2020). **"Optimal best-arm identification in linear bandits".** In: **Advances in Neural Information Processing Systems** 33.

- Katariya, S., A. Tripathy, and R. Nowak (2019). **“Maxgap bandit: Adaptive algorithms for approximate ranking”.** In: **Advances in Neural Information Processing Systems** 32.
- Kaufmann, E., W. M. Koolen, and A. Garivier (Dec. 2018). **“Sequential Test for the Lowest Mean: From Thompson to Murphy Sampling”.** In: **Advances in Neural Information Processing Systems (NeurIPS) 31**.
- Kocák, T. and A. Garivier (2021). **“Epsilon Best Arm Identification in Spectral Bandits.”.** In: **IJCAI**.
- Kone, C., E. Kaufmann, and L. Richert (2023). **“Adaptive Algorithms for Relaxed Pareto Set Identification”.** In: **arXiv preprint arXiv:2307.00424**.
- Moulos, V. (2019). **“Optimal best Markovian arm identification with fixed confidence”.** In: **Advances in Neural Information Processing Systems** 32.

References v

Rakhlin, S. and K. Sridharan (2013). **"Optimization, Learning, and Games with Predictable Sequences".** In: **Advances in Neural Information Processing Systems 26**.

Russac, Y., C. Katsimerou, D. Bohle, O. Cappé, A. Garivier, and W. M. Koolen (Dec. 2021). **"A/B/n Testing with Control in the Presence of Subpopulations".** In: **Advances in Neural Information Processing Systems (NeurIPS) 34**.

Tirinzoni, A., M. Pirotta, M. Restelli, and A. Lazaric (2020). **"An asymptotically optimal primal-dual incremental algorithm for contextual linear bandits".** In: **Advances in Neural Information Processing Systems** 33.