Incremental Multiple Oracle

Extended Abstract

Carlos Martin
Carnegie Mellon University
Pittsburgh, United States
cgmartin@cs.cmu.edu

Tuomas Sandholm
Carnegie Mellon University
Strategy Robot, Inc.
Optimized Markets, Inc.
Strategic Machine, Inc.
Pittsburgh, United States
sandholm@cs.cmu.edu

ABSTRACT

We present a framework for computing approximate mixed-strategy Nash equilibria of continuous-action games. It is a modification of the traditional double oracle algorithm, extended to multiple players and continuous action spaces. Unlike prior methods, it maintains fixed-cardinality pure strategy sets for each player. Thus, unlike prior methods, only a constant amount of memory is necessary. Furthermore, it does not require exact metagame solving on each iteration, which can be computationally expensive for large metagames. Moreover, it does not require global best-response computation on each iteration, which can be computationally expensive or even intractable for high-dimensional action spaces and general games. Our method incrementally reduces the exploitability of the strategy profile in the finite metagame, pushing it toward Nash equilibrium. Simultaneously, it incrementally improves the pure strategies that best respond to this strategy profile in the full game. We test our method on various continuous games. It obtains approximate mixed-strategy Nash equilibria with low exploitability.

KEYWORDS

game theory; game solving; equilibrium finding; continuous games

ACM Reference Format:
Carlos Martin and Tuomas Sandholm. 2026. Incremental Multiple Oracle: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/ea0097NFSI9839

1 INTRODUCTION

Research on computing *Nash equilibria (NE)* in games has mostly focused on settings with finite, discrete action spaces. However, many games involving space, money, or time have continuous action spaces. These include security games in continuous spaces [11–13], resource allocation games [6], network games [8], simulations of military scenarios and wargaming [17], and video games [3, 23]. Also, even if the action space is discrete, it may be fine-grained enough to treat as continuous in order to improve the computational efficiency of equilibrium finding [4, 5, 7]. The usual approach to computing an equilibrium of a game with continuous action spaces involves discretizing the action space. That entails a loss in solution quality [14] and does not scale well to high-dimensional spaces. Therefore, other approaches are called for.

We present a new algorithm for computing approximate mixed-strategy NE in continuous-action games. It is a modification of the double oracle algorithm, extended to multiple players and continuous action spaces. The method incrementally reduces the exploitability of the strategy profile in the finite metagame, pushing it toward NE. Simultaneously, it incrementally improves the pure strategies that best respond to this strategy profile in the full game. We evaluate the method on various continuous-action games, showing that it obtains approximate mixed-strategy NE with low exploitability. Unlike prior methods, this method runs in constant memory, because it maintains fixed-cardinality pure strategy sets for each player. Furthermore, it does not require exact metagame solving on each iteration, which can be computationally expensive for large metagames. Moreover, it does not require global best-response computation on each iteration, which can be computationally expensive or intractable for general games.

2 RELATED RESEARCH

McMahan et al. [19] introduced *Double Oracle (DO)*, an algorithm that computes NE of two-player zero-sum normal-form games. Adam et al. [1] extended DO to two-player zero-sum games with *continuous* compact strategy sets. Kroupa and Votroubek [15] extended that to n players, under the name *Multiple Oracle (MO)*. It starts with nonempty finite supports for each player. On each iteration, it performs the following two steps. First, it computes an equilibrium of the finite metagame consisting of the restriction of the full game to the finite subsets. Second, for each player, it computes a *best response (BR)* to this equilibrium in the *full* game, and adds it to the player's finite subset. The BR computation is based on global solvers for special classes of utility functions.

3 PROPOSED METHOD

The MO algorithm has some limitations. First, metagame solving on each iteration can be expensive, especially when the metagame is large. Second, global BR computation on each iteration can be expensive or even intractable (e.g., for high-dimensional spaces or general utility functions). Third, the memory it requires is not fixed, but grows over time. To address these limitations, we introduce a new algorithm, which we call *Incremental Multiple Oracle (IMO)*. Unlike MO, this algorithm maintains fixed-cardinality supports for each player, thus requiring only constant memory. Furthermore,

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/ea0097NFSI9839

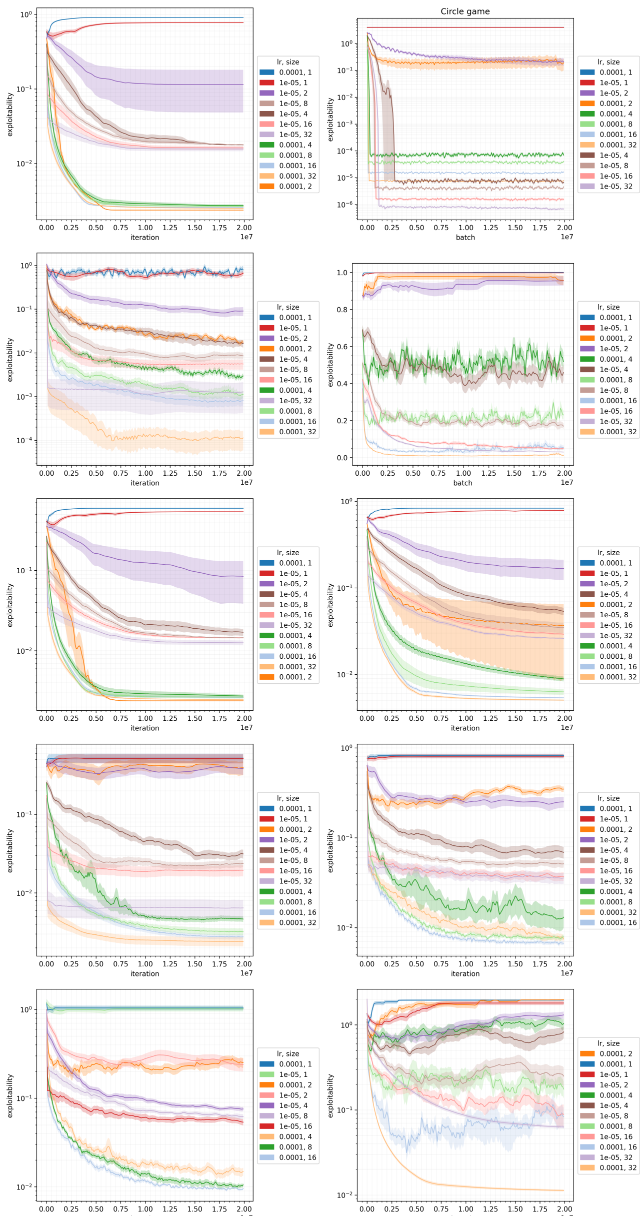

Figure 1: Experimental results. Left to right, top to bottom: Interval game, circle game, Glicksberg–Gross game, Colonel Blotto, security game, all-pay auction, and chopstick auction.

it replaces the exact or global oracles of the MO algorithm with incremental or approximate ones. On each iteration, the algorithm works as follows. First, it constructs the metagame of utilities that is induced by the current supports of the players. For each player, it adjusts the probability assigned to each pure strategy in order to decrease the exploitability of the induced mixed strategy profile in the metagame. This is called *exploitability descent (ED)* [16, 18]. Second, it improves approximate BRs to the induced strategy profile in the full game. Specifically, it adjusts the pure strategies that happen to be BRs to the current approximate equilibrium, in order to increase the corresponding player's utility against that equilibrium. For comparison, MO maintains a set of strategies that is expanded on each iteration with approximate BRs to the meta-strategies of the other players. These strategies are *static*, that is, they do not change after they are added. In contrast, our algorithm *dynamically* improves them over the course of training. Thus the supports do not need to be augmented on each iteration, but instead improve autonomously over time. Our algorithm is the first, to our knowledge, that tackles this setting in full generality, making as few assumptions as possible (e.g., no BR or metagame equilibrium oracles).

4 EXPERIMENTS

In our experiments, we use 10^5 iterations per epoch, and 16 trials per experiment. In our plots, solid lines show the mean across trials, and bands show its standard error. In each legend, "lr" means the learning rate and "size" means the size of each player's support. The X axis shows the number of iterations of the algorithm. The Y axis shows the exploitability of the learned strategy profile *in the full game*. The exploitability in the full game is approximated via a simple maximization of utility over a fine-grained discretization of the action space. Each experiment ran on one NVIDIA A100 SXM4 40GB GPU on a computer cluster running Springdale Linux 8.6. A brief description of each game is as follows. Let $\beta = 20$. **Interval game** [20]: Two players pick points on the unit interval and $u_1 = -u_2 = (x-y)^2$. **Circle game**: Two players pick points on the unit circle and $u_1 = -u_2 = \|x-y\|_2^2$. **Glicksberg–Gross game** [9]: Two players pick points on the unit interval and $u_1 = -u_2 = \frac{(1+x)(1+y)(1-xy)}{(1+xy)^2}$. **Security game** [11–13]: Two players pick points on the unit interval and $u_1 = -u_2 = e^{-\beta(x-y)^2}$. **Continuous Colonel Blotto** [10]: Players pick points on the standard 2-simplex (each of the 3 vertices is a battlefield) and $u = \mathrm{sum}_{\mathrm{items}}(\mathrm{softmax}_{\mathrm{players}}(\beta a))$. **Complete-information all-pay auction** [2]: Players pick points on the unit interval and $u = \mathrm{softmax}(\beta a) - a$. **Chopstick auction** [21, 22]: 3 chopsticks are sold simultaneously in separate first-price sealed-bid auctions. A player gets utility 1 if it wins at least 2 chopsticks, and 0 otherwise. As the plots show, when the support size is sufficiently large, the algorithm yields approximate NE with low exploitability.

5 CONCLUSION

We introduced an algorithm for computing approximate mixed-strategy NE of continuous-action games. It is a modification of MO, extended to continuous action spaces and multiple agents. It gradually improves a metagame equilibrium while simultaneously and gradually improving pure strategies against it. It mitigates the need for exact equilibrium and/or global best-response computation on each iteration. We tested our algorithm on various continuous-action games and observed its convergence to approximate NE.

ACKNOWLEDGMENTS

This material is based on work supported by the Vannevar Bush Faculty Fellowship ONR N00014-23-1-2876, NSF grants RI-2312342 and RI-1901403, ARO award W911NF2210266, and NIH award A240108S001.

REFERENCES

[1] Lukáš Adam, Rostislav Horčík, Tomáš Kasl, and Tomáš Kroupa. 2021. Double oracle algorithm for computing equilibria in continuous games. *AAAI Conference on Artificial Intelligence (AAAI)* 35, 6 (2021), 5070–5077.

[2] Michael Baye, Dan Kovenock, and Casper de Vries. 1996. The all-pay auction with complete information. *Economic Theory* 8, 2 (1996), 291–305.

[3] Christopher Berner, Greg Brockman, Brooke Chan, Vicki Cheung, Przemysław Debiak, Christy Dennison, David Farhi, Quirin Fischer, Shariq Hashme, Chris Hesse, et al. 2019. Dota 2 with Large Scale Deep Reinforcement Learning. arXiv:1912.06680 [cs.LG]

[4] Émile Borel. 1938. *Traité du calcul des probabilités et ses applications*. Applications aux jeux des hazard, Vol. 4. Gauthier-Villars, Paris.

[5] Bill Chen and Jerrod Ankenman. 2006. *The Mathematics of Poker*. ConJelCo, Pittsburgh, PA.

[6] Sam Ganzfried. 2021. Algorithm for computing approximate Nash equilibrium in continuous games with application to continuous Blotto. *Games* 12, 2 (2021), 47.

[7] Sam Ganzfried and Tuomas Sandholm. 2010. Computing equilibria by incorporating qualitative models. In *Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 183–190.

[8] Papiya Ghosh and Rajendra P. Kundu. 2019. Best-shot network games with continuous action space. *Research in Economics* 73, 3 (2019), 225–234.

[9] Irving Glicksberg and Oliver Gross. 1953. *Notes on games over the square*. Princeton University Press, Princeton, NJ, 173–182.

[10] Oliver Alfred Gross and R. A. Wagner. 1950. *A continuous Colonel Blotto game*. RAND Corporation, Santa Monica, CA.

[11] Nitin Kamra, Fei Fang, Debarun Kar, Yan Liu, and Milind Tambe. 2017. Handling continuous space security games with neural networks. In *IWAISe: First International Workshop on Artificial Intelligence in Security*. Springer Cham, Cham, Switzerland, 17.

[12] Nitin Kamra, Umang Gupta, Fei Fang, Yan Liu, and Milind Tambe. 2018. Policy learning for continuous space security games using neural networks. *AAAI Conference on Artificial Intelligence (AAAI)* 32, 1 (2018), 1103–1112.

[13] Nitin Kamra, Umang Gupta, Kai Wang, Fei Fang, Yan Liu, and Milind Tambe. 2019. DeepFP for finding Nash equilibrium in continuous action spaces. In *Decision and Game Theory for Security*. Springer-Verlag, Berlin, Heidelberg, 238–258.

[14] Christian Kroer and Tuomas Sandholm. 2015. Discretization of Continuous Action Spaces in Extensive-Form Games. In *Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), Richland, SC, 47–56.

[15] Tomáš Kroupa and Tomáš Votroubek. 2022. Multiple oracle algorithm to solve continuous games. In *International Conference on Decision and Game Theory for Security (GameSec)*. Springer-Verlag, Berlin, Heidelberg, 149–167.

[16] Edward Lockhart, Marc Lanctot, Julien Pérolat, Jean-Baptiste Lespiau, Dustin Morrill, Finbarr Timbers, and Karl Tuyls. 2019. Computing approximate equilibria in sequential adversarial games by exploitability descent. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. International Joint Conferences on Artificial Intelligence Organization, Marina del Rey, CA, USA, 464–470.

[17] Alberto Marchesi, Francesco Trovò, and Nicola Gatti. 2020. Learning Probably Approximately Correct Maximin Strategies in Simulation-Based Games with Infinite Strategy Spaces. In *Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), Richland, SC, 834–842.

[18] Carlos Martin and Tuomas Sandholm. 2025. ApproxED: Approximate Exploitability Descent via Learned Best Responses. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems* (Detroit, MI, USA) *(AAMAS '25)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1454–1463.

[19] H. Brendan McMahan, Geoffrey J. Gordon, and Avrim Blum. 2003. Planning in the presence of cost functions controlled by an adversary. In *International Conference on Machine Learning (ICML) (ICML'03)*. AAAI Press, Washington DC, USA, 536–543.

[20] Pablo Parrilo. 2006. Polynomial games and sum of squares optimization. In *IEEE Conference on Decision and Control (CDC)*. IEEE, Piscataway, NJ, 2855–2860.

[21] Balazs Szentes and Robert Rosenthal. 2003. Beyond Chopsticks: Symmetric equilibria in majority auction games. *Games and Economic Behavior* 45, 2 (2003), 278–295.

[22] Balázs Szentes and Robert W. Rosenthal. 2003. Three-object two-bidder simultaneous auctions: chopsticks and tetrahedra. *Games and Economic Behavior* 44, 1 (2003), 114–133.

[23] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* 575 (2019), 350–354.