Technische Universität München
School of Computation, Information and Technology

Multi-Agent Reinforcement Learning for the Computation of Market Equilibria

Nils Kohring

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Julien Gagneur

Prüfende der Dissertation:

1. Prof. Dr. Martin Bichler
2. Prof. Dr. Florian Matthes
3. Prof. Ioannis Panageas, Ph. D.

Die Dissertation wurde am 28.06.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 23.05.2024 angenommen.

To my late grandfather Günter.

Abstract

Understanding and analyzing both the dynamics of self-interested agents in markets and the possible resulting equilibrium outcomes are of the utmost importance for economic theory. Yet, a general framework for equilibrium computation, particularly in auction theory, is still lacking. Equilibrium outcomes react sensitively to changes in the underlying properties and assumptions, such as the information structure, the agents' utility preferences, and the pricing mechanism. Previous methods for analytically or numerically solving for the equilibrium typically are subject to two fundamental limitations. They either rely on exploiting particular market properties, which limits their applicability to more realistic scenarios, or they are computationally demanding, which limits their scalability to realistic market sizes. Therefore, except for relatively small and restricted economic settings, we lack the knowledge of equilibria and the tools to compute them. This thesis proposes a general and scalable approach for equilibrium computation based on reinforcement learning methods. We model the agents' strategies by neural networks that learn to bid optimally through repeated self-play. The application of such learning dynamics, where all agents simultaneously adapt their behavior, is challenging in general games. On top of that, the technical implementation for markets specifically is involved. This is mainly due to the agents' discontinuous utility functions, which make the application of gradient-based methods more challenging, and the agents' continuous valuation and action spaces, which necessitate the search for optimal strategies in an infinite-dimensional function space. The classes of gradient-based and particle-swarm-based approaches and their respective trade-offs regarding accuracy and computational efficiency are examined. We provide strong numerical results where these methods successfully compute approximate equilibria in markets ranging from small single-item auctions to larger combinatorial auctions with multiple heterogeneous goods and double auctions, where buyers and sellers act strategically. We successfully extend the algorithms to markets where the bidders' valuations for the goods are correlated and to markets with more general preference structures, such as those that account for risk aversion. Finally, we show local convergence of the gradient-based approach under additional regularity conditions. With this, we pave the way for employing reinforcement learning methods in settings of economic theory, which ultimately may assist economists in better understanding market dynamics and their equilibria.

Keywords: Auction Theory • Equilibrium Computation • Machine Learning • Multi-Agent Reinforcement Learning

Zusammenfassung

Die Analyse der Dynamik von selbstinteressierten Agenten in Märkten sowie der resultierenden Gleichgewichte ist von höchster Bedeutung für die Wirtschaftstheorie. Dennoch existiert kein allgemeingültiger Ansatz, solche Gleichgewichte zu berechnen, insbesondere in der Auktionstheorie. Gleichgewichte reagieren sehr empfindlich auf Änderungen in den Markteigenschaften und -annahmen, wie zum Beispiel der Informationsstruktur, den Nutzenfunktionen und den Preismechanismen. Bisherige Ansätze für das analytische oder numerische Berechnen von Gleichgewichten haben im Allgemeinen zwei fundamentale Probleme. Entweder beruhen sie auf der Ausnutzung von speziellen Markteigenschaften, was ihre Anwendbarkeit für realistischere Märkte limitiert, oder sie erfordern erheblichen Rechenaufwand, was die Skalierbarkeit zu größeren Märkten limitiert. Abgesehen von sehr kleinen ökonomischen Modellen, sind daher weder Gleichgewichte noch Methoden diese zu berechnen, bekannt. In dieser Dissertation stellen wir einen allgemeinen und skalierbaren Ansatz zur Gleichgewichtsberechnung basierend auf Methoden des Reinforcement Learning vor. Die Strategien von Agenten werden durch neuronale Netze modelliert, die durch wiederholtes Ausprobieren und Adaptieren optimales Bieten lernen. Die Anwendung solcher Lerndynamiken, in denen Agenten gleichzeitig ihr Verhalten anpassen, ist in Spielen generell herausfordernd. Darüber hinaus ist die technische Umsetzung, speziell in Märkten, erschwert. Dies ist zum einen durch die unstetigen Nutzenfunktionen der Agenten bedingt, was die Anwendung von gradientbasierten Methoden erschwert, und zum anderen durch die kontinuierlichen Werte- und Aktionsbereiche der Agenten, welche die Suche nach optimalen Strategien in einem unendlich dimensionalen Funktionenraum bedingen. Die Klassen von gradientenbasierten Algorithmen und Partikelschwarmoptimierung werden mit ihren jeweiligen Abwägungen auf Genauigkeit und Rechenbedarf untersucht. Die numerischen Ergebnisse sind durchweg positiv, da diese Lernverfahren erfolgreich Annäherungen der Gleichgewichte berechnen. Dies gelingt in kleinen Auktionen mit einem Gut sowie in größeren kombinatorischen Auktionen mit mehreren heterogenen Gütern und in Double-Auction-Märkten, in denen sowohl die anbietenden als auch die bietenden Agenten strategisch agieren. Wir erweitern die Anwendung dieser Algorithmen erfolgreich auf Märkte, in denen die Bewertungen der Güter abhängig voneinander sind und auf Märkte mit allgemeingültigeren Nutzenfunktionen, z. B. auf solche die Risikoaversion abbilden. Des Weiteren zeigen wir lokale Konvergenz des gradientenbasierten Ansatzes unter zusätzlichen Regularitätsannahmen. Insgesamt zeigen wir damit einen Weg zur Anwendung von Reinforcement Learning in mikroökonomischen Modellen auf, der auf lange Sicht Wirtschaftswissenschaftler und -wissenschaftlerinnen dabei unterstützen kann, Marktdynamiken und ihre Gleichgewichte besser zu verstehen.

Stichwörter: Auktionstheorie • Gleichgewichtsberechnung • Maschinelles Lernen • Multi-Agent Reinforcement Learning

Acknowledgements

I am grateful to a number of supportive people without whom this dissertation would not have been possible. Firstly, I want to sincerely thank my supervisor, Martin. Working with him was always both a valuable learning experience and a fruitful exchange. I am thankful for his guidance and supervision throughout my dissertation project, and the opportunities working as part of his research group gave me. Furthermore, I want to thank Martin for the helpful suggestions regarding substance and exposition on an early draft of this dissertation. Also, I am thankful to the examination committee and especially Prof. Matthes and Prof. Panageas for taking the time and effort to review my thesis.

Secondly, I want to thank the members of the Chair of Decision Sciences and Systems. The cooperative atmosphere while discussing new ideas and conducting joint research projects was an excellent and inspiring environment. In particular, I want to thank my main project partners, Stefan, with whom I worked on this dissertation's first major piece of research and corresponding software project, and Fabian, with whom I worked on subsequent projects. Furthermore, I want to thank the remaining coauthors of my work: Paul, Max, and Matthias. Furthermore, I thank Anja and the remaining colleagues at the chair for our successful coordination of administrative and teaching-related matters.

Lastly, I want to thank my parents Christine and Arne, my siblings Amelie and Lennart, and my girlfriend Teresa for their continued support throughout this academic endeavor.

Contents

Chapter 1

Introduction

Economic theory is the study of relationships in markets by using formal mathematical constructs. Considering the rules and regulations of a market, one strives for an understanding of the dynamics of various economic elements, such as the allocation of goods and the behavior of agents. How do new technological advances, market entries and exits, or changes by the policymaker influence the market dynamics, such as prices or supply and demand?

Economists have always debated the amount and type of regulatory interventions necessary. Over the past decades, some desiderata, such as ensuring market efficiency, stability, and fairness, were not explicitly enforced by regulation, but delegated to the markets' pricing dynamics. Discussions on these topics are often driven by ideology, and assessments are difficult to conduct. However, some recent studies aim to objectively and quantitatively evaluate economic regulation (Parker and Kirkpatrick, 2012; Vannoni and Morelli, 2021). It remains largely an open question how various incentives may influence market participants in implicitly agreeing upon, ideally, a favorable outcome or equilibrium.

Analytically predicting equilibria has only been achieved under very specific market designs. Any extensions of existing models or relaxations of their assumptions can fundamentally change the equilibrium behavior and may render methods of analytically deriving equilibria inapplicable. Knowledge of equilibria enables economists and policymakers to compare several scenarios and prevent market failures. For a systematic approach to such questions, auction theory provides the mathematical framework by modeling the interaction of buyers and sellers under different information structures (Krishna, 2009). The focus on an algorithmic perspective on such problems has received more attention in recent years. This was mainly driven by the availability of ever more computational power and market data as well as the emergence of new digital markets (Tardos and Vazirani, 2007) and algorithmically guided agents (Calvano et al., 2020). Examples of such markets include online display advertisement auctions for sponsored search or on social media, public spectrum auctions for mobile network frequencies, or financial exchanges where bids and offers must be matched. The domains of market and mechanism design are more specifically concerned with the influencing factors of agents' utilities and the setup of markets where agents reveal their preferences truthfully in equilibrium. It turns out that this requires rigorous assumptions, and even the outcomes of the famous Vickrey-Clarke-Groves mechanism (Groves, 1973, most general form) fail to achieve such favorable properties in many settings (Conitzer and Sandholm, 2006; Guo and Conitzer, 2009). Therefore, economists usually face a trade-off between multiple design criteria, such as incentivizing truthfulness or welfare maximization, that do not preclude one another (Bichler, 2017), and access to equilibria is crucial in this decision-making process.

As a computational approach to learning from interaction, *reinforcement learning* (RL) has been successfully applied to a wide range of applications in recent years (Mnih et al., 2015; Silver et al., 2016). RL agents explore their available actions but must exploit those that proved good in earlier rounds via trial-and-error and without human guidance. These actions may determine not only current but also future utilities. Most algorithms in the RL literature focus on single-agent learning. In contrast, the theory of learning in games studies the processes in which multiple, usually self-interested, agents are iteratively trying to maximize their utilities by adapting to the environment and their opponents (Fudenberg et al., 1998).

1.1 Research Question

In light of the importance of equilibria in models of auction theory and our limited knowledge of them, the objective of this dissertation is the computation of such equilibria. Few computational approaches to *learning to bid* exist for continuous-action games with incomplete information. Most models drastically restrict the auctions to finite action spaces, independent private valuations, single-object auctions, or quasi-linear utilities. This reduces the problem's complexity and allows the application of conventional learning algorithms. Two of the most notable approaches are from Bosshard et al. (2020) and Balseiro and Gur (2019). In the former, bidding strategies are learned in discretized combinatorial auctions under the assumptions of risk neutrality and independent valuations. Crucially, the introduced error by discretizing the action space (and valuation space) is hard to quantify in general, and there is no guarantee that an ever finer discretizing granularity will lead to a better approximation of the continuous game (Waugh et al., 2009). The latter has a different approach by heavily restricting the action space to linear bid functions in sequential auctions. Both approaches to simplifying the equilibrium problem allow the establishment of convergence guarantees and corresponding bounds on the utility loss when following the learned strategies compared to optimal bidding.

We aim to overcome these limitations by proposing a generally capable RL framework for computing optimal strategic behavior in market environments. We model the agents' strategies by neural networks, which adapt via trial-and-error during repeated simulations of the market. The main finding is strong evidence for consistent equilibrium convergence. But we further investigate under which circumstances learning dynamics converge as well as their limitations in terms of the market's complexity and scalability to larger scenarios.

There are some fundamental challenges to this computational endeavor of the equilibrium problem. It has been shown that finding or even approximating equilibria is a problem of high computational complexity (Daskalakis et al., 2009). Furthermore, there are results on algorithms not converging to a steady state, and the currently known conditions on guaranteed convergence are prohibitively restrictive (Hsieh et al., 2021; Mertikopoulos and

Zhou, 2019). They are rarely met beyond game theoretic toy examples (such as potential or two-player, zero-sum games). Additionally, there are two main technical challenges. Firstly, state-of-the-art RL methods are prohibitively sample-inefficient. Hence, they require access to a highly performant simulation engine of the market. Secondly, most machine learning algorithms are built on gradient-based updating rules that almost exclusively rely on back-propagation. However, this procedure is not applicable to first-order gradient estimation in markets because it assumes a continuous objective. The simplest example of this issue is the binary decision of whether or not an agent is allocated an item and the resulting jump in the agent's utility. In view of this, this dissertation's contributions can be summarized as follows.

1.2 Contribution

In the first publication (Bichler et al., 2021, see chapter 4), we have proposed a new RL method for approximating market equilibria, called *neural pseudogradient ascent* (NPGA). Agents employ neural networks as their strategies and let them submit competing bids. They adjust their strategies by repeated interaction and subsequently following the gradient in the direction of higher expected utility. Thereby, the networks reach equilibrium without having to solve the corresponding equilibrium condition explicitly. The article focuses on symmetric auctions, that is, settings where the bidders face exactly the same situation ex-ante, e.g., they have the same prior valuation distributions and preferences. However, unlike previous work, NPGA can handle diverse correlations between the bidders' preferences and general utility functions, such as those that account for risk aversion. The analysis of convergence is approached from a theoretical and an empirical standpoint. The former uses the current understanding of learning dynamics and establishes local convergence under relatively strict conditions. The latter consists of a suite of experiments in specific auctions where convergence to equilibrium is observed and a verification technique for the learned strategies. By utilizing advanced conditional sampling techniques and an exhaustive search of the discretized action space for approximate best responses, a close proximity of the learned strategies to the maximal utility can be verified. The learning and verification procedures are applicable to an extensive range of environments. Despite being optimized for parallel execution on *graphics processing units* (GPUs), it remains computationally challenging considering the dimensionality of multi-agent auction games, mainly driven by the discretization granularity of the verifier and the sampling size. NPGA provides a generic approach for learning to bid in auctions with different distributional assumptions, valuation interdependencies, or risk attitudes.

The second publication (Bichler et al., 2023, see chapter 5) extends the empirical analysis of NPGA to a range of asymmetric markets. This extension is crucial for modeling more realistic markets. These include larger combinatorial auctions from the family of local-local-global auctions that are used to model high-stakes spectrum auctions. Having access to approximate equilibria computed by NPGA helps to understand phenomena such as more aggressive bidding by buyers with lower valuations, which is essential from a practitioner's standpoint. We find empirically that NPGA performs just as well as in symmetric auctions.

In the third publication of this dissertation (Bichler et al., 2022, see chapter 6), the above learning method, as well as another one based on a discretization of the market, has been extended to models of bilateral bargaining. This is the basic model of double auctions, as they can be observed in financial markets where multiple sellers and buyers come together. We again find robust convergence empirically and present a local convergence guarantee. All settings have been implemented in a vectorized fashion such that hundreds of thousands of games can be simulated in parallel. This makes training the deep neural network policies exceptionally efficient. The combined codebase of these first three publications establishes the most expansive suite of implemented market mechanisms.

We also proposed and analyzed two alternative updating approaches. The discrete nature of allocating indivisible goods to economic agents renders the game non-differentiable and performant first-order methods inapplicable. The default algorithmic choices, such as the REINFORCE or NPGA algorithm, rely on zeroth-order gradient feedback and hence suffer from high variance and computational costs. In the fourth publication of this thesis (Kohring et al., 2023, see chapter 7), we construct a surrogate market model in which the application of first-order gradient estimators becomes possible. As we show empirically and theoretically, the bias resulting from smoothing the utility function and the empirical variance can be controlled for in practice, making this procedure a substantially more performant and robust option for learning to bid. The surrogate approach exploits the structure of the utility function; thus, it remains open to what extent it can be generalized to markets beyond the simultaneous sale of independent goods, such as full-scale combinatorial auctions. Lastly, in (Kohring et al., 2022, not included workshop paper), each bidder deploys a particle swarm that heuristically searches for the optimal parameters of the bidder's neural network strategy. It replaces the conventional gradient-based updating and comes with a greater ability to escape local optima, as it keeps track of multiple candidate solutions (so-called particles) but with less mathematical justification. Only some limited stability analysis has been conducted for the single-agent case. Empirically, we find the results to be on a par with the gradient-based NPGA algorithm, both in terms of accuracy and robustness as well as the computational needs.

1.3 Outline

The remainder of this dissertation is structured as follows. First, we will outline the concepts of game and auction theory (chapter 2) that set the stage for the equilibrium learning algorithms considered in this work. Before presenting these algorithms and applying them to learning in markets specifically, the fundamentals of learning in games are briefly revisited by introducing classic learning dynamics (chapter 3). At last, after presenting the publications themselves, we summarize the contributions and conclude with open questions and future work in chapter 8.

Chapter 2

Fundamentals of Game Theory

Game theory dates back to the work of von Neumann and Morgenstern (1944) and Nash (1951). It concerns itself with the strategic interactions among rational agents that are assumed to follow some exogenous objective. That is, agents are aware of possible alternatives, form expectations about uncertainties, and choose actions after some process of optimization over their objectives. Thereby, game theory allows the creation and subsequent analysis of abstract representations of real-life problems involving multiple parties. A game defines the rules of the interactions by stating a set of possible actions for players and defining how actions influence the players' individual objectives. We refer the interested reader to (Osborne and Rubinstein, 1994) for more details on the underlying game theory concepts.

Economic theory leverages much of this formalism. It constructs games representing parts of the economy or individual markets and focuses on elements of economic interest, such as market efficiency and welfare. Typically, markets are modeled as non-cooperative games where agents possess private information.

2.1 Types of Games

A complete and exhaustive taxonomy of all types of games is elusive as there are many dimensions to consider, some of which are conflicting with one another. We present the core concepts relevant to the microeconomic models considered in this thesis as well as for learning in games. With this, the main research problem and question of this thesis will be stated formally. To put these economic models into perspective, let us briefly take a look at the most common points of differentiation.

Cooperative and Non-Cooperative Games

Most work on game theory analyzes competitive or non-cooperative games. In its most extreme, one player's win is proportional or equal to another player's loss. Such games are called zero-sum, and equilibrium computation reduces to well-studied saddle-point problems. At the other end of the spectrum, the objectives are perfectly aligned, and agents effectively have a joint goal and a strong incentive to cooperate.

Strategic and Extensive Form Games

Historically, there has been a strong focus on strategic games, where players choose their complete plans of action before the game starts. The most basic case is a single-shot game, where agents are simultaneously asked to submit their sole action, and the game terminates thereafter by assigning the corresponding utilities. For relatively large and sequential games, it has proven to be more useful to consider games in their extensive form. The extensive form arranges the sequence of possible events in a tree-like structure, and the players consider one action at a time (Osborne and Rubinstein, 1994). Markets modeled as single-shot games are the primary focus of this thesis, and extensions to sequentially conducted markets are only outlined briefly.

Information Structure in Games

In games of complete information, players are fully informed about the characteristics of the game and the players. In contrast, in games with incomplete information, players may have private information that is not available to others. As we will see in section 2.2, auction theory emphasizes games of incomplete information as this tends to be a central assumption for modeling economic interactions.

Game Properties

In addition to the above properties of games, some notable other properties include

- whether the game is discrete or finite in the number of possible actions and outcomes or continuous as in Definition 2.1,
- the type of players involved, i.e., if the set of players is finite (as in Definition 2.1) as opposed to mean-field games, where players are modeled as a continuum of an infinitely large population, and
- whether the game exhibits any symmetries, e.g., whether the players face the same strategic decisions under the same conditions (Cheng et al., 2004).

With a focus on auction games, we will mainly consider non-cooperative, simultaneous move games with continuous action spaces and a finite number of players in this thesis.

Continuous game

Definition 2.1 (Continuous Game)**.** *The triple $\mathcal{G} = (\mathcal{N}, \mathcal{X}, u)$ is called a game, where*

- *$\mathcal{N} = \{1, \dots, N\}$ is the finite set of players with $N \in \mathbb{N}_{\geq 2}$,*
- *$\mathcal{X} = \prod_{i \in \mathcal{N}} \mathcal{X}_i$ is the set of action profiles with $\mathcal{X}_i = \mathbb{R}^{d_i}$ being the action space for player $i \in \mathcal{N}$ and $d_i \in \mathbb{N}_{\geq 1}$, and*
- *$u = (u_1, \dots, u_N)$ with $u_i : \mathcal{X} \to \mathbb{R}$ being the utility function of player i which maps an action profile $x \in \mathcal{X}$ to its associated utility.*

A game is called finite if $\mathcal{X}$ is finite instead. For example, the process of action bucketization transforms a continuous game into a finite one, where some structure is lost in a trade-off for more tractability.

Going forward, we will make use of the shorthand notation $(x_i, x_{\text{-}i}) = (x_1, \ldots, x_N) \in \mathcal{X}$ for actions and analogously for other concepts defined for all agents. This helps to formulate the game from the perspective of player i and combine the opponents under the index -i. In the following, we will extend Definition 2.1 by introducing the concept of incomplete information. Agents then have private information they can base their decisions on.

2.2 Auction Theory

Economists try to understand the structures and dynamics of strategic behavior in different markets. Auction theory, as a subfield of game theory, provides the formal means to model markets. This primarily comprises games of incomplete information (Klemperer, 1999; Krishna, 2009). Applications range from energy and commodity markets, bond issues by public utilities, public-to-private auctions such as spectrum auctions, display advertisements, or e-commerce auctions to the process of procurement in supply chains.

The origins of auction theory date back to work by Vickrey (1961) in the early 1960s, in which he proposed modeling auctions as non-cooperative games of incomplete information and noticed that participants use their private information to their advantage. Agents are assumed to be strategic; that is, they may bid less than they are willing to pay. This involves each bidder receiving a privately observed valuation, conditioned on which the bidder acts on the market. A mechanism determines the allocation of goods and the corresponding transaction prices. Bids are submitted so as to maximize individual expected utility. Especially when the number of participants is relatively small, their strategies highly impact the market dynamics. Mechanism design is the field of study concerned with analyzing and setting up markets that optimize certain desiderata, such as yielding efficient outcomes or maximizing total welfare or seller revenue, which are of interest to the policymakers or the market participants (Bichler, 2017). In contrast, the bidders' main question is, given the mechanism designed by the policymaker, what is their best strategy in order to maximize individual utility? This raises the general need for access to equilibria of strategically acting market participants. As alluded to above, analytical equilibria remain largely unknown because mathematical solution methods require expert knowledge or may even be inapplicable, motivating the need for computational methods.

We will formally introduce a sealed-bid auction as an extension of and in conformity with the above definition of continuous games. We will only present the single-item and single-sided case (i.e., only buyers act strategically in a competition for the good) to minimize the notational clutter. This constitutes the starting point for much of auction theory.

Definition 2.2 (Auction Game)**.** *The quintuple $\mathcal{G} = (\mathcal{N}, \mathcal{V}, \mathcal{A}, f, u)$ is called a first-price sealed-bid single-item auction, where* *Auction game*

- $\mathcal{N} = \{1, \ldots, N\}$ *is the finite set of players or bidders,*

- *$\mathcal{V} = \prod_{i\in\mathcal{N}} \mathcal{V}_i$ is the product space of valuation profiles with $\mathcal{V}_i \subset \mathbb{R}_{\geq 0}$ being the space for bidder $i \in \mathcal{N}$,*
- *$\mathcal{A} = \prod_{i\in\mathcal{N}} \mathcal{A}_i$ is the product space of action or bid profiles with $\mathcal{A}_i \subset \mathbb{R}_{\geq 0}$ being the action space for bidder $i \in \mathcal{N}$,*
- *$f : \mathcal{V} \to \mathbb{R}_{\geq 0}$ is the joint prior valuation density and is assumed to be bounded and atomless (with marginals f_i), and*
- *$u = (u_1, \ldots, u_N)$ with $u_i : \mathcal{V}_i \times \mathcal{A} \to \mathbb{R}$ being the utility function of bidder i which maps a valuation v_i and an action profile b to i's associated utility:*

  $$u_i(v_i, b) = v_i\, x_i(b) - p_i(b) = \begin{cases} v_i - p_i(b) & b_i > \max_{j\neq i} b_j, \\ 0 & else. \end{cases} \tag{2.1}$$

  Here, x_i and p_i define the allocation and first-price payment rule, respectively:

  $$x_i(b) = \begin{cases} 1 & b_i > \max_{j\neq i} b_j, \\ 0 & else, \end{cases} \qquad p_i(b) = \begin{cases} b_i & b_i > \max_{j\neq i} b_j, \\ 0 & else. \end{cases} \tag{2.2}$$

  Ties are broken uniformly at random when needed.

Thus, most of the market is incorporated in the utility function u. It captures the auction *mechanism* that allocates the item to a specific bidder and determines its price. The market is called a *sealed-bid auction* because there is a single bidding round, and bidders can base their decisions only on their observed valuations and not anyone else's bids. And it is called a *first-price auction* because the price corresponds to the highest bid. Hence, we call this a first-price sealed bid (FPSB) auction. This contrasts, for example, second-price sealed bid auctions (SPSB), where the highest bidding bidder wins but pays the second-highest bid. Considering the literature on game theory, this game specifies a continuous *Bayesian game*.

Mechanism

FPSB auction

Bayesian game

Now, one is not only interested in optimal behavior for one particular sample of prior valuations but, more broadly, in an optimal strategy that reasons over the uncertainty of the opponents' behaviors and the bidder's private information. Therefore, we define a bidding strategy for bidder i as $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$ that assigns each valuation one action or bid. We denote the resulting set of strategy profiles by $\Sigma = \prod_{i\in\mathcal{N}} \Sigma_i$ and a strategy profile $\beta \in \Sigma$ by $\beta = (\beta_1, \ldots, \beta_N) = (\beta_i, \beta_{\text{-}i})$. Although large parts of game theory and RL consider *mixed strategies* (that map valuations or observations to a distribution over actions), in auction theory, it is generally sufficient to reduce the space to *pure strategies* (Krishna, 2009), as done here.

Ex-post utility

The utility u_i from Equation 2.1 is commonly referred to as *ex-post utility* because it can only be evaluated at the end of a game, that is, once the valuations have been drawn, bids submitted, and the allocation and payments determined. Based on the ex-post utility, one can define the *interim utility*

Interim utility

$$\bar{u}_i(v_i, b_i, \beta_{\text{-}i}) \; := \; \mathbb{E}_{v_{\text{-}i}\sim f_{\text{-}i}}[u_i(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))] \tag{2.3}$$

as expectation of i's utility, when having valuation v_i and submitting bid b_i, over the opponents' valuations $v_{\text{-}i}$ and their subsequent bids $\beta_{\text{-}i}(v_{\text{-}i})$. Going a step further, with the *ex-ante utility*

Ex-ante utility

$$\tilde{u}_i(\beta) := \mathbb{E}_{v_i \sim f_i}[\bar{u}_i(v_i, \beta_i(v_i), \beta_{\text{-}i})] = \mathbb{E}_{v \sim f}[u_i(v_i, \beta(v))], \tag{2.4}$$

one can characterize utilities at all different stages of the game. The bidders' objective is now to decide on strategies β that maximize their ex-ante utility.

With this, we are set to discuss the main challenges of applying and analyzing learning to bid and equilibrium computation.

1. Auctions are neither fully cooperative nor competitive: Despite the profit of one bidder tending to decrease those of other bidders (e.g., all bidders compete for the same goods), by collectively bidding low, they decrease the prices and increase the winner's profit. This can be considered a form of collusion and exemplifies that auctions are general-sum games.

2. One can simplify the computation of symmetric equilibria by learning a single strategy for all players in a symmetric game, where the identity of an agent does not matter. However, there exist many auctions that consider asymmetries for the participants. These include different demand structures in terms of the valuation priors (e.g., some bidders may be interested in a different set of goods than their competitors) or different utility preferences (e.g., individual attitudes towards risk aversion). We investigate how learning to bid can be applied to symmetric markets in (Bichler et al., 2021, see chapter 4) and to more challenging asymmetric markets in (Bichler et al., 2023, see chapter 5).

3. In mathematical optimization, concavity of the objective function is a common requirement for global convergence. Analogous but more demanding conditions for the convergence of multi-agent reinforcement learning have been established. In auctions, however, the ex-ante utilities are not globally concave even under quite simple parameterizations of the bid functions. This makes the convergence analysis more challenging.

4. More than two parties are typically involved in auction markets. Consequently, the large literature on learning in two-player games and its positive results on convergence are inapplicable.

What is more, we face the following technical challenges in addition to the above general hurdles.

5. One of the main challenges is the discontinuity of a bidder's utility in his or her action at the bid level around the highest opponent bid. Bidding just below the highest opposing bid leads to a utility of zero, whereas a bid just above leads to a high positive utility (assuming the bidder is not bidding above his or her valuation). This issue is thoroughly analyzed in (Kohring et al., 2023, see chapter 7).

6. The equilibrium search space is an infinite-dimensional space of bid function profiles. We will overcome this by parameterizing the strategies by neural networks and thereby reducing the game to a finite-dimensional version as defined in Definition 2.1. This obviously reduces the space of bid functions and excludes less regular strategies. Nevertheless, we provide a bound on the utility error based on the networks' expressiveness (Bichler et al., 2021, see chapter 4, Lemma 3).

7. What is more, as bidders only have partial information on the environment's state, they must act under uncertainty. Their expected ex-ante utilities depend on the opponents' strategies, which makes calculating the expected values infeasible analytically. Therefore, we will approximate the ex-ante utilities via Monte Carlo integration over sampled ex-post utilities.

We will elucidate all of these points in subsequent parts and propose and discuss solution methods.

The auction model of Definition 2.2 can readily be extended to more plausible scenarios. These will only be briefly outlined here and not formally introduced in an attempt to keep this introductory section focused on the main building blocks of learning to bid. For instance, the utility function may also capture preference structures that seem plausible to the application, such as risk aversion, or one may introduce correlations in the bidders' valuations (Milgrom and Weber, 1982). From a practical standpoint, it seems reasonable that bidders competing for the same good or bundle of goods also have similar valuations for it. For notational convenience, the above model assumes risk-neutral bidders and *independent private values* (IPV). Nonetheless, both extensions are considered and discussed in (Bichler et al., 2021, see chapter 4). Other straightforward extensions include auctions of multiple goods, such as multi-unit or more general combinatorial auctions. Then, the valuation $\mathcal{V}_i$ and bid spaces $\mathcal{A}_i$ are of higher dimension, which exacerbates learning the bid functions $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$. Such markets will be analyzed in (Bichler et al., 2023, see chapter 5). Furthermore, the model can be extended to double auctions, where not only bidders are interested in buying goods, but also sellers strategically determine their offers as considered in (Bichler et al., 2022, see chapter 6). We will also take a look at reverse auctions in (Bichler et al., 2023, see chapter 5) that model procurement processes where multiple sellers make offers. This is motivated by applications within industrial supply chains. Other extensions to Definition 2.2 include all-pay auctions (Ewert et al., 2022), contests, or considering budget-constrained bidders.

Independent private values

2.3 Solution Concepts

This section first introduces the necessary solution concepts before algorithmic frameworks for optimizing gameplay are considered. As we have seen, this comes with additional challenges compared to single-agent learning and mathematical optimization.

Nash Equilibrium

The *Nash equilibrium* (NE) is the starting point of most discussions on solution concepts. Informally, no player has the incentive to deviate from his or her strategy unilaterally in an NE. More formally, we state:

Nash equilibrium

Definition 2.3. *Let $\mathcal{G} = (\mathcal{N}, \mathcal{X}, u)$ be a game and $\varepsilon \geq 0$. The action profile $x^\star \in \mathcal{X}$ is said to be a local ε-Nash equilibrium of $\mathcal{G}$ if there exist open sets $U_i \subseteq \mathcal{X}_i$ such that $x_i^\star \in U_i$ and*

$$u_i(x_i^\star, x_{\text{-}i}^\star) + \varepsilon \;\geq\; u_i(x_i, x_{\text{-}i}^\star) \tag{2.5}$$

for all $x_i \in U_i$ and all $i \in \mathcal{N}$. It is called global if $U_i = \mathcal{X}_i$ for all i and exact or simply NE if $\varepsilon = 0$.

For fixed opponent actions $x_{\text{-}i}$, we call any action x_i that achieves maximal utility a *best response*. With this, an NE can be interpreted as all agents best responding to their opponents. The definition is straightforwardly extended to Bayesian games, such as the auction from Definition 2.2, by considering the expected utility of the players across the prior distributions:

Best response

$$\tilde{u}_i(\beta_i^\star, \beta_{\text{-}i}^\star) + \varepsilon \;\geq\; \tilde{u}_i(\beta_i, \beta_{\text{-}i}^\star) \tag{2.6}$$

for all strategies $\beta_i \in \Sigma_i$ and all $i \in \mathcal{N}$. Such a strategy profile $\beta^\star$ is called an (ex-ante) *ε-Bayes-Nash equilibrium* (ε-BNE). Similarly, in an ε-BNE, no bidder is able to deviate and gain ε or more expected utility.

ε-Bayes-Nash equilibrium

Attempts to analytically derive the BNE strategies usually involve stating the distribution of the highest opponent valuation, assuming symmetric strategies, and then solving the resulting differential equation for the inverse bid function.

Example 2.1. *Consider the single-item FPSB auction of Definition 2.2 with N bidders. Let the prior valuations be independently and uniformly distributed on the unit interval. Following the above procedure, one verifies $\beta_i(v_i) = \frac{N-1}{N} v_i$ to be a symmetric BNE in this auction (Krishna, 2009, Example 2.1). Intuitively, the prices rise with increasing levels of competition.*

The main goal of this thesis is the computation of such equilibrium strategies or approximations thereof, that map each valuation to a bid leading to the highest expected utility such that no bidder can deviate profitably from his or her strategy. Considering the auction from Definition 2.2, we employ a parameterized bid function for bidder i via

$$\beta_i(v_i) \;=\; \pi_{\theta_i}(v_i), \tag{2.7}$$

where $\theta_i \in \Theta = \mathbb{R}^d$ are the parameters of the neural network π. Let $\theta = (\theta_1, \ldots, \theta_N)$ and $\pi_\theta = (\pi_{\theta_1}, \ldots, \pi_{\theta_N})$. Importantly, this reduces the infeasible search in the infinite-dimensional space of bid functions to a finite-dimensional search for optimal parameters. However, depending on the network's architecture, some equilibrium strategy profiles may be excluded.

Assuming the availability of the expected utilities allows us to interpret the parameterized auction in its ex-ante form as a complete information continuous game as in Definition 2.1. We call

$$\mathcal{G}_{\text{proxy}} = (\mathcal{N}, \Theta, \tilde{u}) \tag{2.8}$$

Proxy game

the *proxy game* of the auction $\mathcal{G}$ with utilities $\tilde{u}_i(\pi_\theta)$ for all i in accordance with (Bichler et al., 2021, chapter 4, Definition 2). With this perspective, we are looking for a set of parameters $\theta^\star$ that satisfies the Nash condition of Definition 2.3 for the ex-ante utilities $\tilde{u}$. Furthermore, the ex-ante utilities are usually more regular (particularly differentiable), unlike the discontinuous ex-post utilities. Loosely speaking, when taking the expected utility over all the possible valuations and respective bids by the opponents, the discontinuity of winning or losing an item is averaged out. In practice, these expected utilities are not available and, instead, must be approximated via sampling. We will come back to this observation in later sections on learning in games.

Correlated Equilibrium

Correlated equilibrium

Coarse correlated equilibrium

A relaxation of NE is given by *correlated equilibrium* (CE). These are motivated by some simple learning methods leading to them. Informally, the players choose their actions following some public correlation mechanism in a CE. Importantly, CE have a lower computational complexity than NE.[1] Additionally, the set of *coarse correlated equilibria* (CCE) has been introduced. CCE can be considered the weakest solution concept and may contain undesirable, strictly dominated strategies. We have the inclusions NE $\subset$ CE $\subset$ CCE. CCE are motivated by being the consequence of many simple learning methods, where players adapt their strategies according to how much they regret historical actions.

A strategy is called no-regret if it leads to a utility at least as high as any fixed strategy in retrospect. Hence, playing according to a CCE can be understood as being no-regret (Cesa-Bianchi and Lugosi, 2006). Furthermore, in cases where the (C)CE is unique, it must coincide with an NE. Consequently, any methods converging to (C)CE will find an NE under these circumstances. This is an important observation that may explain the convergence of learning to bid in some settings.

Existence and Uniqueness

Nash (1951) proved that every (mixed strategy extension of a) finite game admits an NE. Extending this result, Debreu (1952) showed the existence of NE in continuous games with compact action spaces. And Athey (2001) proved the existence of pure strategy BNE for multiple auction types of incomplete information by introducing and verifying the so-called single-crossing condition. Essentially, this requires all participants to have non-decreasing strategies in their valuations. Assuming concavity of the utility functions, Rosen (1965) gave a sufficient condition for Nash equilibrium existence and uniqueness. Ui (2008) generalized these results in smooth games to CE uniqueness. Obviously, the results only apply to

[1] These complexity results were derived for finding mixed strategy equilibria in finite games.

a limited set of games, and we know of some auction games with multiple equilibria. More substantial strategy space restrictions are needed for equilibrium uniqueness, even in symmetric single-item SPSB auctions with two bidders. Besides the equilibrium of all bidders truthfully revealing their valuations, there exists a second "ill-behaved" BNE with constant strategies, where one bidder bids zero and the other bids one regardless of their valuations.

Solving for the Nash Equilibrium

For relatively simple models of independent private valuations, such as the symmetric FPSB auctions from Example 2.1 (Holt Jr, 1980; Riley and Samuelson, 1981) or bilateral bargaining (Leininger et al., 1989), closed-form solutions have been derived by writing the optimality condition of Equation 2.6 for the inverse bid function as an *ordinary differential equation* (ODE). Marshall et al. (1994) and Bajari (2001) use numerical algorithms for computing optimal strategies in IPV FPSB single-item auctions, and Hubbard and Paarsch (2014) considered auctions with asymmetries and risk-aversion. Campo et al. (2003) extended this beyond the IPV model to affiliated values for single-item auctions. The convergence properties of these approaches are unknown, and they have been criticized for instability (Fibich and Gavish, 2011). Further, it remains open if these methods can successfully be applied to more general markets, such as combinatorial auctions.

Chapter 3

Learning in Games: Theory and Algorithms

By pursuing optimal decision-making in dynamic environments, we have established criteria for desirable solutions in the last section. At the core of learning in games is the question of the consequences of autonomous agents simultaneously learning in a common environment. As such, the game itself evolves over time, making independent optimization for the agents a complicated affair as agents can only adapt their own strategies, yet their utilities depend on all agents' strategies. What is more, there exists a hierarchy of solution concepts which translates to a hierarchy of computational complexity. This stands in contrast to single-agent learning or mathematical optimization, where the convergence of gradient ascent to local optima is essentially guaranteed. Unfortunately, the famous Nash equilibrium (Definition 2.3), as the arguably most desirable solution concept, has been shown to be of high computational complexity. Specifically, it is PPAD-complete already for two-agent finite games (Daskalakis et al., 2009). Such complexity results extend to games of incomplete information. To be precise, it was shown that finding a BNE or best response in simultaneous single-item second-price auctions, in which bidders have non-trivial combinatorial valuations, is PP-hard (Cai and Papadimitriou, 2014). The authors also show that the computation of a Bayesian extension of the weaker notion of a CE is NP-hard. Currently, two common conclusions on these complexity results are discussed. Firstly, it may suggest the usage of a specialized algorithm for games of a computationally manageable subclass. Secondly, some researchers have suggested that the dynamics are possibly interesting by themselves. According to them, one should focus more on the process of repeated interaction itself instead of its possible terminal states, such as static equilibria (Papadimitriou and Piliouras, 2019). A good point of reference for this debate and learning in games more generally is the book by Fudenberg et al. (1998).

Motivated by the empirical success of deep learning and single-agent RL, there has been a resurgence of interest in *multi-agent reinforcement learning* (MARL) (Zhang et al., 2021). RL-based approaches have outperformed humans in some video games (Mnih et al., 2015) and board games with prodigious state- and action-spaces via a combination of deep learning and Monte Carlo tree search (Silver et al., 2016). There has been no shortage of newly proposed algorithms from the domains of machine learning, operations research, and multi-agent systems, mainly driven by the diverse landscape of games and their inherently different properties. Yet, their theory is not well understood. The main challenges in MARL include the environment becoming non-stationary due to the opponents' changing strategies and the exponentially increasing joint action space in the number of players. MARL is usually subdivided into value- and policy-based approaches. The former methods first compute the players' values for particular states of the environment, based on which they try to move to more promising states (with higher values). The latter methods directly optimize for a strategy that maps observations to the best actions, which is generally more robust but less sample efficient. Both can be extended to leverage function approximation for large and continuous state and action spaces. We will introduce and discuss the suitability of policy-based methods to economic applications below. Most importantly, any algorithm for learning in auctions must handle incomplete information and continuous action spaces.

Note that this chapter does not claim to be a general introduction to the topic of learning in games. Instead, the concepts are presented from the perspective of an economic theorist, focused on learning to bid in the auction game from Definition 2.2 and extensions thereof.

3.1 Algorithms

The first attempts at algorithmic learning in games date back to work by Brown (1951) and the solver by Lemke and Howson (1964) for finite bimatrix games. These and other classic approaches are mainly concerned with finite and full information games and do not scale well to large or continuous action spaces. For example, the computation of exact best responses becomes insurmountable, requiring global maximization as a subroutine at each iteration. These limitations motivate the focus on gradient-based methods utilizing function approximation as simple and computationally attractive alternatives. These are at the center of this dissertation's analysis. This section starts by revisiting the classic approaches and highlights whenever an algorithm has been applied or explicitly extended for learning to bid.

We highly recommend the habilitation by Mertikopoulos (2019) and the references therein for an overview of the fundamentals of multi-agent learning and Shalev-Shwartz (2007) for a concise look at the online learning literature.

3.1.1 Classic Learning Dynamics

The direct application of classic tabular methods to discretized auctions quickly becomes infeasible due to the exponential growth of the valuation and action space for finer resolutions in multi-agent games. However, they comprise the conceptual core of much of the learning in games literature. Thus, the following overview puts the algorithms considered in this dissertation into perspective and enables assessing when alternative approaches are viable. Thereby, it demarcates the contribution of this thesis from existing work.

Fictitious play

With *fictitious play*, Brown (1951) introduced one of the most basic procedures for adaptive behavior in finite games. Here, players best respond to the historical frequency of their opponents' actions. Some convergence results are known, such as for two-player or zero-sum games (Robinson, 1951; Miyasawa, 1961), and multiple variants, such as smooth fictitious play, have been introduced over the decades (Hofbauer and Sandholm, 2002).

> **Application 3.1.** *Rabinovich et al. (2009, 2013) propose a generalization of fictitious play that learns BNE in auctions with continuous valuations and discrete actions. They do so by utilizing the anonymous nature of auctions, in the sense that outcomes are independent of the identity of the players and only depend on the bids. Consequently, the finite number of opponents can be interpreted as a continuum of anonymous players representing the same amount of competition. In contrast, Gemp et al. (2022) are in the pursuit of automated mechanism design. They employ bidders with fictitious play in an inner loop and use a gradient-based method in an outer loop for the objective of designing a seller-optimal all-pay auction.*

Best response dynamics

In contrast to fictitious play, where equal weight is put on all historical actions, *best response dynamics* only consider the most recent actions. That is, agents respond with their pure best response to their opponents' actions in the last round of play.

> **Application 3.2.** *There have been multiple approaches to applying best response dynamics to learning to bid. Reeves and Wellman (2004) propose an iterative best response approach that learns piecewise-linear pure strategies in two-bidder auctions. Bosshard et al. (2020) learn in IPV combinatorial auctions via an iterated best response procedure, and they are additionally able to verify closeness to BNE. Dütting and Kesselheim (2022) apply best response dynamics to combinatorial auctions with item bidding (i.e., multiple simultaneous single-item auctions) on restricted prior valuations. Although these dynamics are slow to converge or may even fail to do so, they provide social welfare guarantees.*

Follow-the-regularized-leader

In a similar spirit, under the *follow-the-regularized-leader* (FTRL) procedure, players select actions that are optimal in hindsight while considering a regularization term that prevents too drastic updates (as they may occur under best response dynamics). Under standard concavity and Lipschitz assumptions on the utility, it is shown to achieve no-regret, which implies convergence to CCE (Shalev-Shwartz, 2007; Flokas et al., 2020). The main drawbacks of this class of algorithms are the requirement of full access to the objective functions and the need for solving an optimization problem in each iteration.

> **Application 3.3.** *Daskalakis and Syrgkanis (2016) propose and analyze a variant of the follow-the-leader scheme in simultaneous auctions with fixed valuations and risk-neutral bidders. They provide theoretical bounds for market efficiency and the gap to optimal welfare. Balseiro and Gur (2019) learn in repeated auctions via so-called adaptive pacing strategies. By limiting the bidders' strategies to clipped linear functions, where the clipping is conducted to satisfy the budget constraints, they restrict the game and strategic possibilities. Still, they can establish a convergence guarantee*

> *and corresponding bounds on the utility loss.*

Multiplicative weights

Let us also consider the *multiplicative weights* method (Auer et al., 1995; Arora et al., 2012) for finite games. Weights on the past utilities of all actions are maintained, and future actions are simply chosen randomly with probability proportional to these weights. At the end of each iteration, they are updated based on the utility of the associated action. Over time, the algorithm tends to assign higher weights to actions of higher utility. This allows the establishment of tighter regret bounds compared to gradient ascent in some situations (Shalev-Shwartz, 2007).

Dual averaging

Another class of algorithms is given by the *dual averaging* method. Unlike FTRL, it does not require full information feedback but instead is based on first-order gradient feedback. The gradients are accumulated in the dual space (instead of only considering the most recent gradient evaluation) and then projected back to the feasible region (Nesterov, 2009; Mertikopoulos, 2019). We will conduct an in-depth comparison of using gradient ascent in the parameter space of neural networks representing the bid functions to the application of dual averaging in a fully discretized auction game in (Bichler et al., 2022, see chapter 6).

> **Application 3.4.** *Kolumbus and Nisan (2022) use the multiplicative weights algorithm for repeated auctions with finite valuations and actions. They provide a convergence analysis in the resulting finite game. Feng et al. (2021) study the convergence of an extension of the multiplicative weights algorithm in discretized repeated auctions with unknown valuation distributions. Their choice of model is motivated by online advertisement auctions. They also propose an extension to deep Q-learning for better scalability to finer discretizations.*

> **Application 3.5.** *Online advertising auctions are a key motivation for learning automated bidding strategies. Weed et al. (2016) apply a form of online learning to such repeated second-price auctions. They provide regret bounds, but their analysis is limited to the IPV model and bidders with an unlimited budget. Further, they show that discretizing the action space in auctions leads to regret growing at least linearly.*

Regret matching

Counterfactual regret minimization

At last, let us mention a modern approach to learning in finite games of incomplete information. Hart and Mas-Colell (2000) propose *regret matching* where players try to reach equilibrium by counting regrets over the history of play and adapting future play inversely proportional to the regrets. Zinkevich et al. (2007) extend this to games of sequential play by introducing *counterfactual regret minimization* (CFR). CFR achieves an average overall regret that is linear in the number of possible observations. It is at the core of current state-of-the-art poker bots (Brown and Sandholm, 2018), but considerable engineering efforts are required to scale this computationally and memory-demanding tabular method to the full game size with a sufficiently fine discretization.

3.1.2 Simultaneous Gradient Ascent

Gradient-based methods are the algorithm of choice in statistical learning theory due to their low computational costs per iteration, albeit their relatively slow convergence rate. Under mild conditions, one can find the global optimum of concave functions and local optima of non-concave functions (Lee et al., 2016). At each iteration of gradient ascent, a step is taken in the direction of the steepest ascent of the payoff by slightly adjusting the actions (in the dual space). When the domain is restricted, a projection step back to the feasible region follows (sometimes also called mirroring).

For an action profile $x \in \mathcal{X}$ in the game from Definition 2.1, let

$$\nabla u(x) \;=\; (\nabla_{x_1} u_1(x), \ldots, \nabla_{x_N} u_N(x)) \tag{3.1}$$

be the *simultaneous gradient* of the utilities, where $\nabla_{x_i} u_i$ denotes the derivative of u_i with respect to x_i. Now in *simultaneous gradient ascent*, agents independently and myopically follow the gradient of their respective utility to update their actions.[1] In the context of learning in games, gradient ascent can be interpreted as a regularized best response with momentum. Then, in analogy to single-agent learning, necessary and sufficient conditions for local optimality have been established:

Simultaneous gradient ascent

Proposition 3.1 (Ratliff et al. (2016), Proposition 1)**.** *Let $u_i \in C^2(\mathcal{X}, \mathbb{R})$ for all $i \in \mathcal{N}$. If $x^\star \in \mathcal{X}$ is a local NE, then $\nabla u(x^\star) = 0$ and the second partial derivative of u_i with respect to x_i is negative semi-definite for all $i \in \mathcal{N}$.*

Convergence to local NE is all we can hope for in general non-concave games with gradient-based approaches. Proposition 3.1 seems like a straightforward extension of the single-agent case. However, the question of how uncoupled multi-agent dynamics behave is much more involved. Here, uncoupled dynamics are understood in these sense that agents only have access to their own utilities and the partial derivatives with respect to their own actions.

In the case of learning to bid, recall that some auction games are known to have multiple equilibria and that modeling the players' strategies by neural networks is usually performed with an over-parameterization, i.e., there may be multiple parameter configurations for the same strategy. Thus, even if convergence were to be guaranteed, which equilibrium would be reached? To answer this, a setting with multiple known equilibria will be analyzed in (Bichler et al., 2023, see chapter 5, Subsection 6.1.1). The discussion on the convergence of simultaneous gradient ascent is continued in section 3.2.

Remark 3.1. *Letcher et al. (2019a) establish a result similar to that of Proposition 3.1.*

Remark 3.2. *There are two dimensions to consider for the type of feedback or oracle available to a learning scheme.*

[1] Following the literature on online (convex) optimization, some authors call this procedure *online gradient ascent* as discussed in (Zinkevich, 2003).

1. *Occasionally, the payoffs of a game cannot be accessed exactly, but there may be a certain amount of noise in the measurements. Most commonly, this is due to the stochasticity of the environment or the opponents. In Bayesian games such as auctions, one takes a sample of finite size from the valuation distribution of the agents. Hence, noticeable noise will be in the players' utility estimates, especially for small sample sizes.*

2. *We consider the order of feedback. Being able to evaluate the payoffs is referred to as zeroth-order feedback, whereas having access to its gradient is considered first-order feedback. Constructing a gradient estimate based on zeroth-order feedback is less desirable in terms of computational efficiency and usually requires multiple payoff evaluations. The next section investigates different gradient estimation techniques.*

Remark 3.3. *Let us also mention the continuous-time viewpoint on learning dynamics. The continuous dynamics can be considered the limiting state of iterative updating procedures with an infinitesimal learning rate. This makes possible the application of some techniques from evolutionary game theory and allows for easier analytical analysis. In particular, some regret bounds can be tightened in the continuous time framework (Kwon and Mertikopoulos, 2017).*

To apply gradient ascent (or some other updating rule) to learning to bid, either the auctions must be discretized, or the strategies must be parameterized via function approximation. The following section considers the latter option. This approach comes in particularly handy for large or continuous action spaces as it reduces the infinite-dimensional search space (finding optimal bid functions) to a finite-dimensional one (finding optimal sets of parameters). The most important implementations will also be highlighted.

3.1.3 Policy Optimization

A policy defines a function that maps game observations to actions (or distributions thereof). Typically, policies are parameterized by neural networks. The main hurdle for practical usage is the necessity of a highly efficient simulator capable of supplying enough data for the sample-inefficient training of neural networks.

Let us consider the parameterized auction game $\mathcal{G}_{\text{proxy}}$ from Equation 2.8. The exact gradient update step for bidder i at iteration t takes the form

$$\theta_i^t \;=\; \theta_i^{t-1} + \eta \cdot \nabla_{\theta_i^{t-1}} \tilde{u}_i\big(\pi_{\theta^{t-1}}\big), \tag{3.2}$$

for a step size $\eta > 0$. Unfortunately, neither can we calculate the exact ex-ante utilities nor their gradients. We can only approximate the utilities via sampling the ex-post utilities, which gives noisy unbiased estimates of the ex-ante utilities.

With deep learning theory still in its infancy, the theoretical analysis of deep MARL is very limited. Coming from a supervised or offline learning problem formulation, there are two key challenges regarding the training of neural networks. Firstly, the environment is highly dynamic, and once the policy is updated, subsequently playing according to this changed policy will result in a different data distribution. This is even more severe when there are multiple learners involved who are continuously updating their strategies. This invalidates the common assumption of independent and identically distributed samples for stochastic gradient ascent. That is why many RL algorithms use replay buffers that effectively alleviate the issues of correlation between sampled game outcomes. Secondly, most games or RL applications do not provide a differentiable loss or utility function. Informally, they are considered black boxes that take in a policy and return a utility without revealing a functional structure. The gradient estimate is influenced by policy updates that change the state and action distributions and the resulting utilities. More specifically, in auctions, the ex-post utility is discontinuous in the agent's bid. At a bid magnitude around the highest opponent bid, bidding just below that value results in losing the auction and a utility of zero, whereas bidding just above results in a positive utility, assuming that bidder i is not overbidding. This prevents the application of backpropagation for first-order gradient estimation via Monte Carlo sampling. Let us now consider different ways of estimating the gradients, where this discussion will be continued.

REINFORCE

REINFORCE
Proximal policy optimization

The REINFORCE algorithm (Williams, 1992) is central to modern RL and lays the conceptional foundation for actor-critic methods such as the famous *proximal policy optimization* (PPO), which was introduced by Schulman et al. (2017) and also found widespread adoption in continuous control and robotics.

At its core, the *policy gradient theorem* is applied, which allows rewriting the policy's gradient in terms of the action distribution instead of the inaccessible gradient of the utility function via the *log derivate trick*. For that, one is necessarily interested in learning mixed strategies. We will write $\pi_{\theta_i}(\,\cdot\,|v_i)$ for the *probability density function* of bidder i's bids given the valuation v_i. This conditional distribution is assumed to be Gaussian, but any type of absolute continuous distribution can be utilized. The opponents stick to playing pure strategies $\beta_{\text{-}i}$ in the following for ease of notation. Following the derivations of Mohamed et al. (2020, Section 4) adapted to learning to bid, the policy gradient is given as

$$\nabla_{\theta_i}\tilde{u}_i(\pi_{\theta_i},\beta_{\text{-}i}) = \nabla_{\theta_i}\mathbb{E}_{v\sim f}\mathbb{E}_{b_i\sim\pi_{\theta_i}(\,\cdot\,|v_i)}\left[u_i(v_i,b_i,\beta_{\text{-}i}(v_{\text{-}i}))\right] \tag{3.3}$$

$$= \nabla_{\theta_i}\int_{\mathcal{V}} f(v)\int_{\mathcal{A}_i}\pi_{\theta_i}(b_i|v_i)\cdot u_i(v_i,b_i,\beta_{\text{-}i}(v_{\text{-}i}))\,db_i\,dv. \tag{3.4}$$

We can now apply the Leibniz rule and interchange integration and differentiation. Importantly, this is valid even for the discontinuous utilities in auctions (Flanders, 1973). u_i was discontinuous in b_i when considering pure strategies, whereas now, u_i is independent of θ_i and the product of π_{θ_i} and u_i is continuous in θ_i after the interchange (Equation 3.5). Formally, we can write

$$\nabla_{\theta_i}\int_{\mathcal{V}} f(v)\int_{\mathcal{A}_i}\pi_{\theta_i}(b_i|v_i)\cdot u_i(v_i,b_i,\beta_{\text{-}i}(v_{\text{-}i}))\,db_i\,dv \tag{3.4}$$

$$= \int_{\mathcal{V}} f(v)\int_{\mathcal{A}_i}\nabla_{\theta_i}\,\pi_{\theta_i}(b_i|v_i)\cdot u_i(v_i,b_i,\beta_{\text{-}i}(v_{\text{-}i}))\,db_i\,dv \tag{3.5}$$

$$= \mathbb{E}_{v\sim f}\mathbb{E}_{b_i\sim\pi_{\theta_i}(\,\cdot\,|v_i)}\Big[\nabla_{\theta_i}\log\pi_{\theta_i}(b_i|v_i)\cdot u_i(v_i,b_i,\beta_{\text{-}i}(v_{\text{-}i}))\Big], \tag{3.6}$$

where the log derivate trick, $\nabla_{\theta_i}\pi_{\theta_i}(b_i|v_i) = \pi_{\theta_i}(b_i|v_i)\nabla_{\theta_i}\log\pi_{\theta_i}(b_i|v_i)$, is applied in the last step to regain an expected value. The resulting expected gradient can be approximated by sampling from the prior distribution and utilizing standard backpropagation. Concluding, the evaluation of the utility gradient can be circumvented by transferring the "gradient flow" through the action probabilities instead and only evaluating the utility itself.

> **Application 3.6.** *Tan et al. (2022) employ an actor-critic RL method and empirically validate the approach in the continuous valuation auction setting.*

The REINFORCE estimate falls in the class of zeroth-order methods because it only relies on evaluating the utility and not on any of its derivatives. The estimate can be of high variance, especially once the policy collapses to an almost pure strategy, i.e., when it assigns most of the probability mass to a single approximate best response. This is frequently the case in games considered in auction theory because they are known to have pure strategy BNE. Thus, the variances of the learned action distributions converge to zero, which results in ever larger gradient magnitudes. This shortcoming renders the REINFORCE algorithm not to be the estimator of choice. A good reference on the details of this estimator and a comparison to alternative approaches provides the survey paper by Mohamed et al. (2020).

Remark 3.4. REINFORCE *can be proven to converge locally under standard stochastic approximation conditions for offline or single-agent learning (Bhandari and Russo, 2019). They also provide conditions for preventing convergence to suboptimal policies. Giannou et al. (2022) prove a convergence rate of $\mathcal{O}(^1/\sqrt{n})$ for the* REINFORCE *algorithm for finite games under appropriate step sizes.*

Deep deterministic policy gradient

Alternatively, *deep deterministic policy gradient* (DDPG) (Silver et al., 2014) is an approach capable of learning pure strategies directly. This is possible by concurrently learning the Q-function (mapping pairs of observations and actions to the player's estimated utility) and a policy. As an off-policy method, it tends to use data more effectively than on-policy methods, such as REINFORCE or PPO, but the usage of two networks per learner adds a layer of complexity to the architecture and training design.

> **Application 3.7.** *Jin et al. (2018) consider online advertising auctions and apply a method based on a multi-agent version of DDPG.*

Evolution Strategies

Evolution strategies

An alternative zeroth-order approach for estimating the policy gradient is based on *evolution strategies* (ES). These were utilized by Salimans et al. (2017) in the context of RL and originally proposed by Spall (1992). They give an asymptotically unbiased estimator, even when the utility gradient is inaccessible, or the utilities are discontinuous, and tend to circumvent the problem of high variance that the REINFORCE algorithm suffers from. Hence, ES are used in our NPGA algorithm (Bichler et al., 2021, see chapter 4).

ES sample nearby parameter configurations of the network such that the local utility surface can be approximated in a manner similar to finite difference approximation. For a small $\sigma > 0$, consider a pure strategy π_{θ_i} and

$$\nabla_{\theta_i} \tilde{u}_i(\pi_{\theta_i}, \beta_{\text{-}i}) = \nabla_{\theta_i} \mathbb{E}_{v\sim f}\Big[u_i(v_i, \pi_{\theta_i}(v_i), \beta_{\text{-}i}(v_{\text{-}i}))\Big] \tag{3.7}$$

$$\approx \nabla_{\theta_i} \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)} \mathbb{E}_{v\sim f}\Big[u_i(v_i, \pi_{\theta_i+\sigma\varepsilon}(v_i), \beta_{\text{-}i}(v_{\text{-}i}))\Big] \tag{3.8}$$

$$= \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)} \mathbb{E}_{v\sim f}\Big[\frac{\varepsilon}{\sigma} u_i(v_i, \pi_{\theta_i+\sigma\varepsilon}(v_i), \beta_{\text{-}i}(v_{\text{-}i}))\Big]. \tag{3.9}$$

As this term directly approximates the ex-ante utilities, which can be assumed sufficiently well-behaved and continuous (Bichler et al., 2022, see chapter 6, Assumption 1), one again overcomes the issue of discontinuous ex-post utilities. In practice, Equation 3.9 can be estimated via the utility-weighted parameter configurations where the prior is sampled and the expected utility approximated for each set of parameters. The additional sampling of parameters and resulting computational costs are one of the main disadvantages of this approach.

> **Application 3.8.** *Li and Wellman (2021) use a similar gradient estimation technique to the one used by NPGA. They empirically evaluate the algorithm in two simultaneous sealed-bid auctions but do not make any theoretical considerations. Noti and Syrgkanis (2021) employ simultaneous gradient ascent to learn in repeated sponsored search auctions and provide some empirical results.*

First-Order Policy Gradient

If available, first-order gradient estimates are often favorable compared with zeroth-order estimates due to lower variances (Ghadimi and Lan, 2013; Suh et al., 2022). The wide use of zeroth-order methods, such as REINFORCE and PPO, is mainly due to the fact that RL environments are given as black boxes without explicit access to gradients. In auction games, the discrete allocation of goods additionally makes the ex-post utility function discontinuous. As a result, the first-order Monte Carlo gradient estimate is inapplicable, as noted in (Bichler et al., 2021). Therefore, we propose a surrogate market in (Kohring et al., 2023) where allocations (and corresponding payments) are smoothed to reestablish the continuity of the ex-post utility. For single-item auctions, we establish:

Theorem 3.1 (Kohring et al. (2023), informal Theorem 4.2)**.** *The first-order estimate of the interim utility's gradient under the smoothed version of the auction from Definition 2.2 with utilities u^{SM} is unbiased under some regularity conditions. That is,*

$$\nabla_{\theta_i} \overline{u}_i^{SM}(v_i, b_i, \beta_{\text{-}i}) \;=\; \mathbb{E}_{v_{\text{-}i}\sim f_{\text{-}i}}\Big[\nabla_{\theta_i} u_i^{SM}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))\Big], \tag{3.10}$$

for all $i \in \mathcal{N}$, $v_i \in \mathcal{V}_i$, and $b_i \in \mathcal{A}_i$.

Furthermore, we verify this change to only introduce a bounded bias on the underlying game dynamics by proving that any approximate equilibrium in the surrogate auction also constitutes one in the original auction (Kohring et al., 2023, Theorem 4.6). With additional theoretical and empirical bounds on the sample variance (which increases with the smoothing strength), we conclude this approach to be superior to previous methods. We showed a significant improvement to NPGA and REINFORCE in performance and computational costs in variously sized markets. However, it remains open if and how this technique can be extended beyond independent single-item auctions to general combinatorial auctions where bids on bundles of items are submitted.

3.1.4 Advanced Policy-Based Methods

Neural fictitious self-play

Policy space response oracles

Regret policy gradient

This section briefly outlines non-gradient-based and more advanced updating procedures. Some tabular methods have also been extended to incorporate parameterized policies. Such hybrid methods enable alternative updating schemes instead of greedily following the direction of locally higher utility as standard policy gradient methods do. Noteworthy is *neural fictitious self-play* (Heinrich and Silver, 2016), which approximates fictitious play by learning a best response against the average of past policies. Other approaches include *policy space response oracles* (Lanctot et al., 2017), which keeps a population of policies and iteratively adds approximate best responses for each player as well as *regret policy gradient* (Srinivasan et al., 2018), which makes a connection between CFR and policy gradients, where the gradient-update rule is adjusted according to regret matching. However, no guarantees on its convergence are available yet.

Swarm Optimization

Particle swarm optimization

Particle swarm optimization (PSO) is a heuristic that is not gradient-based and was introduced in the 1990s by Kennedy and Eberhart (1995). It maintains a fixed-sized set of solution candidates throughout the optimization process. All candidates (so-called particles) adjust their parameters based on their current utilities, momentum, and information exchange shared by other particles. By maintaining multiple candidates, it is considered more capable of escaping suboptimal local optima; however, a good theoretical understanding of the swarm dynamics is still lacking, even in the single-agent case.

In our MARL application, each agent is represented by a separate swarm of particles, and they repeatedly interact in self-play. In contrast to PSO, NPGA only considers a single candidate, and in each iteration, a fixed number of parameter samples is evaluated in close proximity to the current candidate to estimate the gradient. Therefore, a comparable number of objective function evaluations is needed. We found the performance and the computational needs of PSO to be competitive with NPGA's gradient-based learning (Kohring et al., 2022).

Coupled Dynamics

Learning with opponent-learning awareness

Learning with opponent-learning awareness (LOLA) was proposed by Foerster et al. (2018). As the name suggests, each agent tries to anticipate the learning of the other agents. Its update rule includes a second-order correction term that accounts for anticipated policy updates by the opposing agents. However, LOLA may change the stationary points of the original dynamics, which inspired a new iteration of the algorithm (Letcher et al., 2019b). Also, *symplectic gradient adjustment* (SGA) has been introduced by Letcher et al. (2019a). They decompose the utility function into its well-behaved potential and so-called Hamiltonian part and add a correction term to the updating rule that is based on the Hessians. The provided convergence guarantees of this divide-and-conquer approach still require assumptions on the structure of the utility functions.

Symplectic gradient adjustment

This dissertation does not further investigate coupled approaches due to the empirical success of uncoupled dynamics such as NPGA or PSO. In a student project, we found LOLA applicable to learning in auctions. Yet, the additional computational burden of computing second-order terms tends to be not worth the effort (Liu, 2021).

3.2 Convergence to Equilibrium

It is well-established that unconditioned global Nash convergence is out of reach. Results from single-agent learning do not generalize, and multi-agent learning trajectories may end up in limit cycles or chaos (Sanders et al., 2018; Hsieh et al., 2021). A simple example illustrates the failure of uncoupled gradient dynamics:

Example 3.1. *Consider the two-player bilinear continuous game with objectives* $u_1 : \mathbb{R}^2 \to \mathbb{R}$, $(x_1, x_2) \mapsto x_1 x_2$ *and* $u_2 = -u_1$. *Following the individual gradient components leads to cyclic orbits and prevents convergence to the unique critical point (NE) located at the origin.*

This phenomenon results from the misalignment of utilities and has no counterpart in single-agent learning or optimization. Mertikopoulos et al. (2019) have rigorously analyzed this behavior in zero-sum games by considering recurrences in the gradient orbits and leveraging results from the literature on *dynamical systems*. Hart and Mas-Colell (2003) provide an impossibility result and argue that non-convergence is intrinsic to learning being uncoupled. So, as providing global convergence guarantees for general-sum games remains challenging or even impossible in some games, the literature has considered, on the one hand, more restricted classes of games and, on the other hand, weaker notions of convergence.

Firstly, it is well-known that algorithms with the no-regret property lead to the weaker equilibrium notion of CCE when considering convergence in empirical frequencies (Freund and Schapire, 1999; Hartline et al., 2015; Cesa-Bianchi and Lugosi, 2006). Considering the empirical frequencies refers to taking the average strategy over all historic strategies instead of considering the actual strategy played in the last iteration.

Secondly, for special families of games, such as two-player zero-sum games or *potential games* (Monderer and Shapley, 1996), there exist uncoupled dynamics that do converge to NE. The simultaneous gradient constitutes a conservative vector field in potential games, effectively reducing the problem to finding the maximum of a single function. For example, Hofbauer and Sandholm (2002) show global convergence of a variant of fictitious play in zero-sum and potential games. Notwithstanding, these are relatively strong restrictions and hard to verify in continuous games with infinite-dimensional strategy spaces, such as auctions. The notion of a game being potential can be relaxed in some sense by the concepts of *monotonicity* (Rosen, 1965) and by *variational stability* (Mertikopoulos and Zhou, 2019). The latter of these essentially ensures that all individual utility gradients point in the direction of the equilibrium. Hence, one is guaranteed to move toward equilibrium when following the gradients with an appropriate step size.

Remark 3.5. *Some convergence results have been extended to settings of stochastic or noisy feedback. This usually involves assuming zero mean and finite variance errors; see (Mertikopoulos and Zhou, 2019, Theorem 4.7), (Chasnov et al., 2020, Section 4), and (Mazumdar et al., 2020, Theorem 4.3). This is important for our sampling-based approaches to auction games, where ex-ante utilities are only approximated.*

Remark 3.6. *Following the alternative methodology of considering coupled dynamics, as Letcher et al. (2019a) do, local convergence to NE can be proved in a slightly broader class of games.*

3.2.1 Local Convergence

Local convergence is all one can hope for under general non-concave objectives when employing some form of (stochastic) gradient ascent in mathematical optimization (Bottou, 1998; Karimi et al., 2016). In this spirit, Chasnov et al. (2020) establish local convergence results for learning in games based on the refinement of differential equilibria. In the following, let $\mathcal{X}_i = \mathbb{R}^d$ and the utilities be twice differentiable, $u_i \in C^2(\mathcal{X}, \mathbb{R})$, for all agents i.

Differential Nash equilibrium

Definition 3.1 (Ratliff et al. (2016), Definition 3)**.** *An action profile $x^\star \in \mathcal{X}$ is called a differential Nash equilibrium if $\nabla u(x^\star) = 0$ and the matrix of second partial derivatives of $u(x^\star)$ is negative-definite for all $i \in \mathcal{N}$.*

This narrower definition of equilibrium rules out some degenerate strategy profiles in terms of the utility structure, such as plateaus of utility. In the cycle-game from Example 3.1, the region of attraction is the singleton of the NE, which explains the non-convergence. As the name suggests, the region of attraction of a point (or set of points) is simply a neighborhood of all points such that their trajectories lead to said point (or set). With the notion of *stability* of differential NE (see (Chasnov et al., 2020) for details), the following local guarantee can be established:

Proposition 3.2 (Chasnov et al. (2020), informal Proposition 2)**.** *Consider a game $\mathcal{G}$ satisfying some regularity conditions on its utilities and second partial derivatives. Let $x^\star \in \mathcal{X}$ be a stable differential Nash equilibrium. Suppose players use gradient-based learning with some step size constraint. Then, $x \to x^\star$ for x in the region of attraction of $x^\star$.*

3.2.2 Convergence in Markets

As we have seen in section 2.3, the uniqueness of an NE is a strong condition, and we know of markets that admit multiple equilibria. For example, the reverse auction considered in

(Bichler et al., 2023, see chapter 5, Section 6.4) has a winner-takes-all and a pooling equilibrium. The single-item auction with asymmetric priors from (Kaplan and Zamir, 2015) also has multiple BNE. So even if convergence can be established, which equilibria will be reached? Because games satisfying the condition of global monotonicity can exhibit no more than a single equilibrium, these auctions with multiple BNE trivially violate the monotonicity condition. Hence, global monotonicity does not explain the convergence of gradient dynamics in general. Furthermore, we made the violation of monotonicity explicit in a double auction market (Bichler et al., 2022, see chapter 6, Appendix F).

With these impossibility results established, let us go forward with a local analysis. We have transferred the local convergence result of Proposition 3.2 to establish local convergence in a restricted version of bilateral trade in (Bichler et al., 2022, see chapter 6, Proposition 2). We limited the neural network policies to consist of a single neuron without a non-linear activation function (i.e., we restrict the strategy space to linear functions), which allowed keeping track of the functional forms of utilities and their derivatives. Still, this strengthens the argument for local convergence in auction games.

Concluding, we conjecture that many auction games are monotonic or their equilibria attracting, at least on large subsets of their strategy spaces. Another aspect that possibly explains the robust convergence across many markets, despite the absence of global monotonicity or stability, may be the alignment of the set of NE with the set of CCE. This would mean that the learning dynamics converge to CCE and only end up in NE, as the sets are perfectly aligned. Dütting et al. (2014) investigated this idea and showed the uniqueness of CE in a class of full-information games that include procurement auctions and Bertrand competitions.

3.2.3 Verification

As we have seen, a priori certifying convergence in auctions is usually infeasible. That motivated our development and implementation of a verification method. On a high level, it computes approximate best responses by exhaustively trying out alternative actions and comparing the resulting utilities via Monte Carlo sampling. Formally, we are interested in the loss of not playing a best response strategy against the current opponents,

$$\tilde{\ell}_i(\beta) \;=\; \sup_{\beta'_i \in \Sigma_i} \tilde{u}_i(\beta'_i, \beta_{\text{-}i}) - \tilde{u}_i(\beta_i, \beta_{\text{-}i}). \tag{3.11}$$

Clearly, we cannot possibly try out all alternative strategies β'_i, and neither are bounds easy to establish without further restrictions on the strategy space and utility function. To estimate $\tilde{\ell}_i$, we first draw n_{batch} samples from i's prior, and, for each of these samples, we find the approximate (interim) best response from a discrete set of bids of size n_{grid}:

$$\hat{\ell}_i(\beta) \;=\; \frac{1}{n_{\text{batch}}} \sum_{h=1}^{n_{\text{batch}}} \max_{j \in \{1,\ldots,n_{\text{grid}}\}} \left(\frac{1}{n_{\text{batch}}} \sum_{k=1}^{n_{\text{batch}}} u_i\left(v_i^h, b^j, \beta_{\text{-}i}(v_{\text{-}i}^k)\right) \right.$$
$$\left. - \, u_i\left(v_i^h, \beta_i(v_i^h), \beta_{\text{-}i}(v_{\text{-}i}^k)\right) \right). \tag{3.12}$$

This equation also highlights the high computational demand of verification: For all n_{batch} valuations of bidder i, n_{grid} alternative actions by i are evaluated against n_{batch} actions by the opponents (each corresponding to samples from their priors and subsequent evaluations of their bid functions). The calculation becomes incredibly prohibitive for a larger number of agents and higher dimensional action spaces.

Details on this procedure, which includes an extension to handle interdependent prior valuations (which requires advanced conditional sampling methods), can be found in (Bichler et al., 2021, see chapter 4, Section Evaluation Criteria). In (Bichler et al., 2023, see chapter 5, Section 2 of the Online Supplement), we have analyzed the performance of the utility loss considering different sample and grid sizes.

Remark 3.7. *Bosshard et al. (2020) propose a similar verification approach for the discretized auctions they consider. Assuming linear utilities (risk neutrality), bounded valuation spaces, and an IPV model allows them to quantify a theoretical upper bound on the utility loss in full combinatorial auctions (Bosshard et al., 2020, Theorem 2).*

Reducing the Computational Demand

From the perspective of each bidder, an anonymous auction can be interpreted as competing against an aggregate opponent who bids according to the maximum distribution of the bids of the actual opponents. This observation holds true for all auction formats where only the highest opponent bid is considered, as in the case of the first- and second-price payment rules. It was already noted by Rabinovich et al. (2009), where results from games with a continuum of anonymous players were applied to auctions. When this maximum distribution is available, sampling it instead breaks the curse of dimensionality in the number of agents as the game is effectively reduced to two players. This insight can be leveraged both during learning and during verification. For example, in the case of N symmetric bidders with uniform priors, it is common knowledge that the distribution of the maximum valuation simplifies to a Beta distribution with parameters $N - 2$ and 1.

Chapter 4

Zeroth-Order Learning in Symmetric Markets

Peer-Reviewed Journal Paper

Title: Learning equilibria in symmetric auction games using artificial neural networks.

Authors: Martin Bichler, Maximilian Fichtl, Stefan Heidekrüger, Nils Kohring, Paul Sutterer.

In: Nature Machine Intelligence.

Abstract: Auction theory is of central importance in the study of markets. Unfortunately, we do not know equilibrium bidding strategies for most auction games. For realistic markets with multiple items and value interdependencies, the Bayes Nash equilibria (BNEs) often turn out to be intractable systems of partial differential equations. Previous numerical techniques have relied either on calculating pointwise best responses in strategy space or iteratively solving restricted subgames. We present a learning method that represents strategies as neural networks and applies policy iteration on the basis of gradient dynamics in self-play to provably learn local equilibria. Our empirical results show that these approximated BNEs coincide with the global equilibria whenever available. The method follows the simultaneous gradient of the game and uses a smoothing technique to circumvent discontinuities in the ex post utility functions of auction games. Discontinuities arise at the bid value where an infinite small change would make the difference between winning and not winning. Convergence to local BNEs can be explained by the fact that bidders in most auction models are symmetric, which leads to potential games for which gradient dynamics converge.

Citation: Bichler et al. (2021).

This is a License Agreement between Nils Kohring ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	13-Apr-2023
Order License ID	1344735-1
ISSN	2522-5839

Type of Use	Republish in a thesis/dissertation
Publisher	Nature Research
Portion	Chapter/article

LICENSED CONTENT

Publication Title	Nature Machine Intelligence
Article Title	Learning equilibria in symmetric auction games using artificial neural networks
Date	01/01/2018
Rightsholder	Springer Nature BV
Publication Type	e-Journal

Start Page	687
End Page	695
Issue	8
Volume	3

REQUEST DETAILS

Portion Type	Chapter/article
Page Range(s)	1-9
Total Number of Pages	9
Format (select all that apply)	Electronic
Who Will Republish the Content?	Author of requested content
Duration of Use	Life of current edition
Lifetime Unit Quantity	Up to 499

Rights Requested	Main product
Distribution	Worldwide
Translation	Original language of publication
Copies for the Disabled?	No
Minor Editing Privileges?	No
Incidental Promotional Use?	No
Currency	EUR

NEW WORK DETAILS

Title	Multi-Agent Reinforcement Learning for the Computation of Market Equilibria
Instructor Name	Prof. Dr. Martin Bichler

Institution Name	Technical University of Munich
Expected Presentation Date	2023-09-01

ADDITIONAL DETAILS

The Requesting Person/Organization to Appear on the License	Nils Kohring

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Full article
Editor of Portion(s)	N/A
Volume of Serial or Monograph	3
Page or Page Range of Portion	687-695

Title of the Article/Chapter the Portion Is From	Learning equilibria in symmetric auction games using artificial neural networks
Author of Portion(s)	Bichler, Martin; Fichtl, Maximilian; Heidekrüger, Stefan; Kohring, Nils; Sutterer, Paul
Publication Date of Portion	2021-08-09

RIGHTSHOLDER TERMS AND CONDITIONS

If you are placing a request on behalf of/for a corporate organization, please use RightsLink. For further information visit http://www.nature.com/reprints/permission-requests.html and https://www.springer.com/gp/rights-permissions/obtaining-permissions/882. If the content you are requesting to reuse is under a CC-BY 4.0 licence (or previous version), you do not need to seek permission from Springer Nature for this reuse as long as you provide appropriate credit to the original publication. https://creativecommons.org/licenses/by/4.0/ STM Permissions Guidelines STM Permissions Guidelines (2022) - STM (stm-assoc.org) will complement the Terms & Conditions on this page CCC Payment T&Cs (copyright.com)

Marketplace Permissions General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach

not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such “clipping” or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit "publishing ventures" where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- a statement to the effect that such copy was made pursuant to permission,
- a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and
- a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also

ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service. Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery).*** The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or “clipping” does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Last updated October 2022

nature machine intelligence

ARTICLES

https://doi.org/10.1038/s42256-021-00365-4

Learning equilibria in symmetric auction games using artificial neural networks

Martin Bichler ✉, Maximilian Fichtl, Stefan Heidekrüger, Nils Kohring and Paul Sutterer

Auction theory is of central importance in the study of markets. Unfortunately, we do not know equilibrium bidding strategies for most auction games. For realistic markets with multiple items and value interdependencies, the Bayes Nash equilibria (BNEs) often turn out to be intractable systems of partial differential equations. Previous numerical techniques have relied either on calculating pointwise best responses in strategy space or iteratively solving restricted subgames. We present a learning method that represents strategies as neural networks and applies policy iteration on the basis of gradient dynamics in self-play to provably learn local equilibria. Our empirical results show that these approximated BNEs coincide with the global equilibria whenever available. The method follows the simultaneous gradient of the game and uses a smoothing technique to circumvent discontinuities in the ex post utility functions of auction games. Discontinuities arise at the bid value where an infinite small change would make the difference between winning and not winning. Convergence to local BNEs can be explained by the fact that bidders in most auction models are symmetric, which leads to potential games for which gradient dynamics converge.

The literature on machine learning largely focuses on single-agent learning. Multi-agent learning has become more popular recently due to the advent of generative adversarial networks and applications in complex competitive game playing[1–3]. Although complete-information games have seen some progress, equilibrium learning for incomplete-information (also known as Bayesian) games with continuous action spaces is in its infancy. For complete-information games, the worst-case complexity of finding Nash equilibria is known[4], and a number of learning algorithms have been developed for finding equilibria in specific normal-form games such as zero-sum games[5–7]. Auctions arguably form the best-known and practically most relevant application of Bayesian games, central to modern economic theory[8,9] and with a multitude of applications in the field. The derivation of Bayes Nash equilibrium (BNE) strategies for the first- and second-price sealed-bid auction in the independent private values model led to a comprehensive theoretical framework for the analysis of single-item auctions, a landmark result of economic theory[10,11].

Although single-item auctions in this model are well understood, we only know equilibrium strategies for very few multi-item auction environments. For example, no explicit characterization of BNE strategies is known for first-price sealed-bid auctions of multiple homogeneous goods (multi-unit auctions), nor for first-price sealed-bid combinatorial auctions in which bidders can submit bids on packages of goods[11]. Value interdependencies turn out to be even more challenging[12]. In fact, very little is known about BNE strategies in standard auction formats with multiple objects for sale and value interdependencies. Even for single-object auctions, the specification of equilibria can end up in a system of partial differential equations and no closed-form solution is available[13]; however, such environments are important to understand. In fact, the Nobel Memorial Prize in Economic Sciences that was awarded to Paul Milgrom and Robert B. Wilson in 2020 highlighted their contribution to auctions with interdependent values[14].

Numerical techniques to compute BNEs can be very valuable. Although there has been substantial recent work on imperfect-information finite-dimensional extensive-form games such as Poker or other card games[15–18], relatively few papers focus on continuous-type and -action Bayesian games such as auctions. The few initial attempts make strong restrictions such as finite action spaces, single-object auctions, or independent private values with uniform priors and quasilinear utilities[19–25]. The motivation for such restrictions is the computational hardness of equilibrium computation.

We know of the existence of a mixed Nash equilibrium for finite, complete-information games and that computation is PPAD hard[4]. For Bayesian games with continuous types and actions, we neither know whether (possibly mixed) BNEs exist in the general case nor do we know how hard they are to find if they exist. Cai and Papadimitriou[26] showed that finding a BNE in simultaneous auctions for individual items and bidders with independent private values is already hard for PP, a complexity class above the polynomial hierarchy and close to PSPACE, and we know little about the complexity of finding BNEs in other multi-item auctions. Even approximating equilibria in these auction games is NP hard[26].

The theory of learning in games examines what kind of equilibrium arises as a consequence of a process in which agents are trying to maximize their own payoff by adapting to the actions played by other learning agents[27]. Research on equilibrium learning has largely focused on complete-information normal-form games. So far there is no comprehensive characterization of games that are learnable, but there are some important results. For example, it is well-known that no-regret dynamics converge to a coarse correlated equilibrium in arbitrary finite games[28–31] in their average history of play. Coarse correlated equilibria encompass the set of correlated equilibria. The latter is a non-empty convex polytope that in turn contains the convex hull of the game's Nash equilibria such that we get Nash equilibria ⊂ correlated equilibria ⊂ coarse correlated equilibria. By contrast to correlated equilibria, coarse correlated equilibria may contain strictly dominated (pure) strategy profiles with positive probability. This means that although CCEs are learnable via no-regret algorithms, they are a rather weak solution concept[32]. The question is therefore when learning dynamics converge to a Nash equilibrium. A different relaxation of Nash equilibria is given by local equilibria[33] that only

Department of Computer Science, Technical University of Munich, Garching, Germany. ✉e-mail: bichler@in.tum.de

require stability when allowing agents to make infinitesimal—rather than arbitrary—adjustments to their strategies.

Bayesian auction games have received little attention in equilibrium learning until recently. Given how hard it is to find BNEs even in simple simultaneous single-item auctions in the worst case[26], it is far from obvious that no-regret dynamics can find a BNE in continuous-type and -action Bayesian games. Recent work used deep learning for auction design[34–37] but it did not attempt to find BNEs in auctions. Challenges in computing NEs in general-sum games have also led to alternative solution concepts[38]. Apart from this, artificial intelligence and machine learning are increasingly used to predict strategic behaviour of humans[39] or outcomes of auctions in the field[40], as well as for other problems in automated market design, for example, discovery of socially optimal tax policies[41].

We introduce neural pseudogradient ascent (NPGA) as a method to learn ex ante equilibrium bid functions in symmetric Bayesian auction games with continuous-type and action-spaces. The method is generic in that it allows for different types of value interdependencies and utility functions (for example, accommodating risk aversion). Neural networks are used to represent the bid functions of the players, and the agents learn via self-play. Unfortunately, using neural self-play in this environment is not straightforward: although we assume the expected utility of the players (over the distribution of other players' types) are differentiable in the chosen action, a key challenge is that in auctions, their ex post utilities (which are based on specific realizations of types) have discontinuities. Only the latter, however, can be directly observed in the data generated from self-play. As a result, standard ways of gradient computation (that is, backpropagation from the observed data) fail and would result in constant-zero bids by all bidders. We address this problem by deriving pseudo-gradients via evolutionary strategy optimization rather than exact gradients via standard learning methods.

Given the computational hardness of BNE computations in general Bayesian auction games[26], it is not obvious that gradient-ascent schemes such as ours would converge to BNEs. To prove convergence of NPGA to local equilibria, we leverage the fact that the vast majority of auction games described in the literature assume symmetric bidders and equilibrium bid functions[11]. This leads to a potential game, and gradient dynamics converge to local Nash equilibria in potential games. Although there can also be asymmetric equilibria, such equilibria are often unnatural and the symmetry assumption encompasses a very large set of interesting auction environments. An example of such an asymmetric equilibrium is given in a second-price auction when one player bids the upper bound of the distribution whereas all of the others bid constant zero, independent of their respective private valuations.

In our experiments we illustrate NPGA via a combinatorial auction in the local–local–global (LLG) model[42], which has received considerable attention due to the use of core-selecting combinatorial auctions for spectrum sales worldwide[43]. In the LLG model, core-selecting auctions with risk-neutral bidders are known to be economically inefficient. It is one of the few multi-object auction models in which correlation among bidder valuations has been investigated analytically with quasilinear utility functions, but this is not the case for risk aversion. Yet such multi-object environments with interdependencies and non-quasilinear utility functions have not been explored in the scarce literature on equilibrium computation. Using NPGA, we can show that risk aversion mitigates the inefficiencies that arise in the equilibrium of risk-neutral bidders, while correlation among the bidders' valuations has little impact. This result is of independent interest to policymakers. In the Supplementary Information we discuss further experiments in a number of additional environments to demonstrate the versatility of the method.

To apply NPGA, we neither need to specify the equilibrium as a system of differential equations, nor do we need to derive complex conditional type distributions in settings with interdependencies. As a result, NPGA provides a convenient method to explore symmetric sealed-bid auction models and study the BNEs that arise with different types of interdependencies, distributional assumptions or different levels of risk aversion.

The algorithm

We will now introduce the necessary notation before stating the algorithm and discussing its convergence properties.

Notation. An incomplete-information or Bayesian game is given by a sextuplet $G=(\mathcal{I},\mathcal{V},\mathcal{O},\mathcal{A},f,\mathbf{u})$. Here $\mathcal{I}=\{1,\ldots,n\}$ denotes the set of agents participating in the game. The joint probability density function $f:\mathcal{V}\times\mathcal{O}\to\mathbb{R}_{\geq 0}$ describes an atomless prior distribution over agents' types, given by tuples (o_i, v_i) of observations and valuations. We make no further restrictions on f, thus allowing for arbitrary correlations; f is assumed to be common knowledge and we will denote its marginals by f_{v_i}, f_{o_i} and so on; its conditionals by $f_{v_i|o_i}$ and so on; and its associated probability measure by F. Agent i's private observation is then given as a realization $o_i\in\mathcal{O}_i$, with $\mathcal{O}=\mathcal{O}_1\times\cdots\times\mathcal{O}_n$ being the set of possible observation profiles. Similarly, $\mathcal{V}$ denotes the set of true but possibly unobserved valuations. Crucially, we make this distinction to model interdependencies in settings beyond purely private values or purely common values. Based on o_i, the agent chooses an action or bid, $b_i\in\mathcal{A}_i$, and the set of possible action profiles is given by $\mathcal{A}=\mathcal{A}_1\times\cdots\times\mathcal{A}_n$. For each possible action and valuation profile, the vector $\mathbf{u}=(u_1,\ldots,u_n)$ of F-integrable, individual (ex post) utility functions $u_i:\mathcal{A}\times\mathcal{V}_i\to\mathbb{R}$ assigns the game outcome to each player. Ex ante (before the game), agents neither possess observations nor valuations, only knowledge about f. In the interim stage, agents also observe o_i that provide (possibly partial or noisy) information about their own v_i. Full access to the outcomes $\mathbf{u}(\mathbf{v},\mathbf{b})$ is given only after taking actions (ex post). In our formulation, we do not assume explicit ex post access to any values (for example, v_i,$\mathbf{v}_{-i}$,$\mathbf{b}_{-i}$) beyond the outcome u itself. An index $-i$ denotes a partial profile of all agents but agent i.

Taking an ex ante view, players are tasked with finding strategies $\beta_i:\mathcal{O}_i\to\mathcal{A}_i$ that map observations to bids. We denote the resulting spaces of individual and joint pure strategies by $\Sigma_i\equiv\mathcal{A}_i^{\mathcal{O}_i}$ and $\Sigma\equiv\prod_i\Sigma_i$, respectively. Note that even for pure strategies, the spaces Σ_i are infinite dimensional unless $\mathcal{O}_i$ are finite (in which case they are finite-dimensional but remain infinite for continuous $\mathcal{A}_i$). We will slightly restrict ourselves to square-integrable strategies and equip Σ_i with the inner product $\langle\cdot,\cdot\rangle_{\Sigma_i}:\Sigma_i\times\Sigma_i\to\mathbb{R}$, $(\alpha,\beta)\mapsto\mathbb{E}_{\mathbf{o}}\sim f_{\mathbf{o}}\left[\alpha(\mathbf{o})^T\beta(\mathbf{o})\right]$ and the norm $\|\beta\|_{\Sigma_i}\equiv\sqrt{\langle\beta,\beta\rangle_{\Sigma_i}}$ such that they form Hilbert spaces[44].

The primary Bayesian games we will consider are sealed-bid auctions on m indivisible items. In general combinatorial auctions, we thus have a set $\mathcal{K}$ of possible bundles of items and the valuation- and action-spaces are therefore of dimension $|\mathcal{K}|=2^m$. We always have $o_i=v_i$ in the private values setting, whereas in the common values setting there is some unobserved constant $v_c=v_1=\cdots=v_n$, where o_i can be considered noisy measurements of v_c. Mixed settings are likewise possible. In any case, based on bid profile $\mathbf{b}$, an auction mechanism will determine two things: (1) an allocation $\mathbf{x}=\mathbf{x}(\mathbf{b})=(x_1,\ldots x_n)$, which constitutes a partition of m where bidder i is allocated the bundle x_i; and (2) a price vector $\mathbf{p}(\mathbf{b})\in\mathbb{R}^n$, where the component p_i is the monetary amount bidder i has to pay to receive x_i. Formally, one may consider the individual allocations to be one-hot-encoded vectors $x_i\in\{0,1\}^{|\mathcal{K}|}$. In the standard risk-neutral model, u_i values are then described by quasilinear (QL) payoff functions $u_i^{\text{QL}}(v_i,\mathbf{b})=(x_i(\mathbf{b})\cdot v_i-p_i(\mathbf{b}))$, that is, by how much a player values their allocated bundle minus the price they have to pay. An extension to this basic setting includes risk aversion (RA). Here we model risk-aversion via utilities $u_i^{\text{RA}}=\left(u_i^{\text{QL}}\right)^\rho$

where $\rho \in (0,1]$ is the risk attitude; $\rho = 1$ describes risk neutrality, where smaller values lead to strictly concave, risk-averse transformations of u_i^{QL}. Risk aversion is an established way to explain why bidders in field studies of single-object first-price sealed-bid (FPSB) auctions bid higher than their risk-neutral counterparts in analytical BNEs[45].

For fixed-strategy profiles $\boldsymbol{\beta} \in \Sigma$, we can extend the notion of utility to the interim and ex ante stages and use this to characterize the Nash equilibria of Bayesian games: although other agents follow $\boldsymbol{\beta}$, we define agent i's interim utility as the expected utility of choosing an action b_i conditioned on o_i:

$$\bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i}) = \mathbb{E}_{v_i, \mathbf{o}_{-i}|o_i}\left[u_i(v_i, b_i, \boldsymbol{\beta}_{-i}(\mathbf{o}_{-i}))\right]. \tag{1}$$

We will also introduce the interim utility loss $\bar{\ell}$ that is incurred by not playing a best response b_i':

$$\bar{\ell}_i(o_i; b_i, \boldsymbol{\beta}_{-i}) = \sup_{b_i' \in \mathcal{A}_i} \bar{u}_i(o_i, b_i', \boldsymbol{\beta}_{-i}) - \bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i}). \tag{2}$$

Then, an (interim) ϵ-Bayes Nash equilibrium (ϵ-BNE) is a strategy profile $\boldsymbol{\beta}^* = (\beta_1^*, \dots, \beta_n^*) \in \Sigma$ such that no agent can improve their own interim utility by more than $\epsilon \geq 0$ by unilaterally deviating from $\boldsymbol{\beta}^*$; thus, the following holds in an ϵ-BNE:

$$\forall i \in \mathcal{I},\ o_i \in \mathcal{O}_i: \quad \bar{\ell}_i\left(o_i; \beta_i^*(o_i), \boldsymbol{\beta}_{-i}^*\right) \leq \epsilon. \tag{3}$$

For $\epsilon = 0$, we will call the BNE exact, or simply drop the ϵ prefix. We will also need the ex ante utility (defined as $\tilde{u}_i(\beta_i, \boldsymbol{\beta}_{-i}) = \mathbb{E}_{o_i \sim f_{o_i}}[\bar{u}_i(o_i, \beta_i(o_i), \boldsymbol{\beta}_{-i})]$), which can be interpreted as the expected utility over all of f for a particular β_i against fixed opponents $\boldsymbol{\beta}_{-i}$. Similarly, we will define ex ante loss $\tilde{\ell}_i(\beta_i, \boldsymbol{\beta}_{-i})$ and ex ante ϵ-BNEs analogously to equations (2) and (3). Note that now we can interpret the ex ante state of the Bayesian game as a complete-information game $\tilde{G} = (\mathcal{I}, \Sigma, \tilde{\mathbf{u}})$ with an infinite-dimensional action space Σ that is identical to the strategy space of the Bayesian game. Every exact (interim) BNE also clearly constitutes an exact ex ante BNE. The reverse holds almost surely, that is, any ex ante equilibrium fulfills equation (3), except possibly on a set $O \subset \mathcal{O}$ with $F(O) = 0$. To see this, one may consider the equations $0 = \tilde{\ell}_i(\boldsymbol{\beta}^*) = \mathbb{E}_{o_i}\left[\bar{\ell}_i(o_i; \beta_i^*(o_i), \boldsymbol{\beta}_{-i}^*)\right]$ and the fact that $\bar{\ell}_i(o_i, \boldsymbol{\beta}) \geq 0$ by definition. Importantly, this almost sure equivalence of ex ante and (interim) BNEs holds for $\epsilon = 0$ but not for strictly positive ϵ: given an ex ante κ-BNE, equation (3) (with $\epsilon = \kappa > 0$) must only hold in expectation but may be violated with strictly positive probability. To delineate this difference between ex ante and interim approximate equilibria, we will write κ and ϵ to denote their respective approximation bounds.

Due to the known computational hardness of computing NEs and BNEs, one is often interested in relaxations of equilibria that may be easier to find in some circumstances. For example, in local BNEs, the loss requirement is relaxed to only consider best responses from a neighbourhood of the equilibrium strategy profile: we call $\boldsymbol{\beta}^*$ a local ex ante BNE if there exists an open set $\emptyset \neq W_i \subset \Sigma_i$ such that $\beta_i^* \in W_i$ and $\tilde{u}_i(\beta_i^*, \boldsymbol{\beta}_{-i}^*) \geq \tilde{u}_i(\beta_i', \boldsymbol{\beta}_{-i}^*)$ for all agents i and all alternative strategies $\beta_i' \in W_i$. If all utility functions u_i are strictly concave in i's action, the game admits a unique global BNE[46] and no other local BNEs.

Smoothness of the (ex post) utilities is a standard assumption in the analysis of Bayesian games[46], but this is commonly violated in auctions due to the discrete nature of $\mathbf{x}$. Instead let us introduce a weaker notion of smoothness at the interim stage, which lends itself for theoretical analysis while being consistent with auction games.

Definition 1 (interim-smooth Bayesian game). We call a Bayesian game with continuous types $\mathcal{V}_i \times \mathcal{O}_i$ and actions $\mathcal{A}_i \subseteq \mathbb{R}^K$ interim smooth if: (1) the interim utilities $\bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i})$ are continuously differentiable with respect to their second argument for each $i \in \mathcal{I}$ and any $o_i \in \mathcal{O}_i$, $\boldsymbol{\beta}_{-i} \in \Sigma_{-i}$; (2) all partial derivatives are uniformly bounded by a finite constant $Z < \infty$:

$$\forall i, o_i, \boldsymbol{\beta}_{-i}, b_i, k \in [K]: \quad \left\| \frac{\partial \bar{u}_i}{\partial b_{ik}}(o_i, b_i, \boldsymbol{\beta}_{-i}) \right\| \leq Z; \tag{4}$$

and (3) the ex post utilities are F-square-integrable: there exists $S < \infty$, such that for all $i \in \mathcal{I}$, $\boldsymbol{\beta} \in \Sigma$:

$$\mathbb{E}_{v_i, \mathbf{o}}\left[u_i\left(v_i, \beta_i(o_i), \boldsymbol{\beta}_{-i}(\mathbf{o}_{-i})\right)^2\right] \leq S \tag{5}$$

To see why the assumption of interim differentiability is justified, consider that ex post utilities in auctions are generally piecewise smooth. Non-differentiability only occurs at the bid profiles in which the auctioneer is indifferent between multiple possible $\mathbf{x}$. In theory, one could therefore interpret the interim expected utility as a lottery over many smooth ex post utility functions that each describe a particular $\mathbf{x}$. The choice probabilities for these are given by $P(\mathbf{x}|b_i, o_i, \boldsymbol{\beta}_{-i})$, bidder i's Bayesian belief that $\mathbf{x}$ will be chosen if they bid b_i. If $\boldsymbol{\beta}_{-i}$ are continuous and f is atomless, these probabilities—and therefore the interim expected utilities as a whole—are smooth in b_i.

In interim-smooth Bayesian games, we write $\nabla \bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i}) \equiv (\partial \bar{u}_i(o_i, b_i, \boldsymbol{\beta}_{-i}) / \partial b_{ik})_k$ and call it the interim payoff gradient. Furthermore, when G is interim-smooth, the ex ante gradients $\nabla_{\beta_i} \tilde{u}_i(\beta_i, \boldsymbol{\beta}_{-i}) \in \Sigma_i$ are also guaranteed to exist and given by the Gateaux derivatives in the Hilbert spaces Σ_i.

Finally, symmetric models are prevalent in auction theory[11]. We will call a Bayesian game symmetric if all players' $i, j \in \mathcal{I}$ marginal prior-type distributions are identical (but not necessarily independent), that is, $f_{v_i, o_i} = f_{v_j, o_j}$, as are their individual utilities (almost surely, up to tiebreaking): $u_i(\beta_i, \boldsymbol{\beta}_{-i}) = u_j(\beta_i; \boldsymbol{\beta}_{-i})$, with probability 1. The literature primarily discusses[11] equilibria that are likewise symmetric, that is, where $\boldsymbol{\beta}^* = (\beta_1^*, \beta_1^*, \dots \beta_1^*)$. We will refer to auctions that are both symmetric and interim-smooth as symmetric and smooth auction games.

NPGA. Our numerical technique to learn BNEs, NPGA, is based on neural networks and repeated self-play, in which players continually update strategies in response to observed game outcomes, that is, all agents follow the game dynamics. By game dynamics, we mean the vector field of the simultaneous gradients of the ex ante utility functions of all players. The goal will be to find an ex ante BNE $\boldsymbol{\beta}^*$ for a continuum of observations $\mathbf{o}$ that bidders can draw. In other words, we search for a profile of equilibrium bid functions in infinite-dimensional spaces. We will first introduce the procedure in the general case before showing convergence for symmetric and smooth auction games in the 'Convergence' section.

We start by taking the infinite-dimensional, complete-information game interpretation $\tilde{G} = (\mathcal{I}, \Sigma, \tilde{\mathbf{u}})$ mentioned in the previous section. To implement gradient ascent in the Hilbert space Σ, we replace the bid functions by neural networks called policy networks that are parametrized by finite-dimensional parameter vectors $\theta_i \in \Theta_i \subseteq \mathbb{R}^{d_i}$. This lets us define a finite-dimensional approximation of $\tilde{G}$, which we will call the proxy game.

Definition 2 (proxy game). Let $G = (\mathcal{I}, \mathcal{V}, \mathcal{O}, \mathcal{A}, f, \mathbf{u})$ be a Bayesian game with ex ante utilities $\tilde{u}_i$ and let its strategy functions be implemented by neural networks: $\beta_i(o_i) \equiv \pi_i(o_i; \theta_i)$, where θ_i are the networks' parameters chosen from finite-dimensional vector spaces $\Theta_i \subseteq \mathbb{R}^{d_i}$. Set $\Theta \equiv \prod_i \Theta_i$ and (with slight abuse of notation) write $\tilde{u}_i(\theta_i, \boldsymbol{\theta}_{-i}) \equiv \tilde{u}_i\left(\pi_i(\cdot; \theta_i), \boldsymbol{\pi}_{-i}(\cdot; \boldsymbol{\theta}_{-i})\right)$. We then call the resulting finite-dimensional complete-information game on parameters, $\Gamma = (\mathcal{I}, \Theta, \tilde{\mathbf{u}})$, the proxy game of G.

Common neural network architectures have been shown to be able to approximate any sufficiently regular function arbitrarily well[47]; thus, this choice of function approximation enables the learning of a wide variety of bid functions with minimal structural constraints. Neural networks also demonstrably achieve good performance in machine learning settings with very high-dimensional input vectors, as is the case in larger auctions with many items. Using neural networks we therefore effectively reduce the problem from finding an infinite-dimensional vector in Σ to finding finitely many (d_i) weights and biases of the neural networks, and we can now perform gradient ascent in the finite-dimensional parameter spaces.

Each agent aims to maximize the objective function of their network, which is given by $\tilde{u}_i$ and estimated via the empirical sample mean of ex post utilities of a batch of H auctions, where H is a large integer: after playing a batch of games, agents observe their utility, estimate its gradient with respect to θ_i and apply an update to θ_i that is expected to lead to an increase in utility.

Traditionally, gradient estimates in neural networks are computed via backpropagation; however, training neural networks in auction games is challenging as the ex post utility functions of individual auctions are discontinuous, leading to a failure to backpropagate gradients through the empirical objective. We solve this problem by leveraging an evolutionary strategy (ES) optimization technique that effectively smoothes the objective[48,49]. This allows us to derive an adequate estimate of the ex ante payoff gradients even under ex post non-smoothness.

Algorithm 1 (NPGA using ES gradients).

```
Input: agents $i \in \mathcal{I}$ with initial policies $\beta_i^0 := \pi_i(\cdot;\theta_i^0)$ induced
       by initial parameters $\theta_i^0$; ES population size $P$; ES noise
       standard deviation $\sigma$; learning rate $\eta$; batch size $H$
for $t := 1, 2, \ldots$ do
    Sample a batch $(\mathbf{v}_h, \mathbf{o}_h)_{h=1,\ldots,H}$ of valuation and obseravtion
    profiles from the prior $f$
    Calculate joint utility in current strategy profile:
        $\tilde{\mathbf{u}}^{t-1} := \frac{1}{H}\sum_h \tilde{\mathbf{u}}\left(\mathbf{v}_h, \boldsymbol{\beta}^{t-1}(\mathbf{o}_h)\right)$
    for each agent $i \in \mathcal{I}$ do
        Sample $P$ perturbations of agent $i$'s current policy:
            $\pi_{i;p} := \pi_i(\cdot;\theta_p)$
          with $\theta_p := \theta_i^{t-1} + \varepsilon_p$ where $\varepsilon_p \approx \mathcal{N}(0, \sigma^2 I)$ i.i.d. for all
          $p \in \{1, \ldots, P\}$
        For each $p$, evaluate the fitness of $\theta_p$ by playing against current opponents:
            $\varphi_p := \frac{1}{H}\sum_h u_i\left(v_{h,i}, \pi_{i;p}(o_{h,i}), \boldsymbol{\beta}_{-i}^{t-1}(\mathbf{o}_{h,-i})\right) - \underbrace{\tilde{u}_i^{t-1}}_{\text{baseline}}$
        Calculate ES pseudogradient as fitness-weighted perturbation noise:
            $\nabla^{\mathrm{ES}}\tilde{u}_i^{t-1} := \frac{1}{\sigma^2 P}\sum_p \varphi_p \varepsilon_p$
        Perform a gradient update step on the current policy:
            $\Delta\theta_i^t := \eta^t \nabla^{\mathrm{ES}}\tilde{u}_i^{t-1},$
            $\theta_i^t := \theta_i^{t-1} + \Delta\theta_i^t,$
            $\beta_i^t := \pi_i(\cdot;\theta_i^t)$
    end
end
```

We provide the pseudocode of NPGA in *Algorithm 1*. At each time-step t, every agent $i \in \mathcal{I}$ receives a noisy estimate $\hat{\nabla}\tilde{u}_i$ of their individual (ex ante) payoff gradient at the current strategy profile. The noise is an artefact of limited-precision Monte Carlo sampling over $\mathcal{V}$ and $\mathcal{O}$. The agents simultaneously take a step along this gradient estimate to determine the strategies for the next stage and continue playing.

Convergence. In our experimental results below and the Supplementary Information, we find that NPGA always converges very close to the global ϵ-BNE, which was surprising at first given the known results about non-convergence of gradient play to Nash equilibria in general[50], and the locality of gradient-based learning. Non-convergence can be due to conflicting utility functions of players. For example, even in simple two-player zero-sum games with one-dimensional actions, the simultaneous gradient may cycle around the Nash equilibrium[51].

A few observations help explain why NPGA converges to an approximate BNE in a wide range of auction games. First, the vast majority of models studied in the literature are symmetric auction games with symmetric equilibria (see the 'Notation' section). As a result, we no longer need to learn multiple bid functions for each bidder in NPGA, but merely a single symmetric bid function $\beta_1 \in \Sigma_1$ that optimizes the single ex ante utility function $\tilde{u}_1(\beta_1, \ldots, \beta_1)$, which serves as a potential function of the game. Any maximum β_1^* of this potential function directly yields a symmetric pure strategy ex ante BNE $\boldsymbol{\beta}^* = (\beta_1^*, \ldots, \beta_1^*)$. For the finite-dimensional proxy game, we can formalize the claim in the following section.

Definition 3 (potential game). A complete-information game $\Gamma = (\mathcal{I}, \Theta, \tilde{\mathbf{u}})$ is an (exact) potential game[52] if there exists a potential function $\phi : \Theta \to \mathbb{R}$, s.t. for all $i \in \mathcal{I}$, $\theta_i, \theta_i' \in \Theta_i$ and $\boldsymbol{\theta}_{-i} \in \Theta_{-i}$, it holds that

$$\tilde{u}_i(\theta_i, \boldsymbol{\theta}_{-i}) - \tilde{u}_i(\theta_i', \boldsymbol{\theta}_{-i}) = \phi(\theta_i, \boldsymbol{\theta}_{-i}) - \phi(\theta_i', \boldsymbol{\theta}_{-i}). \quad (6)$$

When the auction game is symmetric and we additionally enforce symmetric strategies by sharing a common neural network architecture $\pi(\cdot)$ and common parameter vector $\theta_i \equiv \theta_1$ among all players (symmetric NPGA); it is easy to see that with $\phi \equiv \tilde{u}_1$, the proxy game is an exact potential game. Gradient play provably converges to a pure local Nash equilibria in finite-dimensional, continuous potential games[33]. This leads us to the following proposition.

Proposition 1. In any symmetric and smooth auction game, symmetric NPGA with appropriate gradient update step sizes almost surely converges to a local ex ante κ-BNE.

A formal proof can be found in the Methods.

Empirical evaluation

We illustrate the versatility of NPGA in the context of combinatorial auctions in the well-known LLG environment, which has been an important model for the discussion about spectrum auction formats[43,53]. The NPGA model allows us to analyse how correlation and risk aversion impact the outcome in equilibrium. There are many other interesting environments one can explore. In the Supplementary Information we present further results for single-object auctions with different types of value interdependencies (including common values models), small and larger mult-unit auctions, and a larger combinatorial auction setting with eight items and six bidders. Note that even for a multi-unit auction with three items and bidders, no analytical solutions are known anymore. For single-object, multi-unit and combinatorial auctions with only a few bidders, as reported below, NPGA computes equilibria within hundreds of iterations, each taking a few seconds or less. Larger settings such as multi-unit FPSB auctions with four units and bidders or combinatorial auctions with five items and six bidders reported in the Supplementary Information converged to an approximate BNE with estimated relative utility loss of less than 1% within 15 min; however, the runtime depends on the specific model analysed (for example, the prior distribution, the number of bidders and the auction format).

The LLG model. The LLG model consists of two objects $\{1, 2\}$, two local bidders $i\in\{1, 2\}$ and one global bidder $i=3$, with each only interested in one specific bundle (of the single object i (locals) or both objects (global)[42]. We will simply denote the valuation of each bidder's single bundle by $v_i \in \mathbb{R}$. We consider a private values (but not independent private values) setting with $o_i=v_i$, which allows for correlation. The situation is akin to spectrum sales in countries with regional spectrum licences such as Australia or Canada, where local telecoms compete against operators who provide their services nationwide, and governments have used core-selecting combinatorial auctions. The core of an auction game describes the set of outcomes such that no coalition of bidders (and possibly the auctioneer) can profitably deviate. Core-selecting auction mechanisms enforce this notion of stability by their choice of prices. Although there are hardly any game-theoretical analyses of combinatorial auctions, this model is simple enough to allow for the derivation of analytical results[54]. It was shown that with independent private values and risk-neutral bidders, core-selecting payment rules lead to considerable inefficiencies in equilibrium[42] in combinatorial auctions. The two local bidders attempt to free ride on each other. If one bidder bids less, the other has to bid more to overbid the global bidder. Due to incomplete information, both local bidders could bid too low in total and fail to outbid the global bidder, even if their combined valuations are higher than those of the global bidders. This results in an inefficient outcome. This fact has been used as an argument against core-selecting combinatorial auctions[43].

It is interesting to understand equilibria with different assumptions. For example, it is reasonable to believe that bidder valuations in spectrum auctions are correlated, because telecoms face the same downstream market. The model was recently analysed with different types of correlation[54]; however, with standard core-selecting payment rules, it turns out that correlation alone cannot mitigate the efficiency and revenue loss encountered with independent private values. Risk aversion has not yet been analysed, although it plays a role in the revenue ranking of single-object auctions. By contrast to single-object auctions, it has been unclear how risk-aversion plays out in equilibrium. If one local bidder knows that the other is risk averse and might thus bid higher, they might bid even lower as a result of this knowledge. The environment is not symmetric as there are two local bidders and a global bidder. However, the global bidder has a simple dominant strategy to bid truthfully under certain core-selecting payment rules. The gradient dynamics of the global player's network will then stably approach this dominant strategy regardless of the local bidders' behaviour, and the two local bidders can indeed be considered symmetric whenever $f_{v_1} = f_{v_2}$ and thus form a 'local potential game'. NPGA can therefore be expected to converge to a BNE despite the environment's asymmetry.

Ausubel and Baranov[54] investigate two models of correlation among local bidders' private values and derive analytical BNEs, which we will use as a baseline in our experiments. Let us define the joint prior f to be the five-dimensional uniform distribution of a latent random variable $\boldsymbol{\omega} \sim \mathcal{U}[0, 1]^5$. Then let $v_3=2\omega_3$ be the valuation of the global bidder and

$$v_1(\omega) = w\omega_4 + (1-w)\omega_1, \quad v_2(\omega) = w\omega_4 + (1-w)\omega_2 \tag{7}$$

be the valuations of the local bidders where the weight w is a random variable depending on ω_5 only. The valuations of the local bidders can be thought of as a linear combination of an individual component ω_i and a common component ω_4. Now given an exogenous correlation parameter $\gamma\in[0, 1]$, Ausubel and Bananov[54] propose two different ways to choose w such that $\mathrm{corr}(v_1, v_2)=\gamma$: the Bernoulli weights model:

$$w(\omega) = \begin{cases} 1 & \text{if } \omega_5 < \gamma, \\ 0 & \text{else}, \end{cases} \tag{8}$$

and the constant weights model (which does not require w_5):

$$w(\omega) = \begin{cases} \frac{\gamma - \sqrt{\gamma(1-\gamma)}}{2\gamma - 1} & \text{if } \gamma \neq 1/2, \\ 1/2 & \text{else}. \end{cases} \tag{9}$$

They analytically derive the unique symmetric BNE strategies for multiple bidder-optimal core-selecting payment rules including the nearest-zero (NZ), nearest-VCG (NVCG, named after the Vickrey–Clarke–Groves (VCG) payments) and nearest-bid (NB) rule in the Bernoulli weights model. These rules all choose efficient $\mathbf{x}$ (according to the submitted bids), but select different price vectors $\mathbf{p}$ from the set of core-stable outcomes. For example, the NVCG rule picks the point in the core that minimizes the Euclidean distance to the (unique) VCG payments. Similarly, the NZ point takes the origin of the coordinate system as a reference point, whereas the NB rule minimizes the distance to the vector of submitted bids $\mathbf{b}$. Only the NVCG rule has been used in spectrum sales so far. Apart from these core-selecting payment rules, we will also report the results in FPSB auctions, for which no analytical BNEs are known, as these are used in some spectrum sales[43], and in the VCG mechanism, which is not core-stable but always prescribes truthful bidding as a BNE.

Evaluation criteria. Let us discuss how we will evaluate any learned $\boldsymbol{\beta}$ to certify that it indeed constitutes an (approximate) equilibrium. This evaluation is entirely independent of the learning process of NPGA and tries to answer the question of how good a given strategy is. Whenever we encounter a setting where an analytical equilibrium $\boldsymbol{\beta}^*$ is known, we draw on it for direct comparison. In this case, we sample the BNE utility of each player, $\hat{u}_i(\boldsymbol{\beta}^*) \approx \tilde{u}_i(\boldsymbol{\beta}^*)$, as well as the utility β_i played against the BNE, $\hat{u}_i(\beta_i, \boldsymbol{\beta}^*_{-i}) \approx \tilde{u}_i(\beta_i, \boldsymbol{\beta}^*_{-i})$, with a batch size of 2^{22}. We then report the resulting relative utility loss:

$$\mathcal{L}_i(\boldsymbol{\beta}_i) = 1 - \frac{\hat{u}_i(\beta_i, \boldsymbol{\beta}^*_{-i})}{\hat{u}_i(\beta^*_i, \boldsymbol{\beta}^*_{-i})}. \tag{10}$$

We also report the probability-weighted r.m.s.e. of β_i and β^*_i in the action space, which approximates the L_2 distance $\| \beta_i - \beta^*_i \|_{\Sigma_i}$ of these two functions:

$$L_2(\beta_i) = \left(\frac{1}{n_{\mathrm{batch}}} \sum_{o_i} (\beta_i(o_i) - \beta^*_i(o_i))^2 \right)^{\frac{1}{2}}. \tag{11}$$

This metric circumvents the drawback of $\mathcal{L}_i$ that even a strategy with a loss very close to zero could be arbitrarily far from the actual BNE in strategy space.

When no analytical BNE is available for certification of the learned bid function, we aim to compute the ex ante utility loss $\tilde{\ell}_i(\beta_i, \boldsymbol{\beta}_{-i}) = \sup_{\beta'_i \in \Sigma_i} \tilde{u}_i(\beta'_i, \boldsymbol{\beta}_{-i}) - \tilde{u}_i(\beta_i, \boldsymbol{\beta}_{-i})$. Evaluating this supremum exactly in function space Σ_i is not tractable and approximations are computationally expensive. Our estimator $\hat{\ell}_i$ of $\tilde{\ell}_i$ relies on finding approximate interim best responses. To do so, we place an equidistant grid indexed with $w=1, \ldots, n_{\mathrm{grid}}$ over the action space $\mathcal{A}_i$ ranging from zero to the maximum valuation for all dimensions. For o_i and each of the alternative bids b_w, we evaluate the interim utility $\bar{u}_i(o_i, b_w, \boldsymbol{\beta}_{-i})$ against the current opponent strategy profile. This is challenging as it requires access to the distribution of i's true valuation and the opponents' observations, both conditioned on

Table 1 | Convergence results of NPGA in risk-neutral combinatorial LLG auctions with a correlation of $\gamma = 0.5$ among local bidders' valuations. We report mean and s.d. of experiments over ten runs

Auction game	L_2	$\mathcal{L}$	$\hat{\mathcal{L}}$
LLG Bernoulli NZ	0.011 (0.005)	0 (0)	0.007 (0.007)
LLG Bernoulli VCG	0.008 (0.003)	0.001 (0)	0.007 (0.005)
LLG Bernoulli NVCG	0.016 (0.016)	0 (0)	0.008 (0.007)
LLG Bernoulli NB	0.021 (0.021)	0.001 (0)	0.009 (0.008)
LLG Bernoulli FPSB	–	–	0.010 (0.008)
LLG constant NZ	–	–	0.011 (0.010)
LLG constant VCG	–	–	0.008 (0.007)
LLG constant NVCG	–	–	0.011 (0.012)
LLG constant NB	–	–	0.013 (0.015)
LLG constant FPSB	–	–	0.009 (0.006)

o_i (see equation (1)). For n_{batch} samples of o_i and n_{batch} samples of $v_i, \mathbf{o}_{-i}|o_i$ for each o_i, we then have

$$\hat{\ell}_i(\boldsymbol{\beta}) = \frac{1}{n_{\text{batch}}} \sum_{o_i} \max_w \lambda_i(o_i, b_w, \boldsymbol{\beta}) \qquad (12)$$

with λ_i being the estimated expected utility gain by deviating from playing according to β_i to playing action b':

$$\begin{aligned}\lambda_i(o_i, b', \boldsymbol{\beta}) = \tfrac{1}{n_{\text{batch}}} \sum_{v_i, \mathbf{o}_{-i}|o_i} \Big(u_i\left(v_i, b', \boldsymbol{\beta}_{-i}(\mathbf{o}_{-i})\right) \\ -u_i\left(v_i, \beta_i(o_i), \boldsymbol{\beta}_{-i}(\mathbf{o}_{-i})\right)\Big).\end{aligned} \qquad (13)$$

For an increasing number of samples and alternative actions, this estimate converges to $\tilde{\ell}_i$. Our estimate for ϵ in an ex ante ϵ-BNE is then $\epsilon \equiv \max_i \hat{\ell}_i$.

The conditional distribution $v_i, \mathbf{o}_{-i}|o_i$ is rarely available upfront. For simple cases one can derive the analytical distributions and draw samples; however, in most programming environments, one is only able to sample from very basic (pseudo)random numbers such as the uniform or normal distribution. For more complicated multivariate conditional distributions, we use the conditional distribution method (for details, see Supplementary Section 3). Based on these estimates, we can compute a relative ex ante utility loss without access to the analytical BNEs:

$$\hat{\mathcal{L}}_i(\boldsymbol{\beta}) \; = \; 1 - \frac{\hat{u}_i(\boldsymbol{\beta})}{\hat{u}_i(\boldsymbol{\beta}) + \hat{\ell}_i(\boldsymbol{\beta})}. \qquad (14)$$

This metric is the average loss incurred by not playing a best response but instead playing the strategy learned via NPGA. Note that we do not need to make any assumption about the utility function or independence of valuations for this estimator.

Due to the multiple levels of Monte Carlo sampling, the estimator $\hat{\mathcal{L}}_i$ has a higher variance than those that rely on an analytical BNE $\boldsymbol{\beta}^*$, even when the performance of NPGA itself is not affected. Our reported estimates are based on $n_{\text{grid}} = 2^{10}$ possible bids for each sampled interim state using a batch size of $n_{\text{batch}} = 2^{12}$, thus each estimate of $\hat{\mathcal{L}}$ is based on $n_{\text{grid}} \cdot n^2_{\text{batch}} \; = 2^{34}$ simulated auctions. To sample that many games efficiently, both NPGA and our evaluation procedures leverage parallelization on GPU hardware. Certification of BNEs is a challenge in all computational approaches to equilibrium computation. A thorough discussion for environments with standard quasilinear utility functions and independent private values are provided in ref. [23].

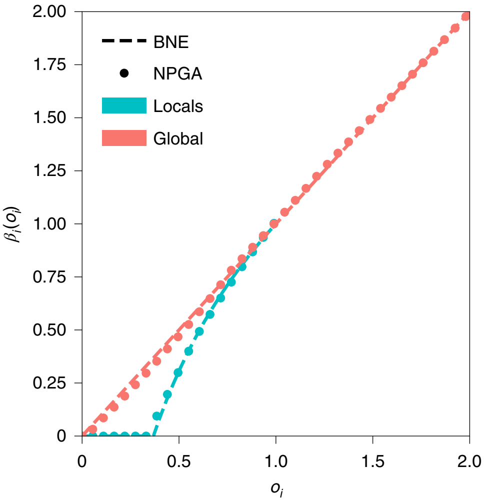


Fig. 1 | Bid functions in the LLG auction with the nearest-zero core payment rule. Bidders are independent and risk neutral. The strategies learned by NPGA (dotted) almost perfectly recover the analytical equilibrium strategies (dashed).

Results. Let us first provide the aggregate convergence results in Table 1, which almost perfectly reproduce the BNE found in ref. [54]. The utility loss is small in all environments and so is the L_2 difference to the analytical BNE wherever it is known. Figure 1 shows the analytical BNE bid function and the NPGA result for a specific setting as an illustrative example. Note that in the FPSB auction, the global bidder does not have a dominant strategy and yet we uncover his equilibrium strategy in spite of the environment being asymmetric.

Next we look at risk aversion. Figure 2 shows that with higher risk aversion, the market efficiency denoted by $\mathcal{E}$ increases for both correlation models in a similar way. Correlation of the local bidders does not influence $\mathcal{E}$ with the widespread VCG nearest payment rule at a precision of ±1% of $\mathcal{E}$. For the highest level of risk aversion of $\rho = 0.1$, $\mathcal{E}$ rose to about 98% from about 84% under risk neutrality; thus, although higher correlation of valuations does not lead to higher $\mathcal{E}$, risk aversion mitigates the efficiency loss, which is important to know for spectrum sales by governments. A similar result has previously been found for an ascending core-selecting auction with a specific tie-breaking rule[55], but the analysis could not yet be extended to the general sealed-bid case.

Similarly, the approximate revenue of the seller can be analysed. In Figure 3 we observe a strong, steady increase of the seller revenue $\mathcal{R}$ with increasing risk aversion and a slight increase with decreasing correlation between the local bidders. Different levels ρ and varying strengths of γ are plotted in the Bernoulli correlation model in the LLG setting with the NVCG payment rule. Results are similar for the constant weights correlation model. Increasing risk aversion has substantial positive impact on revenue, which is important to know for policymakers.

Discussion

Auction theory—and game theory in general—is often very sensitive to model assumptions. Although the results of early studies on auctions in the symmetric independent private values model with quasilinear bidders provided important insights, the assumptions are very restrictive[56]. Value interdependencies and changes in the

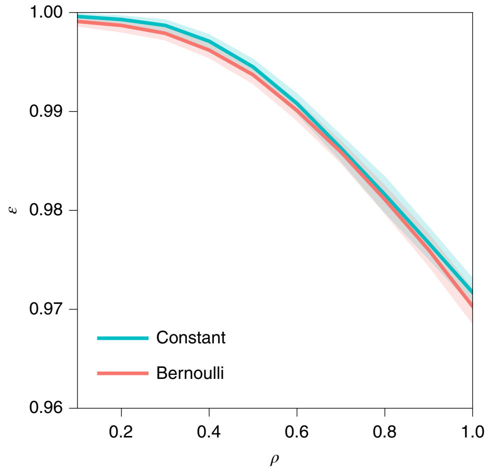


Fig. 2 | The empirical impact of risk-aversion on market efficiency. We depict the market efficiency $\mathcal{E}$ in approximate equilibrium calculated via NPGA for different levels of bidders' risk aversion. The mean (line) and s.d. (shaded bands) of ten runs for each risk-level are depicted.

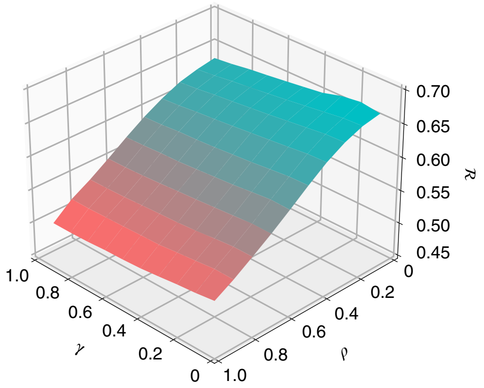


Fig. 3 | The effect of bidder correlation and risk attitudes on seller revenue. The seller revenue $\mathcal{R}$ in approximate equilibrium of LLG auctions with nearest-VCG payments and correlated bidders is shown. The underlying BNEs for each combination of the risk parameter ρ and the correlation strength γ between local bidders have been computed via NPGA.

utility function can have substantial impact on the resulting equilibrium bidding strategies. Although simple single-object auctions in the independent private values model are relatively well understood, we do not know equilibrium bidding strategies for most environments involving multiple objects, interdependencies and different levels of risk aversion to this day.

With NPGA we introduce a numerical technique to compute approximate equilibria in these Bayesian games and show that we converge to a local equilibrium quickly and with high precision. The method can provide a convenient tool for analysts to explore new environments or perform sensitivity analysis with various behavioural assumptions, different priors and value interdependencies. The Supplementary Information provides further experiments to illustrate the versatility of the method.

It is all but clear that gradient dynamics as in NPGA can find global or even local BNEs in auction games. For much simpler min–max games that play an important role in machine learning techniques such as generative adversarial networks, we cannot expect gradient dynamics to find an equilibrium[57]. Convergence of NPGA to approximate local BNEs relies on insights about the symmetry assumptions of bidders in most of the auction models in the literature and their relation to potential games. These assumptions provide the necessary structure for gradient dynamics to converge to local equilibria, and explain our results. Beyond the study of equilibria in games, our techniques can possibly contribute to automated and empirical mechanism design[58,59].

Methods

Proof of Proposition 1. Let G be a symmetric and smooth Bayesian auction game. Per definition, all players in such games have the same marginal type distributions and individual utility functions. Furthermore, assume the auction mechanism to be anonymous: the identity and order of bidders almost surely have no influence on the allocation and payments (tiebreaking on a nullset notwithstanding). Assume that all players play the same strategy β_i. Then, the symmetric ex ante utility function $\tilde{u}_i(\beta_i, \ldots, \beta_i)$ is a potential function and $\tilde{G}$ is a potential game. The same holds for the finite-dimensional proxy game Γ. To use this symmetry, we restrict all players to use the same neural network $\pi(\cdot, \theta)$ with a shared parameter vector $\theta \in \mathbb{R}^d$. Let us first remark that the restriction to symmetric strategies does not alter the gradient vector field in any way, as symmetric strategy profiles also have symmetric gradients.

We draw on a known result that gradient-play with appropriate (summable but not square-summable) step sizes converges almost surely to a local Nash equilibrium in finite-dimensional continuous potential games (see Corollary 4.2 of ref. [33]). It thus remains to be shown that (1) NPGA implements gradient-play in the proxy game Γ and thus finds a local Nash equilibrium θ^* of the proxy game, and (2) that this Nash equilibrium of the proxy game Γ—which restricts the strategy space to neural networks expressible by Θ—is indeed also a BNE of the original unrestricted game G. To show (1) and (2) below, we will rely on some auxiliary lemmata. The proofs of these lemmata are of a technical nature and can be found in Supplementary Section 2. In the following, for a given neural network $\pi(\cdot,\theta)$, we denote its utility and loss in G by $\tilde{u}(\theta)$, $\tilde{\ell}(\theta)$ and in Γ by $\tilde{u}^{\Gamma}(\theta)$, $\tilde{\ell}^{\Gamma}(\theta)$, respectively, where we drop the indices i due to symmetry.

To prove (1), one would need to show that the gradient estimates computed by NPGA have finite variance and at most a small bias with regard to the true gradients of the proxy game Γ. This is not necessarily the case, but let us set $\tilde{u}_i^{\sigma}(\theta_i, \boldsymbol{\theta}_{-i}) \equiv \mathbb{E}_{\varepsilon \sim \mathcal{N}(0,\sigma^2 I)}[\tilde{u}_i(\theta_i + \varepsilon, \boldsymbol{\theta}_{-i})]$ call $\Gamma^{\sigma} = (\mathcal{I}, \Theta, \tilde{\mathbf{u}}^{\sigma})$ the smoothed proxy game and define $\tilde{\ell}^{\sigma}$ analogously. Then Γ^{σ} is likewise a symmetric potential game and we obtain the lemmata described next.

Lemma 1. The gradient estimates ∇^{ES} in NPGA are unbiased and have finite mean squared error with respect to the smoothed utilities $\tilde{u}_i^{\sigma}$ of the game Γ^{σ}.

Lemma 2. For any $\theta \in \Theta$, the loss in Γ is bounded by that in Γ^{σ}:

$$\tilde{\ell}^{\Gamma}(\theta) \leq \tilde{\ell}^{\sigma}(\theta) + 2Z\mathrm{L}\sqrt{d}\sigma$$

where Z is the partial derivative bound from *Definition 1*, d is the number of parameters in the neural network, σ is the standard deviation of the ES perturbations, and the constant L is a property of the neural network architecture π, describing its regularity. By *Lemma 1*, NPGA implements exact gradient play in Γ^{σ} and thus finds a local Nash equilibrium θ^* of that game via the result in ref. [33]. By *Lemma 2*, any Nash equilibrium of Γ^{σ} is an approximate Nash equilibrium of Γ.

For the latter (2), the universal approximation theorem[47] guarantees that a sufficiently large neural network architecture can approximate every $\beta_i \in \Sigma_i$ with arbitrary precision δ. This yields another error bound:

Lemma 3. Let the neural network π be sufficiently expressive, that is for any $\beta_i \in \Sigma_i$ one can find $\theta \in \Theta$ such that $\| \beta_i - \pi(\cdot, \theta)\|_{\Sigma_i} \leq \delta$. Then the loss of θ in G is bounded by that in Γ: $\tilde{\ell}(\theta) \leq \tilde{\ell}^{\Gamma}(\theta) + Z\delta$.

In summary, NPGA almost surely converges to an (approximate) local Nash equilibrium $\boldsymbol{\theta}^*$ of Γ^{σ}, which, by application of local versions of *Lemma 2* and *Lemma 3*, retains a (local) ex ante loss of at most $\kappa = Z(\delta + 2\mathrm{L}\sqrt{d}\sigma)$, thus constituting a κ-BNE of G. In practice, one may choose the parameters δ (via the neural network architecture and size d) and σ sufficiently small such that the error vanishes.

Neural network architecture and hyperparameters. In our implementation, we use fully connected policy networks with two hidden layers of ten nodes each, using SeLU activation in the hidden layers and a ReLU activation function in the output layer. These simple networks are sufficient for the settings here, but even single-layer nets work with a slight decrease in performance. Instead of standard gradient ascent, we apply the Adam optimization algorithm[60] with standard parameters. In each iteration we generate 64 perturbations of the network π_i for ES gradient estimation, using zero-mean Gaussian noise with a standard deviation of $\sigma = 1/d_i$ (as suggested in ref. [49]). We use batch sizes of 2^{17} chosen such that the largest settings would fit into available GPU memory. In the presence of asymmetries or multiple items, degenerate initializations (for example, when some players never win) can impede convergence. To alleviate this and improve comparability, we force close-to-truthful initializations by pre-training the networks towards the truthful strategy using supervised learning (RMSE-loss, 500 steps of vanilla stochastic gradient descent). We did not perform setting-specific hyperparameter tuning to allow for comparable results. There are possibilities to improve the performance of our results when tuning the hyperparameters for a specific environment.

We implemented the auctions using the PyTorch framework[61] with a focus on computing many auctions in parallel. Unless noted otherwise, all experiments were performed on a single consumer-grade Nvidia GeForce RTX 2080Ti GPU with 1,000 iterations for the single-item auctions and 2,000 iterations for the large setting with correlated values ($n=10$) and the multi-unit auctions, where each experiment was run ten times.

Data availability

All data analyses in this study are based exclusively on data generated by our custom simulation framework (see Code Availability). Raw simulation artefacts (all-iteration logs and trained models) will be made available by the corresponding author on request. Source data are provided with this paper.

Code availability

The source code of our simulation framework[62], including instructions to reproduce all models and datasets referenced in this study, is freely available at https://github.com/heidekrueger/bnelearn, licensed under GNU-GPLv3.

Received: 16 December 2020; Accepted: 7 June 2021;
Published online: 9 August 2021

References

1. Brown, N. & Sandholm, T. Superhuman AI for multiplayer poker. *Science* **365**, 885–890 (2019).
2. Daskalakis, C., Ilyas, A., Syrgkanis, V. & Zeng, H. Training gans with optimism. Preprint at https://arxiv.org/abs/1711.00141 (2017).
3. Silver, D. et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362**, 1140–1144 (2018).
4. Daskalakis, C., Goldberg, P. & Papadimitriou, C. The complexity of computing a nash equilibrium. *SIAM J. Comput.* **39**, 195–259 (2009).
5. Brown, G. W in *Activity Analysis of Production and Allocation* (ed. Koopmans, T. C.) 374–376 (Wiley, 1951).
6. Zinkevich, M. Online convex programming and generalized infinitesimal gradient ascent. In *Proc. 20th International Conference on Machine Learning* 928–936 (ICML, 2003).
7. Bowling, M. Convergence and no-regret in multiagent learning. In *Advances in Neural Information Processing Systems* 209–216 (NIPS, 2005).
8. Milgrom, P. R. & Weber, R. J. A theory of auctions and competitive bidding. *Econometrica* **50**, 1089–1122 (1982).
9. Klemperer, P. Auction theory: a guide to the literature. *J. Econ. Surveys* **13**, 227–286 (1999).
10. Vickrey, W. Counterspeculation, auctions, and competitive sealed tenders. *J. Finance* **16**, 8–37 (1961).
11. Krishna, V. *Auction Theory* (Academic, 2009).
12. Bergemann, D. & Morris, S. Robust implementation in direct mechanisms. *Rev. Econ. Stud.* **76**, 1175–1204 (2009).
13. Campo, S., Perrigne, I. & Vuong, Q. Asymmetry in first-price auctions with affiliated private values. *J. Appl. Econom.* **18**, 179–207 (2003).
14. Janssen, M. C. Reflections on the 2020 Nobel Memorial Prize awarded to Paul Milgrom and Robert Wilson. *Erasmus J. Philos. Econ.* **13**, 177–184 (2020).
15. Heinrich, J. & Silver, D. Deep reinforcement learning from self-play in imperfect-information games. Preprint at https://arxiv.org/abs/1603.01121 (2016).
16. Lanctot, M. et al. A unified game-theoretic approach to multiagent reinforcement learning. In *Proc. 31st International Conference on Neural Information Processing Systems* (NIPS, 2017).
17. Brown, N., Lerer, A., Gross, S. & Sandholm, T. Deep counterfactual regret minimization. In *Proc. 36th International Conference on Machine Learning* 793–802 (PMLR, 2019).
18. Brown, N. & Sandholm, T. Superhuman AI for multiplayer poker. *Science* **365**, 885–890 (2019).
19. Reeves, D. M. & Wellman, M. P. Computing equilibrium strategies in infinite games of incomplete information. *Proc. 20th Conference on Uncertainty in Artificial Intelligence* (UAI, 2004).
20. Naroditskiy, V. & Greenwald, A. *Using Iterated Best-response to Find Bayes-Nash Equilibria in Auctions* 1894–1895 (AAAI, 2007).
21. Rabinovich, Z., Naroditskiy, V., Gerding, E. H. & Jennings, N. R. Computing pure Bayesian-Nash equilibria in games with finite actions and continuous types. *Artif. Intell.* **195**, 106–139 (2013).
22. Bosshard, V., Bünz, B., Lubin, B. & Seuken, S. Computing Bayes-Nash equilibria in combinatorial auctions with continuous value and action spaces. In *Proc. 26th International Joint Conference on Artificial Intelligence* 119–127 (IJCAI, 2017).
23. Bosshard, V., Bünz, B., Lubin, B. & Seuken, S. Computing Bayes-Nash equilibria in combinatorial auctions with verification. *J. Artif. Intell. Res.* **69**, 531–570 (2020).
24. Feng, Z., Guruganesh, G., Liaw, C., Mehta, A. & Sethi, A. Convergence analysis of no-regret bidding algorithms in repeated auctions. Preprint at https://arxiv.org/abs/2009.06136 (2020).
25. Li, Z. & Wellman, M. P. Evolution strategies for approximate solution of Bayesian games. In *Proc. AAAI Conference on Artificial Intelligence* Vol. 35, 5531–5540 (AAAI, 2021).
26. Cai, Y. & Papadimitriou, C. Simultaneous Bayesian auctions and computational complexity. In *Proc. 15th ACM Conference on Economics and Computation* 895–910 (ACM, 2014).
27. Fudenberg, D. & Levine, D. K. Learning and equilibrium. *Annu. Rev. Econ.* **1**, 385–420 (2009).
28. Jafari, A., Greenwald, A., Gondek, D. & Ercal, G. On no-regret learning, fictitious play, and Nash equilibrium. *Proc. 18th International Conference on Machine Learning* 226–233 (ICML, 2001).
29. Stoltz, G. & Lugosi, G. Learning correlated equilibria in games with compact sets of strategies. *Games Econ. Behav.* **59**, 187–208 (2007).
30. Hartline, J., Syrgkanis, V. & Tardos, E. No-regret learning in Bayesian games. In *Advances in Neural Information Processing Systems* (eds Cortes, C. et al.) Vol. 28, 3061–3069 (NIPS, 2015); http://papers.nips.cc/paper/6016-no-regret-learning-in-bayesian-games.pdf
31. Foster, D. J., Li, Z., Lykouris, T., Sridharan, K. & Tardos, E. Learning in games: robustness of fast convergence. In *Advances in Neural Information Processing Systems* 4734–4742 (NIPS, 2016).
32. Viossat, Y. & Zapechelnyuk, A. No-regret dynamics and fictitious play. *J. Econ. Theory* **148**, 825–842 (2013).
33. Mazumdar, E., Ratliff, L. J. & Sastry, S. S. On gradient-based learning in continuous games. *SIMODS* **2**, 103–131 (2020).
34. Dütting, P., Feng, Z., Narasimhan, H., Parkes, D. & Ravindranath, S. S. Optimal auctions through deep learning. In *International Conference on Machine Learning* 1706–1715 (PMLR, 2019).
35. Feng, Z., Narasimhan, H. & Parkes, D. C. Deep learning for revenue-optimal auctions with budgets. In *Proc. 17th International Conference on Autonomous Agents and Multiagent Systems* 354–362 (AAMAS, 2018).
36. Tacchetti, A., Strouse, D., Garnelo, M., Graepel, T. & Bachrach, Y. A neural architecture for designing truthful and efficient auctions. Preprint at https://arxiv.org/abs/1907.05181 (2019).
37. Weissteiner, J. & Seuken, S. Deep learning-powered iterative combinatorial auctions. In *Proc. AAAI Conference on Artificial Intelligence* Vol. 34, 2284–2293 (AAAI, 2020).
38. Morrill, D. et al. Hindsight and sequential rationality of correlated play. Preprint at https://arxiv.org/abs/2012.05874 (2020).
39. Hartford, J. S. *Deep Learning for Predicting Human Strategic Behavior*. Ph.D. thesis, Univ. British Columbia (2016).
40. Ghani, R. & Simmons, H. Predicting the end-price of online auctions. In *Proc. International Workshop on Data Mining and Adaptive Modelling Methods for Economics and Management* (CiteSeer, 2004).
41. Zheng, S. et al. The AI economist: improving equality and productivity with AI-driven tax policies. Preprint at https://arxiv.org/abs/2004.13332 (2020).
42. Goeree, J. K. & Lien, Y. On the impossibility of core-selecting auctions. *Theoretical Econ.* **11**, 41–52 (2016).
43. Bichler, M. & Goeree, J. K. *Handbook of Spectrum Auction Design* (Cambridge Univ. Press, 2017).
44. Debnath, L. et al. *Introduction to Hilbert Spaces with Applications* (Academic, 2005).
45. Bichler, M., Guler, K. & Mayer, S. Split-award procurement auctions-can bayesian equilibrium strategies predict human bidding behavior in multi-object auctions? *Prod. Oper. Manag.* **24**, 1012–1027 (2015).
46. Ui, T. Bayesian nash equilibrium and variational inequalities. *J. Math. Econ.* **63**, 139–146 (2016).
47. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Networks* **4**, 251–257 (1991).
48. Wierstra, D. et al. Natural evolution strategies. *J. Mach. Learn. Res.* **15**, 949–980 (2014).
49. Salimans, T., Ho, J., Chen, X., Sidor, S. & Sutskever, I. Evolution strategies as a scalable alternative to reinforcement learning. Preprint at https://arxiv.org/abs/1703.03864 (2017).
50. Benaím, M., Hofbauer, J. & Sorin, S. Perturbations of set-valued dynamical systems, with applications to game theory. *Dyn. Games Appl.* **2**, 195–205 (2012).
51. Letcher, A. et al. Differentiable game mechanics. *J. Mach. Learn. Res.* **20**, 1–40 (2019).
52. Monderer, D. & Shapley, L. S. Potential games. *Games Econ. Behav.* **14**, 124–143 (1996).
53. Bünz, B., Lubin, B. & Seuken, S. Designing core-selecting payment rules: a computational search approach. In *Proc. 2018 ACM Conference on Economics and Computation* 109 (ACM, 2018).

54. Ausubel, L. M. & Baranov, O. Core-selecting auctions with incomplete information. *Int. J. Game Theory* **49**, 251–273 (2019).
55. Guler, K., Bichler, M. & Petrakis, I. Ascending combinatorial auctions with risk averse bidders. *Group Decis. Negot.* **25**, 609–639 (2016).
56. Jehiel, P., Meyer-ter-Vehn, M., Moldovanu, B. & Zame, W. R. The limits of ex post implementation. *Econometrica* **74**, 585–610 (2006).
57. Daskalakis, C., Skoulakis, S. & Zampetakis, M. The complexity of constrained min–max optimization. In *Pro. 53rd Annual ACM SIGACT Symposium on Theory of Computing* 1466–1478 (STOC, 2021).
58. Vorobeychik, Y., Reeves, D. M. & Wellman, M. P. Constrained automated mechanism design for infinite games of incomplete information. In *Proc. 23rd Conference on Uncertainty in Artificial Intelligence* 400–407 (UAI, 2007).
59. Viqueira, E. A., Cousins, C., Mohammad, Y. & Greenwald, A. Empirical mechanism design: designing mechanisms from data. In *Proc. 35th Uncertainty in Artificial Intelligence Conference* 1094–1104 (PMLR, 2020).
60. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2015).
61. Paszke, A. et al. Automatic differentiation in pytorch. In *31st Conference on Neural Information Processing Systems* (NIPS, 2017).
62. Heidekrüger, S., Kohring, N., Sutterer, S. & Bichler, M. *bnelearn: A Framework for Equilibrium Learning in Sealed-Bid Auctions* (Github, 2021); https://github.com/heidekrueger/bnelearn

Acknowledgements

We are grateful for funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; grant no. BI 1057/1-9). We thank V. Bosshard, B. Lubin, P. Mertikopoulos, P. Milgrom, S. Seuken, T. Ui, F. Maldonado and participants of the NBER Market Design Workshop 2020 for valuable feedback on earlier versions.

Author contributions

M.B. conceived and supervised the project and contributed to the overall study design, theoretical analysis of NPGA and writing the manuscript. M.F. contributed to the theoretical analysis of NPGA. S.H. contributed to the design, implementation and optimization of the algorithm and simulation framework, the theoretical analysis, and to the writing of the manuscript. N.K. contributed to the optimization of the algorithm, the design, implementation and optimization of the simulation framework, the theoretical and empirical analysis, and the writing of the manuscript. P.S. contributed to design and implementation of the algorithm and simulation framework, and the empirical analysis.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42256-021-00365-4.

Correspondence and requests for materials should be addressed to M.B.

Peer review information *Nature Machine Intellligence* thanks Pierre Baldi, Neil Newman and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 5

Zeroth-Order Learning in Asymmetric Markets

Peer-Reviewed Journal Paper

Title: Learning equilibria in asymmetric auction games.

Authors: Martin Bichler, Stefan Heidekrüger, Nils Kohring.

In: INFORMS Journal on Computing.

Abstract: Computing Bayesian Nash equilibrium strategies in auction games is a challenging problem that is not well understood. Such equilibria can be modeled as systems of nonlinear partial differential equations. It was recently shown that Neural Pseudogradient Ascent (NPGA), an implementation of simultaneous gradient ascent via neural networks, converges to a Bayesian Nash equilibrium for a wide variety of symmetric auction games. While symmetric auction models are widespread in the theoretical literature, in most auction markets in the field one can observe different classes of bidders having different valuation distributions and strategies. Asymmetry of this sort is almost always an issue in real-world multi-object auctions, where different bidders are interested in different packages of items. Such environments require a different implementation of NPGA with multiple interacting neural networks having multiple outputs for the different allocations the bidders are interested in. In this paper, we analyze a wide variety of asymmetric auction models. Interestingly, our results show that we closely approximate Bayesian Nash equilibria in all models where the analytical Bayes-Nash equilibrium is known. Additionally, we analyze new and larger environments for which no analytical solution is known and verify that the solution found approximates equilibrium closely. The results provide a foundation for generic equilibrium solvers that can be used in a wide range of auction games.

Comment: This publication is a featured article of its IJOC issue.

Citation: Bichler et al. (2023).

CCC Marketplace

This is a License Agreement between Nils Kohring ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	11-Apr-2023	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1343827-1	**Publisher**	INFORMS
ISSN	1526-5528	**Portion**	Chapter/article

LICENSED CONTENT

Publication Title	INFORMS journal on computing	**Language**	English
Article Title	Learning Equilibria in Asymmetric Auction Games	**Country**	United States of America
Author/Editor	Institute for Operations Research and the Management Sciences.	**Rightsholder**	The Institute for Operations Research and the Management Sciences (INFORMS)
Date	01/01/1996	**Publication Type**	e-Journal

REQUEST DETAILS

Portion Type	Chapter/article	**Rights Requested**	Main product
Page Range(s)	1-20	**Distribution**	Worldwide
Total Number of Pages	20	**Translation**	Original language of publication
Format (select all that apply)	Electronic	**Copies for the Disabled?**	No
Who Will Republish the Content?	Author of requested content	**Minor Editing Privileges?**	No
Duration of Use	Life of current edition	**Incidental Promotional Use?**	No
Lifetime Unit Quantity	Up to 499	**Currency**	EUR

NEW WORK DETAILS

Title	Multi-Agent Reinforcement Learning for the Computation of Market Equilibria	**Institution Name**	Technical University of Munich
Instructor Name	Prof. Dr. Martin Bichler	**Expected Presentation Date**	2023-09-01

ADDITIONAL DETAILS

Order Reference Number	N/A	**The Requesting Person/Organization to Appear on the License**	Nils Kohring

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Full article	**Title of the Article/Chapter the Portion Is From**	Learning Equilibria in Asymmetric Auction Games
Editor of Portion(s)	N/A	**Author of Portion(s)**	Bichler, Martin; Kohring, Nils; Heidekrüger, Stefan
Volume of Serial or Monograph	N/A	**Issue, if Republishing an Article From a Serial**	N/A
Page or Page Range of Portion	1-20	**Publication Date of Portion**	2023-09-01

Marketplace Permissions General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such “clipping” or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit "publishing ventures" where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- a statement to the effect that such copy was made pursuant to permission,
- a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and
- a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service.

Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery).*** The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or “clipping” does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Last updated October 2022

INFORMS Journal on Computing

Learning Equilibria in Asymmetric Auction Games

Martin Bichler*, Stefan Heidekrüger, Nils Kohring
Technical University of Munich, Department of Computer Science, 85748 Garching, Germany

bichler@in.tum.de

Computing Bayesian Nash equilibrium strategies in auction games is a challenging problem that is not well understood. Such equilibria can be modeled as systems of nonlinear partial differential equations. It was recently shown that Neural Pseudogradient Ascent (NPGA), an implementation of simultaneous gradient ascent via neural networks, converges to a Bayesian Nash equilibrium for a wide variety of symmetric auction games. While symmetric auction models are widespread in the theoretical literature, in most auction markets in the field one can observe different classes of bidders having different valuation distributions and strategies. Asymmetry of this sort is almost always an issue in real-world multi-object auctions, where different bidders are interested in different packages of items. Such environments require a different implementation of NPGA with multiple interacting neural networks having multiple outputs for the different allocations the bidders are interested in. In this paper, we analyze a wide variety of asymmetric auction models. Interestingly, our results show that we closely approximate Bayesian Nash equilibria in all models where the analytical Bayes-Nash equilibrium is known. Additionally, we analyze new and larger environments for which no analytical solution is known and verify that the solution found approximates equilibrium closely. The results provide a foundation for generic equilibrium solvers that can be used in a wide range of auction games.

Key words: equilibrium learning, neural networks, Bayes-Nash equilibria

1. Introduction

Auction theory is arguably the best-known and practically most relevant application of Bayesian game theory, central to modern economic theory (Klemperer 2000) and with a multitude of applications in the field, ranging from industrial procurement to treasury auctions and spectrum sales (Krishna 2009, Milgrom 2017, Bichler and Goeree 2017). The derivation of Bayesian Nash equilibrium strategies (BNE) for the first-price and second-price sealed-bid auction led to a comprehensive theoretical framework for the analysis of single-item auctions by Nobel laureate William Vickrey, a landmark result of economic theory (Vickrey 1961). Also, the Nobel Prize in Economic Sciences 2020 to Paul Milgrom and Robert Wilson was awarded for contributions to auction theory. However, while single-item auctions are well understood and closed-form BNE strategies are known for a variety of auction formats, we only know equilibrium strategies for a few restricted multi-item auction environments with heterogeneous goods. Even for uniform or discriminatory multi-unit auctions with homogeneous goods and symmetric bidders, we can only characterize properties of the Bayes-Nash equilibrium but do not have a general closed-form solution (Krishna 2009). So, the realm of auction markets where we know a Bayes-Nash equilibrium is very limited.

Equilibrium computation is well-known to be hard even for simple, finite, complete-information games: Finding Nash equilibria in normal-form games is known to be in the complexity class PPAD[1] (Daskalakis et al. 2009). Mathematically, auctions are typically described as Bayesian games. Bidders' valuations are considered samples from some continuous and atomless prior valuation distribution and their strategies are represented by continuous bid functions mapping these valuations to bids. Vickrey (1961) have enabled a deep understanding of common single-item auction formats. However, there still remain many open questions for more involved multi-item auctions such as *combinatorial auctions*, in which players bid on *bundles* of multiple goods simultaneously. We also know little about the existence of Bayesian Nash equilibria in such auction games (Jackson and Swinkels 2005). Importantly, the computational complexity of computing BNE is hardly understood. Typically, for a fully specified setting, we can model the equilibrium problem as systems of nonlinear partial differential equations for which no exact solution theory is known (Klainerman 2010). Given the relevance of auctions, understanding their equilibria is crucial, and numerical methods for computing or approximating such steady states would be a significant step forward in the theory of auctions and also in their design and in applications.

This paper can be viewed in the context of equilibrium learning via gradient dynamics. Whether learning agents' strategies in repeated games converge to equilibria has been studied for complete-information normal-form games (Fudenberg and Levine 2009). In contrast, equilibrium learning in Bayesian auction games is largely unexplored (see Section 2). First, the ex-post utility function is non-differentiable in auctions, which makes it difficult to apply gradient dynamics. Secondly, it is a known fact that multi-agent gradient dynamics do not converge in general games: Convergence to Nash equilibria has only been

[1] The class of *Polynomial Parity Arguments on Directed graphs* (PPAD) problems is believed to be hard and is related to NP.

established for restricted classes of complete-information normal-form games. In summary, it is all but clear how gradient dynamics would be implemented in Bayesian auction games, and even if this was done, whether the algorithm would converge to a BNE in auction games. We draw on Bichler et al. (2021), who recently introduced Neural Pseudogradient Ascent (NPGA), an algorithm that relies on simultaneous gradient ascent of bidders with respect to their ex-ante utility functions. More specifically, NPGA models all players' bidding strategies as neural networks, and trains them via self play based on approximate ex-ante gradients computed from observations of the discontinuous ex-post utility function using evolutionary strategies. It can be applied to a wide range of Bayesian auction games, since it does not require any auction-specific sub-procedures beyond access to simulating auction outcomes. Likewise, its computational steps can exploit massive parallelization and GPU hardware acceleration.

The results by Bichler et al. (2021) focus on *symmetric* auction models, assuming symmetric prior distributions and symmetric equilibrium bidding strategies of the bidders. This allows them to train only a single neural network to provably find the symmetric equilibrium bidding strategy. While symmetric models cover some important auctions in the theoretical literature, many interesting environments include asymmetries. For example, asymmetric priors are a concern for single-object auctions with strong and weak bidders drawn from different distributions, but they are even more prevalent in multi-item auctions where it is unlikely that bidders are interested in the same items with their values drawn from the same distribution. It is also these more general market environments for which the literature on auction theory does not provide analytical equilibrium predictions.

1.1. Contributions

In this paper, we explore a number of challenging environments, models which clearly violate the symmetry assumption. Nothing is known about the convergence and speed of equilibrium learning in such environments where one needs to train multiple neural networks with multi-dimensional outputs modeling different actions of bidders. We show that the NPGA algorithm also converges with multiple neural networks which are required to model asymmetric environments. We explore a wide range of wicked models from the literature where the BNE is known analytically and find that NPGA computes a very close approximation of the Bayes-Nash equilibrium in all of them. In addition, we explore large environments where no analytical solution is known and we can verify empirically that a close approximation to a BNE is found.

We start with a single-object auction with two asymmetric priors (Plum 1992). Apart from this original model, we also analyze one that allows for (rational) overbidding and admits multiple equilibria (Kaplan and Zamir 2015). Here, we only know closed-form equilibrium strategies for uniform prior distributions, for which NPGA finds a BNE. However, we also discuss a specific model with two bidders competing for a single object where the valuations are drawn from a non-linear beta distribution. No equilibrium strategy is known, but we find bidding strategies with a very low estimated utility loss for all players. This indicates that the computed strategy profile is a close approximation of a BNE.

Second, we explore a specific type of multi-unit uniform-price auction of homogeneous goods with two classes of bidders, those with a high and with a low type. The environment is very large with up to 12 units and NPGA is able to compute a sufficiently close equilibrium in under five minutes. Such mechanisms are used in treasury bill auctions but also electricity markets. Demand reduction is a well-known phenomenon in such auctions and it is interesting to observe how it plays out under different model assumptions about the strength of the competitors.

Third, we analyze a combinatorial auction in the well-known local-local-global (LLG) model. The model has two items and three bidders and it has become a standard environment to discuss spectrum auction design and more generally combinatorial auctions (Bichler and Goeree 2017). Two local bidders want to win one item each and they compete against a global bidder interested in the package of both items. Bidders are assumed to only bid for the single item for which they have a strictly positive expected valuation. In this standard LLG model, the local bidders are assumed to have symmetric priors, and NPGA converges quickly to the BNE strategy (Bichler et al. 2021). In contrast to this standard setting, we analyze a variant where one of the local bidders is favored and bidders are not precluded a priori from bidding on bundles for which they do not have a strictly positive value. While a bidder would not actually be interested in *winning* such a bundle, it turns out that sometimes it may nevertheless be rational to submit a positive bid for it. Ott and Beck (2013) showed that, in fact, this version of the local-local-global model has an equilibrium where the second local bidder bids on the package of both items and even overbids—in spite of being interested in a single item only. Such equilibrium strategies are not obvious. Again, we find that NPGA recovers this analytical solution with high precision.

Fourth, we experiment with another reverse combinatorial auction model with two homogeneous objects and two bidders. This model is interesting because there are two pure BNE (Anton and Yao 1992, Kokott et al. 2019). Similar to the analysis of the asymmetric single-object environment (Kaplan and Zamir 2015), NPGA finds an equilibrium, which is also the efficient one.

Finally, we report the results for a large combinatorial auction model with six bidders, belonging to two symmetry classes, and eight items, which has recently been proposed as a challenging problem for equilibrium computation and which, to the authors' knowledge, is the largest combinatorial auction for which an approximate BNE has been computed numerically with a setting-specific algorithm (Bosshard et al. 2020). Going beyond the existing challenge model, we also study NPGA in an even larger extension by introducing an additional seventh bidder belonging to a new third symmetry class. In both these settings, strategy profiles learned by NPGA converge to approximate BNE. Such environments can already be considered very large and beyond what is typically analyzed in auction theory.

Overall, The empirical results we show in this paper provide evidence that gradient dynamics implemented in NPGA are significantly more powerful than expected and they converge in a much wider range of (asymmetric) auction games. This raises hope that gradient dynamics can be used to compute equilibria in a much broader variety of market models and that general auction equilibrium solvers are in reach.

1.2. Organization

In the next section, we discuss related literature. Section 3 introduces preliminaries and notation before we discuss gradient dynamics in the context of auctions in Section 4. Section 5 introduces metrics to evaluate the quality of our results before we report our results in Section 6. Finally, we provide a summary and conclusions in Section 7. The source code and configurations can be found at the repository (Bichler et al. 2023).

2. Related Literature

In what follows, we survey existing hardness results, approaches to equilibrium learning, and initial research on computing approximate Bayes-Nash equilibria.

2.1. Hardness of Equilibrium Computation

The computation of Nash equilibria received significant attention after the initial contribution by John Nash on the existence of such equilibria in complete-information normal-form games (Nash et al. 1950). However, it was shown that the problem is already PPAD-complete for two-agent normal-form games (Daskalakis et al. 2009) and it is hard to approximate (Rubinstein 2016). The computation of Nash equilibria for three or more agents is even FIXP-complete, i.e., complete for the class of search problems that can be cast as fixed-point computation problems (Etessami and Yannakakis 2007).

Determining whether a pure-strategy BNE exists in a finite Bayesian game is NP-complete and these hardness results also hold if there are only two agents and the game is symmetric (Conitzer and Sandholm 2008). Finding a mixed Bayesian equilibrium in a Bayesian game is, of course, PPAD-hard, but might be even harder; however, little is known in general. As indicated in the introduction, Cai and Papadimitriou (2014) show that finding a BNE in simultaneous single-item Vickrey auctions for which the bidders have combinatorial valuations is hard for the class PP (the decision version of ♯P), which is much harder than NP. Even certifying a BNE is PP-hard, which casts doubt on the question of whether BNE can be at all predictive in the field. Additionally, the authors show that it is even NP-hard to find an approximate BNE in the simultaneous Bayesian auction game. Note that environments with continuous action space are not finite games, and the existence result by Nash does not carry over. We are not aware of proof that a possibly mixed Bayesian equilibrium always exists in such games. Athey (2001) showed conditions for pure BNE to exist, Carbonell-Nicolau and McLean (2018) provided conditions that guarantee the existence of a BNE, while Ui (2016) characterized strong payoff-monotonicity as a sufficient condition for uniqueness of BNE in ex-post differentiable continuous-action Bayesian games.

2.2. Equilibrium Learning

Our research is best situated in the literature on equilibrium learning (Fudenberg and Levine 2009). Learning in complete-information normal form games has a long history and has been extensively studied in game theory and, more recently, multi-agent reinforcement learning. One class of methods is formed by *best response dynamics.* The earliest such method, published by Cournot in 1838, has agents play a pure strategy best response against other agents' strategies used in the previous iteration. In Fictitious Play (FP)

(Brown 1951), a best response is instead played against the strategy profile induced by opponents' empirical frequencies of play in all previous iterations. Whenever the *empirical frequencies* of FP converge, the limit constitutes a Nash equilibrium, but the actual (last-iteration) play only converges in special cases of normal form games such as potential games (Monderer and Shapley 1996).

Gradient dynamics constitute another class of equilibrium learning algorithms. Generalized infinitesimal gradient ascent (GIGA) (Zinkevich 2003) or GIGA-WoLF (Bowling 2005) are examples of gradient dynamics in normal form games, where in each iteration, for each agent we move a step along the direction of the utility gradient and then project the resulting point back to the set of feasible mixed strategies. If aggregating over the stages of the process, the agent's regret grows sublinearly, then there is "no regret" asymptotically. GIGA's total regret is $O(\sqrt{T})$, where T is the number of steps in a repeated strategic game. Hazan et al. (2007) have given an algorithm with a total regret of $O(\log(T))$. Complete-information games with continuous action spaces and smooth utility functions have also received some attention in the context of generative adversarial networks (Letcher et al. 2019, Balduzzi et al. 2018, Schäfer and Anandkumar 2019). A common observation in this line of research is that gradient-based learning does not necessarily converge to an equilibrium and may even exhibit cycling or chaotic behavior. However, it often achieves no-regret properties and thereby converges to a weaker form of equilibrium, so called coarse correlated equilibria (CCE). Similar conclusions were drawn for finite-type (and possibly continuous-action) Bayesian games. Here, no-regret learners were shown to converge to Bayesian CCEs (Hartline et al. 2015).

Gradient dynamics are only known to converge to a Nash equilibrium in certain types of normal-form games such as potential games, bilinear games (Singh et al. 2000), and convex games (Mertikopoulos and Zhou 2019). Letcher et al. (2019) explore gradient dynamics in complete-information continuous-action *differential games*. If ex-post payoffs are twice continuously-differentiable, they find properties such that gradient dynamics converge to at least *local equilibria*. Unfortunately, the ex-post utility in our auction games is not differentiable. More importantly, these techniques are defined for complete-information games with finite-dimensional action spaces while we search for strategies over a function space. Unfortunately, a thorough understanding of the convergence and limiting behaviors in general, continuous games is missing. Actually, the analysis of gradient dynamics, in general, can be arbitrarily complex (Andrade et al. 2021).

2.3. Algorithms for Computing Approximate BNE

Earlier approaches to compute approximate BNE in auctions either comprised solving the set of nonlinear differential equations resulting from the first-order conditions of simultaneous maximization of the bidders' payoffs (Marshall et al. 1994, Bajari 2001) or of restricting the action space, e.g., through discretization (Athey 2001). Then, however, one has no guarantees on the quality of the corresponding ϵ-BNE of the original auction game. Armantier et al. (2008) introduced a BNE-computation method that is based on expressing the Bayesian game as the limit of a sequence of complete-information games, but defining this sequence requires setting-specific analysis.

Numerical BNE in more complex combinatorial auctions were first computed by Bosshard et al. (2017, 2020) in two recent papers; in particular, they study the LLG and LLLLGG markets, both of which are also analyzed in this paper. Their algorithm computes point-wise best responses in a linearization of the strategy space via Monte Carlo integration. They prove an an upper bound ϵ on the interim utility loss achieved by their algorithm using a verification method that assumes identical independent priors ($F_{v_i|v_{-i}} = F_{v_i}$) and risk-neutral attitudes of all bidders. High worst-case interim precision comes at a computational cost for more complex environments with multi-minded bidders.

NPGA (Bichler et al. 2021) follows a different approach and is rooted in gradient dynamics rather than best response dynamics. It directly learns the bid functions expressed across the entire value space (as opposed to point-wise) by updating the parameters of the neural networks via ex-ante gradient ascent. NPGA neither requires discretization of the value or action space as in Athey (2001) nor does it rely on twice differentiable payoff or loss functions as required in the literature on differentiable games (Singh et al. 2000, Letcher et al. 2019). Further, it makes no assumptions about the risk attitude or independence of the bidders' valuations. For symmetric auctions, Bichler et al. (2021) show that NPGA converges (at least) to local BNE.

3. Problem Statement and Notation

We next introduce the necessary notation and concepts from Bayesian game theory relevant to our paper.

3.1. Auctions as Bayesian Games

A *Bayesian game* or game with *incomplete information* is defined by the quintuple $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$. The set of players is denoted by $\mathcal{I} = \{1, \dots, n\}$, $\mathcal{A} \equiv \mathcal{A}_1 \times \dots \times \mathcal{A}_n$ is the set of possible action profiles, where agent $i \in \mathcal{I}$ has access to the action set $\mathcal{A}_i$. $\mathcal{V} \equiv \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ we denote the set of *type profiles* and $F\colon \mathcal{V} \to [0,1]$ defines the joint probability distribution over $\mathcal{V}$ that is known to all players. Throughout this paper, F_X will denote the cumulative distribution function of a random variable X. For example, F_{v_i} will denote the marginal distribution of player i's type. In each game, a type profile $v \sim F$ is drawn and all agents i are privately informed of their own types v_i. Based on this private information, each player must then choose an action b_i from $\mathcal{A}_i$. After actions have been chosen, every player will observe their *ex-post* utility according to a function $u_i : \mathcal{A} \times \mathcal{V}_i \to \mathbb{R}$ that notably depends on all agents' actions but only on i's own type.

This paper considers not only sealed-bid auctions of a single object, but also multi-unit auctions and combinatorial auctions with m heterogeneous items, $\mathcal{M} = \{1, \ldots, m\}$. In these auctions, each agent, also called *bidder*, is allocated a bundle $x_i \in \mathcal{K} \equiv 2^{\mathcal{M}}$ of items (possibly $x_i = \varnothing$). In the *private value* setting most commonly studied in auction theory, types $v_i \in \mathcal{V}_i$ can then be interpreted as a vector of *private valuations* that is composed of the valuations the bidder has for all possible bundles: $v_i \equiv (v_i(k))_{k \in \mathcal{K}}$. For a treatment beyond private values (e.g., interdependent bidder types) we refer the interested reader to Bichler et al. (2021). Bidders map these valuations to their individual bids $b_i = \beta_i(v_i)$ according to some *pure strategy* or *bid function* $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$. In line with most work in auction theory, we will focus on pure strategies that choose a specific action with certainty.

In an exclusive-OR (XOR) bid language, a bidder submits bids for every possible bundle but can only win one of the bids. This means that bids are generally in $\mathcal{A}_i \subseteq \mathbb{R}_+^{|\mathcal{K}|}$, and every player must thus submit a total of 2^m scalar bids.

By $\Sigma_i \subseteq \mathcal{A}_i^{\mathcal{V}_i}$ we denote the strategy space of bidder i and by $\Sigma \equiv \prod_i \Sigma_i$ the space of available joint strategies. Note that the spaces Σ_i are infinite-dimensional as a consequence of infinite $\mathcal{V}_i$.

The auctioneer then applies an *auction mechanism* which will determine an allocation x and a price vector p. The allocation determines the bundles of goods $x_i \in \mathcal{K}$ received by each bidder which must be disjoint: $x_i \cap x_j = \varnothing$. Payments $p \in \mathbb{R}^n$ determine a scalar amount of money that each payer will have to pay to the auctioneer in exchange for receiving the bundle x_i. We will rely on the standard environment in auction theory where bidders have a *quasi-linear* utility function given by $u_i : \mathcal{V}_i \times \mathcal{A} \to \mathbb{R}$,

$$u_i(v_i, b_i, b_{-i}) = v_i(x_i) - p_i, \tag{3.1}$$

where the index $-i$ denotes a profile of types, actions, or strategies for all agents but agent i. That is, each bidder's utility is given by her valuation of her allocated bundle, minus the payment she has to make. Quasi-linear utilities correspond to risk-neutral bidders. Note that NPGA is not restricted to the risk-neutral setting. (See e.g. Bichler et al. (2021) or Ewert et al. (2022) for applications in the presence of risk-averse agents.) However, the environments discussed in this paper assume quasi-linear utility, which simplifies notation. We will differentiate between the *ex-ante*, *interim*, and *ex-post* states of the game, where bidders first know only F, then additionally their valuations $v_i \sim F_{v_i}$, and finally also the observed utility $u_i(v_i, b)$, respectively.

3.2. Bayes-Nash Equilibrium

The notion of Nash equilibria (NE) is the central equilibrium solution concept in noncooperative game theory. An action profile b^* is a pure-strategy NE of the complete-information game $G = (\mathcal{I}, \mathcal{A}, u)$ iff no player has any incentive to deviate unilaterally while other agents adhere to the equilibrium: $u_i(b_i^*, b_{-i}^*) \geq u_i(b_i, b_{-i}^*)$ for all $b_i \in \mathcal{A}_i$ and all $i \in \mathcal{I}$. Bayesian Nash equilibria (BNE) generalize this concept to incomplete-information games. To do so, we will need to consider the expected *interim utility* $\overline{u}_i$ of i of a given bid choice $b_i \in \mathcal{A}_i$ over the conditional distribution of opponent valuations v_{-i}, given i's observed type v_i and assuming opponents play fixed strategies $\beta_{-i} \in \Sigma_{-i}$:

$$\overline{u}_i(v_i, b_i, \beta_{-i}) \equiv \mathbb{E}_{v_{-i}|v_i}[u_i\left(v_i, b_i, \beta_{-i}(v_{-i})\right)], \tag{3.2}$$

In our analysis, we will also use the *interim utility loss* of action b_i that is incurred, in hindsight, by not playing a best response action. Given v_i and β_{-i} it is defined as

$$\overline{\ell}_i(b_i; v_i, \beta_{-i}) = \sup_{b_i' \in \mathcal{A}_i} \overline{u}_i(v_i, b_i', \beta_{-i}) - \overline{u}_i(v_i, b_i, \beta_{-i}). \tag{3.3}$$

Typically, $\overline{\ell}_i$ is not actually observable to any agent because it requires knowledge of (a) the opponents' strategies and (b) a corresponding best response.

An interim *ϵ-Bayesian Nash Equilibrium (ϵ-BNE)* is a strategy profile $\beta^* = (\beta_1^*, \dots, \beta_n^*) \in \Sigma$ in which no deviation could yield an interim utility improvement of more than $\epsilon \geq 0$ for any player. Formally, an ϵ-BNE is described as follows:

$$\overline{\ell}_i\left(b_i; v_i, \beta_{-i}^*\right) \leq \epsilon \quad \text{for all } i \in \mathcal{I}, v_i \in \mathcal{V}_i, \text{ and } b_i \in \mathcal{A}_i. \tag{3.4}$$

In a true BNE, where $\epsilon = 0$, every bidder's strategy maximizes her expected interim utility everywhere on her type space $\mathcal{V}_i$ given the opponents' strategies. While this *interim* stage definition of BNE is most common in the literature, we will instead focus on *ex-ante* Bayesian equilibria as strategy profiles that concurrently maximize each player's *ex-ante* expected utility $\tilde{u}$, i.e., at the stage where only the priors F are known, but players have not yet learned their own private valuation. We thusly define $\tilde{u}$ and the *ex-ante utility losses* $\tilde{\ell}$ of a strategy profile $\beta \in \Sigma$ by

$$\tilde{u}_i(\beta_i, \beta_{-i}) \equiv \mathbb{E}_v[u_i(v_i, \beta_i(v_i), \beta_{-i}(v_{-i}))] \tag{3.5}$$

$$= \mathbb{E}_{v_i \sim F_{v_i}}[\overline{u}_i(v_i, b_i, \beta_{-i})], \tag{3.6}$$

and

$$\tilde{\ell}_i(\beta_i, \beta_{-i}) \equiv \sup_{\beta_i' \in \Sigma_i} \tilde{u}_i(\beta_i', \beta_{-i}) - \tilde{u}_i(\beta_i, \beta_{-i}). \tag{3.7}$$

Ex-ante BNE strategy profiles $\beta^* \in \Sigma$ can be characterized by the equations $\tilde{\ell}_i(\beta_i^*, \beta_{-i}^*) = 0$ for all $i \in \mathcal{I}$. Note that interim BNE also constitute an ex-ante equilibria and the reverse holds almost everywhere: every ex-ante equilibrium fulfills Equation 3.4, except possibly on a set of type profiles with measure 0 under F. In this paper, we concern ourselves with finding ex-ante equilibria of auction games.

4. Neural Pseudogradient Ascent

In this section, we introduce Neural Pseudogradient Ascent (NPGA), an algorithm that was recently introduced by Bichler et al. (2021) for Bayesian games with continuous type- and action-spaces. We briefly summarize the algorithm for the paper to be self-contained before we discuss issues around computational hardness and scalability.

4.1. The Algorithm

Intuitively, NPGA simply follows the ex-ante gradient dynamics of the game. However, computing these dynamics is not trivial for auctions, where the ex-post utility functions have discontinuities. Suppose that in each iteration of the learning algorithm players have access to a gradient-oracle $\nabla_{\beta_i} \tilde{u}_i(\beta_i, \beta_{-i})$ with respect to the current joint strategy profile β^t. Then the gradient dynamics would require that each player perform a projected gradient update:

$$\beta_i^t \equiv \mathcal{P}_{\Sigma_i}\left(\beta_i^{t-1} + \Delta_i^t\right) \qquad \text{where} \qquad \Delta_i^t \propto \nabla_{\beta_i} \tilde{u}_i(\beta_i, \beta_{-i}), \tag{4.1}$$

where $\mathcal{P}_{\Sigma_i}(\,\cdot\,)$ projects its argument onto the set of feasible strategies. Some nuances of Equation 4.1 deserve discussion: Importantly, the gradient dynamics are to be understood with respect to the *ex-ante* utility $\tilde{u}$, rather than interim or ex-post utilities. As such, any update iteration aims to marginally improve on the player i's expected utility across all possible type realizations of the game. Furthermore, when computing the gradient oracle $\nabla_\beta \tilde{u}$ via self-play, one may need to rely on access to other players' strategies, but evaluating each player's policy requires only on their own valuation. Finally, $\beta_i \in \Sigma_i$ are functions in an infinite-dimensional function space, so the gradient $\nabla_{\beta_i} \tilde{u}_i$ is itself a *functional* derivative. We formally consider this to be the Gateaux derivative, a generalization of the directional derivative in Euclidean spaces, over the Hilbert space Σ_i equipped with the inner product $\langle \psi, \beta_i \rangle = \mathbb{E}_{v_i \sim F_{v_i}}\left[\psi(v_i)^T \beta_i(v_i)\right]$. This choice of space specifies the projection operation in Equation 4.1 to $\mathcal{P}_{\Sigma_i}(\beta) \equiv \arg\min_{\sigma \in \Sigma_i} \langle \sigma - \beta, \sigma - \beta \rangle$.

To implement these gradient updates in practice, NPGA considers all bidders' strategies to be policy networks $\beta_i(v_i) \equiv \pi_i(v_i; \theta_i)$ specified by some neural network architecture and parameters $\theta_i \in \mathbb{R}^{d_i}$. Importantly, when a suitable neural network architecture is chosen, all relevant θ_i will yield feasible bids, and the projection operation in the update can be neglected as a result. In the empirical part of this study, we restrict ourselves to fully-connected feed-forward neural networks with SeLU activations in the hidden layers (Klambauer et al. 2017) and ReLU activations in the output layer. The latter guarantees satisfaction of nonnegativity of the bids – the only feasibility constraint in the auctions studied below. Note that in contrast to Bichler et al. (2021), we analyze more complex auction models with multi-minded bidders, such that the output layer includes multiple neurons defining bids for different packages of items in a multi-item auction. Importantly, we need to train multiple neural networks that compete rather than only a single one. As network sizes $d_i \in \mathbb{N}$ are finite, the problem of choosing an infinite-dimensional strategy is thus transformed into choosing a finite-dimensional parameter vector θ_i.

As auction allocations x are inherently discrete, the ex-post utilities $u_i(v_i, b_i, b_{-i})$ in auction games have discontinuities and, as a result, are not (sub)differentiable in b_i. While the set of discontinuities is typically a v-nullset, taking the analytical gradient elsewhere nevertheless would yield systematically misleading updates: As an example, consider a first-price sealed-bid auction of a single item where the winner i pays her bid $p_i = b_i$. Players' utility functions are then separated into two intervals: When bidding below the highest other bid, one will lose the auction, have a constant payoff of 0, with $\nabla_{b_i} u_i = 0$ on this interval. Thus there will be no usable learning feedback. When i's bid wins the auction, any further increase in b_i will marginally decrease u_i, $\nabla_{b_i} u_i = -1$. Analytical gradient updates via backpropagation on the ex-post utility will thus always send nonincreasing feedback, until all players finally bid a constant amount of zero no matter their type.

NPGA alleviates this feedback-breakdown of the ex-post gradients by instead estimating the effect of parameter changes on the *ex-ante* utility using finite differences, and computing gradient estimates $\nabla_\theta \tilde{u}$ using a natural evolution strategy (ES) approach Salimans et al. (2017). Given parameters $P \in \mathbb{N}$ and $\epsilon > 0$, we perturb the parameter vector P times $\theta_{i;p} \equiv \theta_i + \varepsilon_p$ using zero-mean Gaussian noise $\varepsilon_p \sim \mathcal{N}(0, \sigma^2)$. NPGA then calculates each perturbation's *fitness*, $\varphi_p \equiv \tilde{u}_i(\pi_i(v_i; \theta_{i;p}), \beta_{-i})$, via Monte Carlo integration, and estimates the gradients as the fitness-weighted perturbation noise $\nabla_\theta^{ES} \equiv \frac{1}{\sigma^2 P} \sum_p \varphi_p \varepsilon_p$. This results in an unbiased estimator of the ex-ante gradients $\nabla_\theta \tilde{u}$ even when the ex-post gradients $\nabla_b u$ are not well-defined. Pseudo-code of NPGA is given in Algorithm 1.

Unlike in Bichler et al. (2021), where the "symmetric" version of NPGA has been analyzed, here we focus on the asymmetric case where agents can differ (in their prior F_{v_i}, or in how the auctioneer treats their bids) and each agent must learn their own optimal bid function. As indicated earlier, this necessitates each bidder to train her own neural network, rather than allowing a simplification of a single *shared* network, which is essential to the theoretical convergence analysis in Bichler et al. (2021). Instead, in each iteration, we iterate over bidders who perform their own individual gradient updates.

In summary, NPGA "implements" Equation 4.1 by parametrizing strategies using neural networks and training them with ES-pseudogradients:

$$\beta_i^t \equiv \pi_i(\,\cdot\,; \theta_i^t) \quad \text{with} \quad \theta_i^t \equiv \theta_i^{t-1} + \Delta_i^t \quad \text{where} \quad \Delta_i^t \propto \nabla_{\theta_i^t}^{ES}. \tag{4.2}$$

The computation of these updates in each iteration only relies on values of the ex-ante utility $\tilde{u} = \mathbb{E}_{v \sim F}[u]$. No further information about the game is necessary. Thus, whenever the joint ex-post utility u can be calculated in a vectorized fashion, $\tilde{u}$ can leverage parallel computations to efficiently perform Monte Carlo integration over $\mathcal{V}$. In practice, this approach lends itself to accelerated computation using GPUs. We built custom vectorized implementations of many common auction mechanisms using the PyTorch framework (Paszke et al. 2017) that allow us to perform the Monte Carlo estimation multiple orders of magnitude faster than prior numerical work on auctions.

Algorithm 1 Neural Pseudogradient Ascent using Evolutionary Strategies

```
input: Initial policy, ES population size $P$, ES noise variance, learning rate, batch size
for $t = 1, 2, \dots$ do
    Sample a batch of valuation profiles from prior
    Calculate joint utility of current strategy profile
    for each agent $i \in \mathcal{I}$ do
        for each $p \in \{1, \dots, P\}$ do
            Perturb agent $i$'s current policy
            Evaluate fitness of perturbation $p$ by playing against current opponents
        end for
        Calculate ES pseudogradient as fitness-weighted perturbation noise
        Perform a gradient ascent update step on the current policy
    end for
end for
```

5. Evaluation

We will provide three metrics for evaluating the quality of the learned strategy profiles β. Whenever an analytical BNE β^* is known, we may simply check whether $\beta \to \beta^*$. To do so, we calculate the agents' utility losses $\mathcal{L}_i$ that result from playing the learned strategy β_i rather than the equilibrium strategy β_i^*.

The *relative utility loss* is then given by

$$\mathcal{L}(\beta_i) \equiv 1 - \hat{u}_i(\beta_i, \beta_{-i}^*)/\hat{u}_i(\beta^*). \tag{5.1}$$

Additionally, we will also measure the distance in strategy space, which tells us how close the learned strategy is to the analytical one:

$$L_2(\beta_i) \equiv \|\beta_i - \beta_i^*\|_{\Sigma_i}. \tag{5.2}$$

Both of these metrics use Monte Carlo integration over a large number of valuations $v \sim F$ to approximate $\hat{u} \approx \tilde{u}$.

When no equilibrium is available for comparison we will instead qualify β by considering the potential gains of deviating from β itself: $\hat{\ell}_i \approx \tilde{\ell}_i(\beta_i, \beta_{-i})$. We will also estimate the "true" epsilon of β, i.e., the smallest ϵ such that β forms an interim ϵ-BNE. This estimator will be denoted by $\hat{\epsilon}$. As we will see, these additional metrics in the absence of analytical solutions are costly: Calculating $\hat{\ell}$ and $\hat{\epsilon}$ relies on a grid $\{b_{i,w} | w = 1, \ldots, n_{\text{grid}}\}$ of equidistant feasible bids for each player i, in order to cover the spaces $\mathcal{A}_i$. For given v_i and b_i, one can then approximate the interim loss $\overline{\ell}$ via

$$\hat{\lambda}_i(v_i; b_i, \beta_{-i}) \equiv \max_{w \in \{1,\ldots,n_{\text{grid}}\}} \frac{1}{n_{\text{batch}}} \sum_{h=1}^{n_{\text{batch}}} u_i\left(v_i; b_{i,w}, \beta_{-i}(v_{h,-i})\right) - u_i\left(v_i; b_i, \beta_{-i}(v_{h,-i})\right). \quad (5.3)$$

Here the batch n_{batch} only runs across opponent valuations v_{-i}. Evaluating $\hat{\lambda}_i$ for a single valuation v_i therefore requires $(n_{\text{batch}} + 1) \cdot n_{\text{grid}}$ simulations of the auction. The ex-ante loss can then be estimated as $\hat{\ell} = \frac{1}{n_{\text{batch}}} \sum_h \hat{\lambda}_i(v_{h,i}; \beta_i(v_{h,i}), \beta_{-i})$.

The worst-case interim loss is then given by $\hat{\epsilon} = \max_h \hat{\lambda}_i(v_{h,i}; \beta_i(v_{h,i}), \beta_{-i})$. Bosshard et al. (2020) proofed that this estimator can be shown to be an upper bound under further assumptions on the mechanism and the strategies. They additionally provide empirical evidence of the approximation quality of the estimator which justifies its usage.

Both computations can use a shared state for the estimations of $\hat{\lambda}$ but nevertheless $\mathcal{O}(n \cdot n_{\text{grid}} \cdot n_{\text{batch}}^2)$ auction simulations are necessary to compute these metrics. In comparison, a learning update in NPGA needs $\mathcal{O}(n \cdot P \cdot n_{\text{batch}})$ simulations only, with the population size $P \ll n_{\text{grid}}$. Due to the high cost of these additional metrics on dense grids $b_{i,w}$, we evaluate the metrics $\hat{\ell}$ and $\hat{\epsilon}$ on smaller batch sizes than $\mathcal{L}$, and only once at the end of an experiment. Finally, to approximate the relative utility loss (Equation 5.1) in the absence of known BNE, we estimate the relative ex-ante utility loss incurred in hindsight by not playing a best response, given as

$$\hat{\mathcal{L}}(\beta_i) \equiv 1 - \frac{\hat{u}_i(\beta)}{\hat{u}_i(\beta) + \hat{\ell}_i(\beta)}. \quad (5.4)$$

We choose this as our main evaluation criterion as its calculation is feasible and its values are comparable across the variety of settings considered.

6. Results

In this section, we report the results of several challenging auction models that allow for various types of asymmetries among bidders and fairly general market environments. In many of these environments, we have analytical solutions which provide unambiguous baselines. Note that these environments already describe some of the most challenging equilibrium problems to solve analytically. For more complex models, closed-form solutions of Bayesian Nash equilibrium strategies are typically not available. We introduce these environments individually and report the results and the runtimes. As we will observe, NPGA converges to approximate equilibria in all presented settings.

We use common hyperparameters across almost all settings (except where noted otherwise): fully connected neural networks with two hidden layers of ten nodes each with SeLU activations (Klambauer et al. 2017), as well as ReLU activations in the output layer. The parameters θ_i are then given by the weights and biases of these networks. The resulting parameter dimensionality d_i for each bidder thus depends on the dimensionality of the input and output layers and ranges from $d_i = 141$, in the single-item settings, to $d_i = 372$ in the 12-item multi-unit setting. All experiments were performed on a single Nvidia GeForce 2080Ti with 11GB of RAM and batch sizes in Monte Carlo sampling were chosen to maximize GPU-RAM utilization: A learning batch size of 2^{18}; primary evaluation batch size (for $\mathcal{L}$, L_2) of 2^{22}; and secondary evaluation batch size $n_{\text{batch}} = 2^{12}$ and grid size $n_{\text{grid}} = 2^{10}$ (for $\hat{\ell}$, $\hat{\epsilon}$). Each experiment was repeated ten times with 2,000 learning iterations each. Section 1 in the online supplement (Bichler et al. 2023) gives insights on the influence of the batch size and the population size, arguably the most important hyperparameters of NPGA. In the single-item auctions, it takes approximately 0.3 seconds to compute each learning iteration, whereas the combinatorial LLG auction takes about 2.0 seconds due to the complexity of the auction mechanism. For the larger LLLLGG and LLLLRRG auctions under the first-price payment rule, the computation takes under one second per iteration. We present a thorough discussion of the factors which influence the computational cost and runtimes in Subsection 6.6.

6.1. Single-Item Auctions with Asymmetric Priors

Our initial analysis focuses on a standard *single-object* first-price sealed-bid (FPSB) auction with asymmetric priors, where bidder valuations are drawn from two different distributions. FPSB auctions have mostly been analyzed with symmetric priors and equilibrium bid functions. Asymmetric prior distributions are harder to analyze analytically compared to symmetric environments, but a few environments with analytical solutions are known. We analyze three different environments, one with two overlapping uniform distributions and a unique BNE, one with two disjunct uniform distributions and multiple BNE, and another one where the priors are non-linear beta functions.

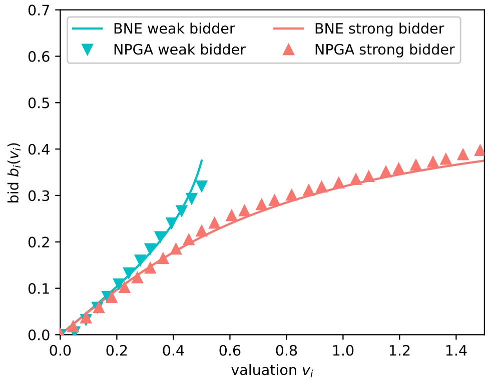


Figure 1 **Equilibrium bid function and strategies learned by NPGA in the asymmetric single-item setting with overlapping valuations.**

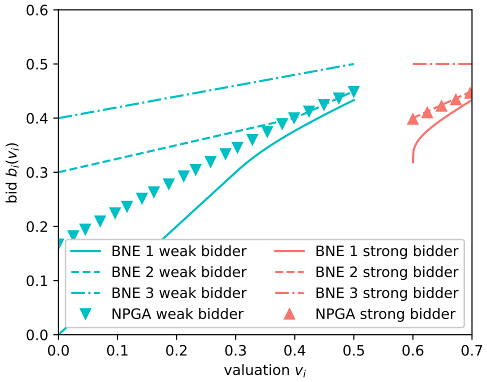


Figure 2 **Equilibrium bid function and strategies learned by NPGA in the asymmetric single-item setting with non-overlapping valuations.**

Table 1 **Average losses achieved in the asymmetric first-price setting with overlapping valuations. Mean and standard deviation are aggregated over ten runs of 2,000 iterations each. The time per iteration is 0.2886 (0.0280) seconds.**

bidder	$\mathcal{L}$	$\hat{\mathcal{L}}$	L_2
strong bidder	0.0024 (0.0026)	0.0178 (0.0037)	0.0104 (0.0055)
weak bidder	0.0074 (0.0031)	0.0524 (0.0134)	0.0128 (0.0029)

6.1.1. Asymmetries Induced by Priors with Different Domains. We first analyze an environment with two bidders who have overlapping uniform prior distributions supported on $(0, 1/2)$ and $(0, 1)$ describing a weak and a strong bidder, respectively. The analysis goes back to Plum (1992). In the BNE, the weaker bidder bids more aggressively than the strong bidder. Figure 1 shows an example of the learned and the analytical BNE bid functions for both bidders. NPGA achieves a relative loss $\mathcal{L}$ below 1% for both types of bidders. Aggregated performance results over ten runs are displayed in Table 1.

This Bayes-Nash equilibrium is unique (Maskin and Riley 2000, Lebrun 2006) given the requirement that bidders may never bid above their observed valuation. Kaplan and Zamir (2015) relaxed this assumption. In their model, the prior distributions are non-overlapping, which results in additional equilibria. In particular, in BNE 1 and 2, which are also depicted in Figure 2, the weaker bidder has incentives to overbid. They conclude that the commonly used assumption of no overbidding, or more generally, the elimination of weakly dominated strategies, should be taken more carefully in asymmetric auctions.

Table 2 Average NPGA losses achieved in asymmetric first-price setting with non-overlapping valuations. Aggregated over ten runs of 2,000 iterations each and compared against the second equilibrium of Kaplan and Zamir (2015). The time per iteration is 0.2856 (0.0221) seconds.

bidder	$\mathcal{L}^{\text{BNE2}}$	$\hat{\mathcal{L}}$	L_2^{BNE2}
strong bidder	0.0080 (0.0097)	0.0104 (0.0012)	0.0109 (0.0085)
weak bidder	0.1687 (0.2310)	0.0229 (0.0140)	0.0544 (0.0161)

We analyzed NPGA in this setting with bidders that have non-overlapping uniform prior distributions, $\mathcal{V} = (0, 0.5) \times (0.6, 0.7)$ (see Table 2). In this model, there are three Bayesian Nash equilibria. Despite the equilibrium selection problem in this game, starting from truthfully initialized strategies, the bidding converges to BNE 2. The stronger bidder is able to decrease her relative utility loss below 1%. Only the weaker bidder has difficulties finding a particular strategy for low valuations because bids in this range are far from competitive for *any* opposing bids and rarely, if ever, win. This strategic disadvantage leads to sparse opportunities to learn in this specific setting, which in turn causes higher relative errors. In fact, only about 1/5 of the sampled data, i.e., the highest valuations of the weak bidder, are relevant for learning.

6.1.2. Asymmetries Induced by Different Prior Densities. For the single-item symmetric FPSB auction with two bidders and assuming uniform priors on $(0, 1)$, the equilibrium strategies and market outcomes are well understood analytically. Apart from the uniform distribution, we also want to analyze asymmetric environments with more complex non-linear prior distributions. Therefore, we analyzed an environment with two bidders whose values are drawn from a beta distribution B with parameters $\alpha, \beta > 0$. Note that for $\alpha = \beta = 1$ the beta distribution equals the uniform distribution. Except for this special case, no analytical equilibrium is known for the asymmetric case. Now we can analyze diverse market outcomes by running NPGA for various combinations of these parameters. As an example, we have selected a valuation prior of $B(0.8, 1.2)$ for the weak bidder and $B(1.2, 0.8)$ for the strong bidder's valuations prior. Note that NPGA has no access to the underlying distributions explicitly, but it learns the opponent's prior implicitly by observing frequencies of the played actions.

As a result of the change in the prior distributions, we already see the change in strategy in Figure 3 compared to the BNE of $\beta(v) = \frac{1}{2}v$ under common uniform priors. As expected,

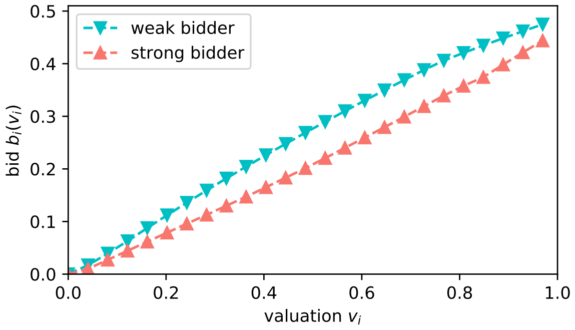


Figure 3 **Learned bid functions after 2,000 iterations for bidders with asymmetric priors:** $B(0.8, 1.2)$ **for the weak bidder with an expected valuation of** 0.4 **and** $B(1.2, 0.8)$ **for the strong bidder with an expected valuation of** 0.6**.**

because of its strategic disadvantage, the weaker bidder bids more aggressively, whereas the stronger bidder can lower its bids. As no analytical equilibrium is available to compare against, we report the approximated utility loss $\hat{\mathcal{L}}$ from Equation 5.4—the amount of a possible utility gain against the opponent—which decreases below 2.37% for the strong bidder and to 3.48% for the weaker bidder. The time per iteration of NPGA of 0.3131 (±0.0220) seconds is comparable with the previous single-item experiments.

6.2. Multi-Unit Auctions with Asymmetric Bidders

This section is concerned with a specific type of multi-unit uniform-price auction with two different classes of bidders for which a closed-form expression of the equilibrium is not available. Such mechanisms are used in treasury bill auctions and also in electricity markets. The environment is very large, with up to 12 units, and NPGA is able to compute a sufficiently close equilibrium in a few minutes.

Demand reduction is an important characteristic of equilibrium bidding strategies in uniform-price auctions (Krishna 2009): Bidders submit bids on fewer items in order to reduce competition, lower the price, and increase their payoffs. The phenomenon of demand reduction can be observed in all our experiments.

In our experiments, we consider two weak bidders with uniform, marginally decreasing valuations on $\mathcal{V}_i = \{v_i \in [0,1]^m : v_{i,1} \geq \cdots \geq v_{i,m}\}$ and one strong bidder with analogously distributed valuations on $[0,2]$. Unlike in general CAs in multi-unit auctions it is sufficient to bid on individual items rather than bundles. Thus, the action space is given by $\mathcal{A}_i = \mathbb{R}^m_+$ and the neural network strategies take m inputs (the marginal valuations) and produce m outputs (the bids for each incremental unit received).

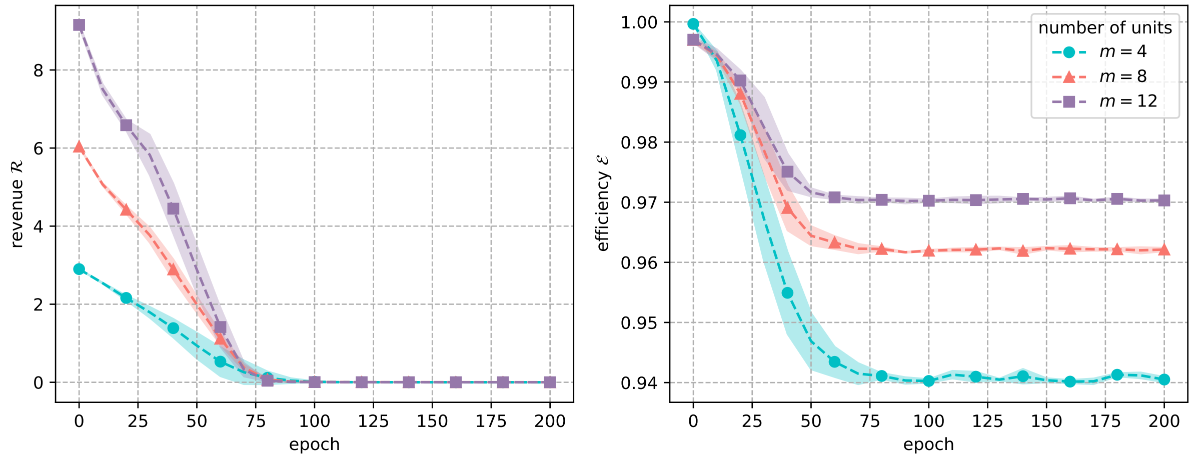


Figure 4 **Revenue $\mathcal{R}$ and efficiency $\mathcal{E}$ during self-play in different asymmetric multi-unit auctions. The means (opaque lines) and the standard deviations (shaded area) are depicted.**

Simulating market sizes with $m \in \{4, 8, 12\}$, one can observe that the agents reduce their demand by lowering bids for multiple goods to zero. For example, in the market with four goods, the strong bidder and weak bidders only bid on two items and one item, respectively. Thus, they learn to collaborate to maximize their payoff. For the remaining demand, the bids are approximately truthful. Similar observations can be made in the other markets for a corresponding higher demand. Figure 4 shows how the seller's revenue decreases to zero when bidders learn to reduce demand and how the efficiency decreases when initialized with truthful bidding strategies. We do not plot the exact bid functions learned due to space constraints. Note that demand reduction happens when the bidders' demand can be easily distributed among the available goods. In other experiments, where there are only very few items and many bidders, the prices stay high.

The approximate utility loss decreases consistently below 1% for all runs.[2] With the default batch size, the experiments with 4, 8, and 12 units took on average about 1.1557 (±0.022), 1.3056 (±0.0207), and 2.405 (±0.0259) seconds, respectively.

6.3. The Asymmetric LLG Model

Next, we focus on the LLG model with three single-minded bidders and two heterogeneous objects or items (Ausubel et al. 2006). This model has received significant attention

[2] Note that we have increased the grid size used for computing the utility loss for the 4, 8, and 12 item case to 2^{14}, 2^{16}, and 2^{22}, respectively. The resulting grid is not as dense as if it was applied in single-dimensional environments.

in the study of core-selecting combinatorial auctions, which are being used in spectrum auctions worldwide (Goeree and Lien 2016, Bichler and Goeree 2017). After the auctioneer has determined the welfare-maximizing allocation, she computes a minimum-revenue core-selecting payment, where winning bidders merely have to pay enough such that no coalition of bidders could potentially deviate together with the auctioneer.

The LLG model is small enough to allow for game-theoretical analysis. Here, the global bidder is interested only in winning the package of both objects while the two local bidders desire exactly one of the objects each. The local bidders thus only need to outbid the global bidder. Both local bidders have an incentive to free-ride on each other, reminiscent of public goods problems. In the original model, the local bidders' priors are symmetric and the global bidder has a simple dominant strategy to bid truthfully in any core-selecting auction. Analytical solutions for core-selecting combinatorial auctions with different payment rules exist (Goeree and Lien 2016, Ausubel and Baranov 2019). Gradient dynamics were shown to achieve very good results in this standard model and approximate the BNE of the local bidders closely (Bichler et al. 2021).

Ott and Beck (2013) introduced a version with asymmetry among the local bidders that causes overbidding by one of the local bidders, which may help explain the outcomes observed in several real-world spectrum auctions. Unlike in the original LLG model, Ott and Beck (2013) define bidder local 2 to be *favored*, meaning that she pays VCG prices[3] for every realization of bids and for every optimal assignment of the items. As a result, bidder local 1 has to pay a higher price. The authors derive an intriguing BNE in which bidder local 1 overbids while both other bidders report their valuations truthfully. More precisely, bidder local 1 places bids for two bundles: the bundle containing only her desired item, as well as the package of both items. Her bid for the package of both items always exceeds the bid for the single desired good, which implies positive demand for the second item even though it provides no additional value to the bidder. This results from an incentive of bidder local 1 to raise the other bidders' payments so that her payment decreases. Such overbidding can increase the prices for opponents, which might lead to high revenues and price differences among bidders. They characterize the exact BNE strategy which is depicted in Figure 5 below. This model is important as it shows that the assumption

[3] Vickrey-Clarke-Groves (VCG) payments are calculated such that each bidder pays for the harm they cause to other bidders by participating in the auction.

Table 3 Results in the asymmetric LLG setting after 2,000 iterations and averaged over ten repetitions. The mean and standard deviation are shown.

bidder	$\mathcal{L}$	$\hat{\mathcal{L}}$	L_2
local 1	0.0005 (0.0005)	0.0119 (0.0107)	0.0353 (0.0082)
local 2	0.0001 (0.0001)	0.0172 (0.0151)	0.1146 (0.0600)
global	0.0000 (0.0000)	0.0058 (0.0054)	0.0281 (0.0112)

that each player only needs to bid for her bundle of interest is, in fact, restrictive, even when the single-mindedness of bidders is common knowledge. Without this assumption, very different equilibrium behavior can emerge as was recently discussed by Bosshard and Seuken (2021).

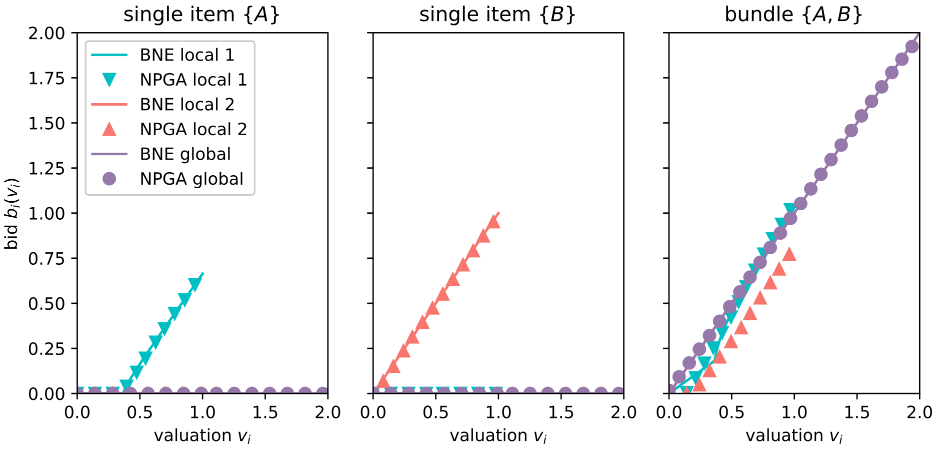


Figure 5 Learned strategies in the asymmetric LLG setting. The left two subplots depict the bids on the individual items that must compete with the bundle bids in the rightmost plot. Bidders 1 and 3 learn to bid almost truthfully and bidder 1 indeed learns to overbid on the bundle as the theory suggests.

Table 3 shows the performance of NPGA in this market. The resulting loss in equilibrium compared to adhering to the analytical BNE strategy $\mathcal{L}$ is well below 0.1% across all agents. Note that the bidders local 2 and global indeed learn to report their valuations truthfully for item B and the bundle, respectively. There is a small deviation from the analytical BNE for bidder local 2, who decreases her bundle bid slightly below her valuation. However, she would not have to bid on the bundle at all in equilibrium: Note that when bidding the same value on item B and the bundle, she would never be allocated the bundle in this auction. As such, the bid on the package of both items learned is irrelevant and the outcome of the auction under the learned NPGA strategies will always be identical in terms of prices and allocations to those in the analytical BNE. Importantly, the bundle bid for bidder local 1 indeed lies above the truthful bid for high valuations, which describes a non-obvious bidding strategy. Notably, NPGA can discover this incentive for overbidding. We point out that there is a minor difference in the NPGA strategy and the BNE in the bundle bid of bidder local 1 for low valuations. However, this difference has a negligible impact on the expected utility of any of the agents.

The time per NPGA update iteration averages at 1.9602 (±0.0358) seconds. Here, we clearly see the computational impact of allocating a bundle of goods, and computing the corresponding prices, as compared to auctions with a single good or multiple goods that are sold individually. The computational workload lies mainly in simulating the auction outcomes and not in learning and updating the strategies themselves, as we will discuss in Subsection 6.6.

6.4. The Split-Award Auction Model

Even in the asymmetric LLG model discussed above, each bidder is only interested in one package. An environment of a combinatorial auction with multi-minded bidders was analyzed in Anton and Yao (1992) and later in Kokott et al. (2019). This model is known to have multiple pure BNE, and it is interesting to understand how NPGA deals with the resulting equilibrium selection problem.

The model is a reverse auction and it is described by the bidders' type (or cost) distribution

$$\mathcal{V}_i = \{v_i \in \mathbb{R}^2 : v_{i,1} \sim F,\ v_{i,2} = C \cdot v_{i,1}\}, \quad i = 1, 2,$$

where $v_{i,1}$ corresponds to the cost of the 50% lot (or items) and the *efficiency parameter* C corresponds to the fraction of total costs for one of the lots. In our experiments we set parameters $F = \mathcal{U}(1.0, 1.4)$ with $C = 0.3$, being consistent to prior experimental work (Kokott et al. 2019). The environment describes diseconomies of scale in the production costs, which make the game strategically interesting.

There are two classes of Bayesian Nash equilibria in this game: First, there is a (single) so-called "winner-takes-all" equilibrium (WTA), which is economically inefficient and in which one bidder wins both items. The other class comprises a continuum of efficient "pooling equilibria" where both suppliers coordinate and reach a common price such that each

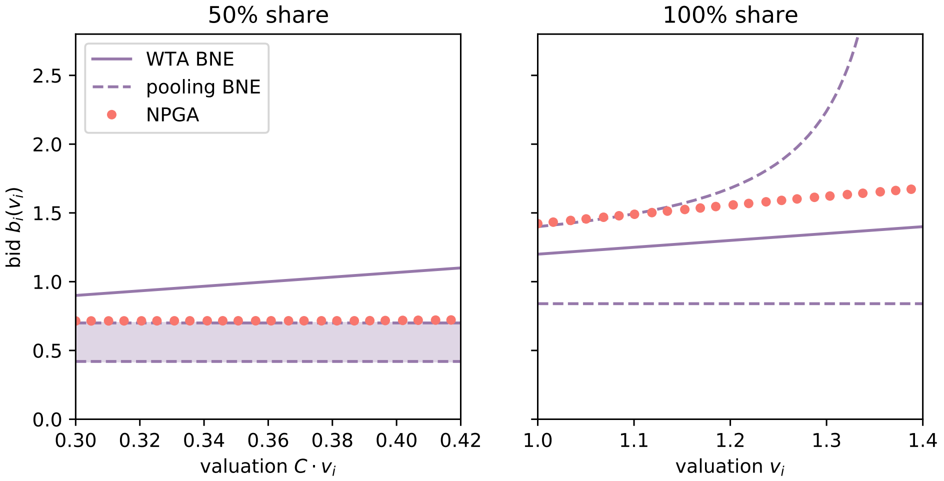


Figure 6 **The figure depicts the winner-takes-all equilibrium (solid line), the bounds for the range of efficient pooling equilibria (shaded), and the NPGA strategies (dotted line) for the first-price split-award auction. As the NPGA strategy is within the continuum of efficient pooling equilibria, two bidders playing according to this strategy always end up with a split contract for one lot each.**

bidder wins one of the goods (Anton and Yao 1992). In Figure 6, this class is represented by the shaded area. In such a pooling strategy, the two bidders select a price independent of their type or value. The payoff-dominant strategy for each bidder is achieved in the pooling equilibrium with the highest bids on a single lot. Apart from these two classes of pure-strategy Nash equilibria, hybrid equilibria are known to exist and there might also be mixed equilibria in nondeterministic strategies, which makes this setting strategically challenging.

Figure 6 depicts the analytically known pure-strategy BNE alongside a strategy learned via NPGA. Running NPGA multiple times, it always converges to a state close to the bidder-optimal pooling BNE: the bidders cooperate in the split equilibrium, where each one wins one lot a high price. NPGA reaches an average utility of 0.384 over ten runs compared to an expected utility of 0.34 in the analytical BNE. This outcome is notable as it requires coordination between the players which is strategically much more challenging than the simple competition to win both items at once, which resembles a single-item auction: To achieve a pooling equilibrium, players must not only submit a high bid for the single-lot, but also need to coordinate on a bid for the two-item bundle, such that deviating from the pooling strategy does not become profitable for the opponent.

Table 4 Results of NPGA after 5,000 iterations in the LLLLGG first-price auction. Results are averages over ten replications and the standard deviation is displayed in brackets.

bidder	$\tilde{u}$	$\hat{\epsilon}$	$\hat{\mathcal{L}}$
globals	0.2366 (0.0040)	0.0235 (0.0026)	0.0171 (0.0006)
locals	0.1793 (0.0012)	0.0241 (0.0024)	0.0230 (0.0006)

With NPGA, the agents learn to bid accordingly on the 100% share in this equilibrium, but this bid becomes subject to minor random changes as there is no "reward signal," that is, the bid does not determine the price. In Figure 6 one can also see that bidding on the 50% lot is very close to the payoff-dominant (highest) pooling bid, whereas the bid on the 100% share lies within the continuum of possible equilibria. The distance in strategy space L_2 decreases to 0.0251, where we only measure the distance of the winning bid as the other bid falls within the continuum of possible BNE bids. The relative ex-ante utility loss $\mathcal{L}$ decreases to 0.0185 and $\hat{\mathcal{L}}$ also falls below 2%. The average time per iteration of 0.4627 (±0.0154) seconds is again much lower than in the combinatorial LLG auction.

6.5. Large Combinatorial Auction Models

Finally, we analyze the LLLLGG model which was introduced by Bosshard et al. (2020) as a benchmark for equilibrium computation, as well as an extension, the LLLLRRG model. There is little hope for analytical solutions to such problems and the fact that the winner determination and payment rules involve NP-hard problems makes them challenging problems for equilibrium computation.

In the LLLLGG model six bidders compete for eight items: Inspired by geographical constraints, four of the bidders are "local" and are interested in two overlapping bundles of two items each. The other two bidders are "global", and each aims to win one of two larger bundles comprising four items each. These bidder classes are asymmetric and no analytical BNE is known. Therefore, we again report the utility loss that we find after learning with NPGA.

As shown in Figure 7, the bidders' utility converges quickly to around 0.24 (local bidders) and 0.18 (global bidders) and the utility losses drop quickly. Due to the computational requirements of this model, we reduced the number of experiments. Both bidders show a small relative ex-ante utility loss of $\hat{\mathcal{L}} < 1.8\%$ and 2.4% for the global and local bidders, respectively. Direct runtime comparisons to other state-of-the-art methods like

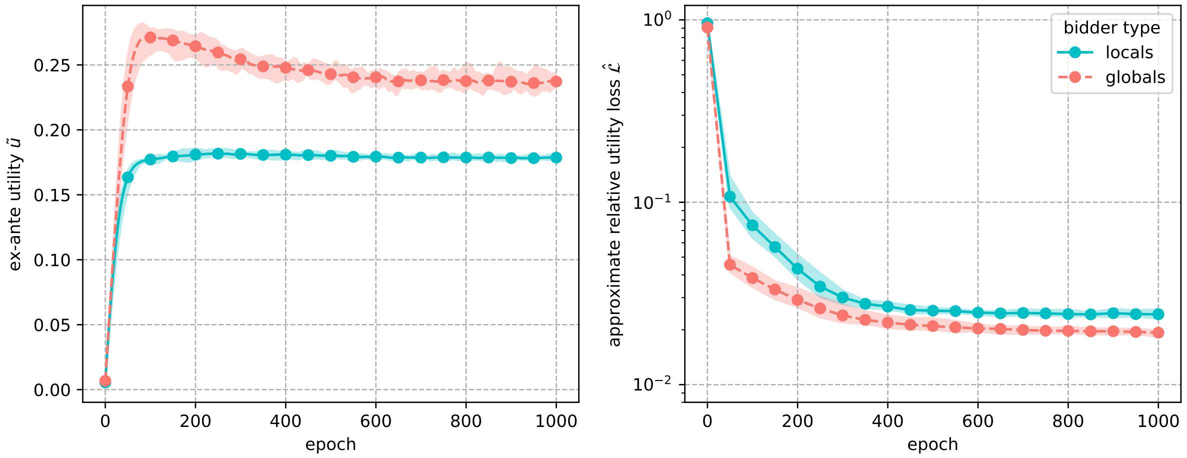


Figure 7 **Ex-ante utility $\tilde{u}$ and loss $\hat{\mathcal{L}}$ of in NPGA self-play in the LLLLGG first-price auction. The shaded area and line show mean and standard deviation over ten repetitions.**

Bosshard et al. (2017, 2020) are difficult due to differences of NPGA and their method in terms of goals (ex-ante vs. "stronger" ex-interim equilibria), implementation (generic vs. setting-specific), and hardware architecture (consumer-grade GPU vs. CPU-cluster). For the LLLLGG first-price auction, Bosshard et al. (2017) report an estimated absolute ex-interim 0.0037-BNE computed in 54,384 CPU-core hours. NPGA, on the other hand, finds an estimated (absolute) ex-ante 0.0042-BNE (absolute ex-interim 0.0241) in 38.3 minutes (corresponding to 0.4616 ($\pm$0.0010) seconds per iteration times 5,000 iterations) on a single GPU ($\approx$ 2,895 CUDA-core-hours).

We also explore a modified version, which we call LLLLRRG, which adds a third class of bidders interested in winning all eight items. Figure 4 in the online supplement (Bichler et al. 2023) depicts the valuation structure. This larger setting has not been explored in the literature previously and, to our knowledge, is the largest combinatorial auction for which a numerical BNE has been computed to date. Note that this environment is highly challenging for equilibrium computation because the auction mechanism needs to solve an NP-hard problem. One iteration of NPGA in this first-price auction takes on average 0.8097 ($\pm$0.0010) seconds on our machine. Table 5 shows the full results under the same hyperparameters as in the LLLLGG experiments. The relative utility loss decreases proportionally to the bidders' strength: to values below 1% for the global bidder and to values below 3.9% for the local bidders.

Table 5 Results (mean and standard deviation) in the LLLLRRG setting after 5,000 iterations and averaged over three repetitions.

bidder	$\tilde{u}$	$\hat{\epsilon}$	$\hat{\mathcal{L}}$
locals	0.0939 (0.0009)	0.0194 (0.0013)	0.0382 (0.0016)
regionals	0.1069 (0.0024)	0.0192 (0.0004)	0.0288 (0.0001)
global	0.4396 (0.0044)	0.0356 (0.0031)	0.0093 (0.0009)

Table 6 Overview of all auction environments with the corresponding NPGA hyperparameters and the resulting number of simulations and runtimes. One NPGA iteration requires $n_{\text{models}} \cdot n_{\text{batch}} \cdot (P+1)$ auction evaluations, where batches are computed in parallel and model perturbations sequentially.

setting	n_{batch}	P	number of iterations	per iteration time (s)	per iteration auctions	total time (h:m:s)	total auctions
single-item uniform overlapping FPSB (6.1.1)	262,144	64	2,000	0.2886	34,078,720	0:09:37.20	68,157,440,000
single-item uniform non-overlapping FPSB (6.1.1)	262,144	64	2,000	0.2856	34,078,720	0:09:31.20	68,157,440,000
single-item beta asymmetric FPSB (6.1.2)	262,144	64	2,000	0.3131	34,078,720	0:10:26.20	68,157,440,000
multi-unit with 4 units (6.2)	262,144	64	2,000	1.1557	34,078,720	0:38:31.40	68,157,440,000
multi-unit with 8 units (6.2)	262,144	64	2,000	1.3056	34,078,720	0:43:31.20	68,157,440,000
multi-unit with 12 units (6.2)	262,144	64	2,000	2.4050	34,078,720	1:20:10.00	68,157,440,000
LLG, adapted VCG (6.3)	131,072	64	2,000	1.9602	25,559,040	1:05:20.40	51,118,080,000
split-award FPSB (6.4)	262,144	64	2,000	0.4627	17,039,360	0:15:25.40	34,078,720,000
LLLLGG FPSB (6.5)	262,144	64	5,000	0.4616	34,078,720	0:38:28.00	170,393,600,000
LLLLRRG FPSB (6.5)	262,144	64	5,000	0.8097	51,118,080	1:07:28.50	255,590,400,000

6.6. Scalability and Computational Costs

Let us now provide a summary of all experiments with their runtimes (Table 6) and a discussion of the computational cost. The runtimes range from a few minutes up to 80 minutes for the most complex scenarios with an NP-hard allocation problem and eight bidders. Let us put these empirical results into perspective. At first sight, it is surprising that we can solve such equilibrium problems at all. As discussed in the introduction, the computational complexity of computing BNE in auction games is, in general, an open problem. The analysis by Cai and Papadimitriou (2014) for a specific asymmetric multi-object auction model proves PP-hardness of exact BNE computation and this suggests that the class of problems is generally very hard. They also show that learning only approximate BNE cannot be polynomial in the number of items to be sold in the auction. As a result, no algorithm can be expected to efficiently compute approximate BNE in the general case. Note that the hardness of approximating a BNE also hinges on the observation that the number of strategies grows quickly in the number of items in their environment. In many auction models, the number of relevant strategies is small even with multiple items. For example, bidders might only be interested in a few out of many items in a combinatorial auction. Even in large combinatorial auctions with many bidders, one can typically limit attention to the strategic analysis of a few pivotal bidders.

In this paper, we have analyzed a number of challenging environments which are significantly more complex than models for which we can derive an equilibrium strategy analytically. For example, combinatorial auctions require solving an NP-hard winner determination problem. Yet, we can solve problems with eight items and seven bidders interested in multiple packages within 67 minutes.

Let us analyze the computational costs of NPGA in more detail. As a zeroth-order method, the vast majority of the computational cost required by NPGA results from calculating samples of the ex-post utilities u across the joint valuation space $\mathcal{V}$, in order to compute estimates of $\tilde{u}$ via Monte Carlo integration (i.e., lines 4 and 8 of Algorithm 1). Here, the main driver of computational cost is the auction mechanism itself, i.e., the cost of computing the winning allocation and the price vector. The role of the remaining computations in NPGA — namely sampling joint valuations $v \in \mathcal{V}$ and noise vectors ε_p, performing forward passes $b_i = \pi_i(v_i; \theta_i)$, aggregating auction sample results into the gradient estimates, and updating the parameters — is negligible in comparison. The cost of computing an approximate BNE using NPGA is thus determined, on the one hand, by the sample efficiency of the algorithm, that is, the number of auction simulations required, and, on the other hand, the computational cost of computing the individual auction samples. As we will see, both of these aspects vary significantly across different auction settings.

First, let us discuss the *computational cost of performing auction simulations* $u_i(v_i, b)$ for given joint valuations v and bids b according to the players' current or perturbed neural net strategies. This complexity varies significantly between auction settings and pricing rules. For example, in a single-item first-price auction, determining the allocation and prices only requires finding a (batch-wise) maximum, which is computable in $O(n)$, whereas computing core prices in a combinatorial auction requires solving a sequence of constrained quadratic problems, which themselves already constitute NP-hard problems (in the number of bidders and items) in general.

For all settings analyzed in this paper, we leverage custom implementations of the auction mechanisms that allow data-parallel simulation on GPUs. As a result, the time to compute auction samples is approximately constant in the batch size as long as an entire batch fits in GPU memory, and grows linearly with batch size thereafter. (The *utility loss estimator*, whose computation is independent of the learning algorithm NPGA, exhibits the same dynamic. Figure 3 in the online supplement (Bichler et al. 2023) depicts the constant-then-linear time complexity as a function of memory footprint.) For the experiments presented in this paper, performing a single iteration of NPGA, which involves computing $P+1$ batches of auctions for each player, takes between 0.3 and 2.4 seconds on a single Nvidia GeForce RTX 2080Ti GPU (see Table 6). While our implementation sequentially computes the utilities for each of the P model perturbations, these operations could easily be parallelized across larger or multiple GPUs.

The other important aspect is the *sample complexity of the algorithm*, which further breaks down into the number of samples needed for gradient estimation in each iteration, and the number of iterations needed to converge to an equilibrium. As discussed above, the asymmetric settings we study differ from those in Bichler et al. (2021) in that no theoretical convergence guarantee is available for simultaneous gradient methods (or any no-regret learner), even asymptotically. Consequently, it is difficult to characterize the number of gradient updates needed to converge to an approximate equilibrium, even if an exact oracle for the ex-ante gradient were available.

The gradient estimation in one iteration of NPGA requires $n_{\text{batch}} \cdot (P+1)$ auction simulations for each player (or class of identical players). Both higher batch sizes and higher population sizes will reduce the variance of the estimator at the expense of higher computational costs. An exemplary analysis of the impact of these hyperparameters on the learned equilibrium outcomes in the LLLLGG setting is presented in the online supplement (Bichler et al. 2023). As the estimation is performed via Monte Carlo integration, it is susceptible to the curse of dimensionality: Given a fixed batch size of samples, the variance of the estimator will increase with the dimensionality of the valuation space $\mathcal{V}$.

Furthermore, the specific prior distributions F may also affect the fidelity of the Monte Carlo estimator: For example, we observe that ceteris paribus, settings with uniform priors exhibit lower variance in the gradient estimator, compared to nonuniform priors. Intuitively, this is because the tails of the distribution, particularly for the highest valuations, play a significant role in the total achievable utility of a player. As a result, more samples are necessary to adequately calculate the utility contribution of these low-density regions. For example, NPGA would require more iterations to achieve the same performance in the asymmetric setting with Beta-distributed priors (Subsection 6.1.2). In practice, one may employ several variance-reduction techniques, such as importance-sampling or low-discrepancy sequences of quasi-random valuation samples to further improve the sample efficiency of Monte Carlo integration (Bosshard et al. 2020). While these methods are conceptually applicable to NPGA, they require setting-specific implementations and are not explored further in this work.

In summary, the key drivers for the total runtime of NPGA are the number of players, number of items, choice of prior distribution, and auction mechanism, as they influence the ability to efficiently calculate low-variance gradient estimates. Importantly, we demonstrate that NPGA, for the first time, finds close approximations of BNE in two of the largest settings to date, namely a 12-item, 3-bidder multi-unit auction (Subsection 6.2), and the 8-item, 7-bidder combinatorial auction with multi-minded bidders ("LLLLRRG", Subsection 6.5).

7. Conclusion

Understanding the result of strategic interaction on markets is a fundamental problem and one that appears everywhere in economics and the management sciences. Equilibrium solution concepts are our primary approach to studying the outcome of games with multiple interacting agents. They help understand fundamental questions about the efficiency of markets, but equilibrium analysis can also provide tangible guidance for bidding in specific markets such as in procurement auctions or in high-stakes spectrum sales and for the design of specific auction mechanisms. Algorithms to compute equilibrium strategies in games would have a substantial impact on theory and practice. However, computing equilibrium in auction games with continuous action space and value distributions turned out very challenging. We know little about the existence of equilibrium in such auctions and do not have a mathematical solution theory for the underlying differential equations in more complex markets. Obviously, an equilibrium solution concept that is intractable is of little value and can hardly serve as a prediction for the outcome of a game. Equilibrium learning provides a reasonable behavioral model of agents in a market. While the implementation of equilibrium learning algorithms in auction games is challenging, we show that NPGA reliably finds equilibrium in a surprisingly wide array of complex auction models. The experimental results reported in this paper show that the gradient-based algorithm implemented in NPGA finds BNE even in asymmetric environments with multiple equilibria. Such asymmetric environments required us to train multiple neural networks with multiple outputs, where convergence to the bidder-optimal equilibrium is far from obvious.

An open question concerns a broader theoretical characterization of Bayesian games in which NPGA converges to a Bayesian Nash equilibrium. However, this is a very challenging theoretical endeavor that is beyond this article. Learning dynamics do not generally obtain a Nash equilibrium (Benaim and Hirsch 1999). A number of recent results on matrix games showed that gradient dynamics can either circle, diverge, or even be chaotic (Sanders et al. 2018). Actually, the study of gradient dynamics in games is akin to studying dynamical systems and characterizing environments, where gradient dynamics converge to a Nash equilibrium (if one exists), can be arbitrarily complex (Andrade et al. 2021).

However, even if we do not know a priori if an algorithm converges, we can verify an approximate BNE ex-post, if the algorithm converges. If we analyze many environments as in this article, we might be able to induce characteristics of auction models that can be learned via NPGA and those that cannot. In our experiments, we found that NPGA always converged to an approximate Bayes-Nash equilibrium in single- and multi-object auctions and we did not encounter cycling or chaotic behavior as was observed for finite games. As such, NPGA provides the foundation for widely applicable equilibrium solvers.

Acknowledgments

We are grateful for funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – BI 1057/9.

References

Andrade GP, Frongillo R, Piliouras G (2021) Learning in matrix games can be arbitrarily complex. *arXiv preprint arXiv:2103.03405* .

Anton JJ, Yao DA (1992) Coordination in split award auctions. *The Quarterly Journal of Economics* 107(2):681–707.

Armantier O, Florens JP, Richard JF (2008) Approximation of Nash equilibria in Bayesian games. *Journal of Applied Econometrics* 23(7):965–981.

Athey S (2001) Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica* 69(4):861–889.

Ausubel LM, Baranov O (2019) Core-selecting auctions with incomplete information. *International Journal of Game Theory* ISSN 1432-1270, URL `http://dx.doi.org/10.1007/s00182-019-00691-3`.

Ausubel LM, Milgrom P, et al. (2006) The lovely but lonely Vickrey auction. *Combinatorial auctions* 17:22–26.

Bajari P (2001) Comparing competition and collusion: a numerical approach. *Economic Theory* 18(1):187–205.

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

Benaim M, Hirsch MW (1999) Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games. *Games and Economic Behavior* 29:36–72, URL `https://escholarship.org/uc/item/4qj1335f`.

Bichler M, Fichtl M, Heidekrüger S, Kohring N, Sutterer P (2021) Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence* 3:687–695.

Bichler M, Goeree JK (2017) *Handbook of spectrum auction design* (Cambridge University Press).

Bichler M, Heidekrüger S, Kohring N (2023) bnelearn-asymmetric Version v2021.0151 URL `http://dx.doi.org/10.5281/zenodo.7407158`, available for download at https://github.com/INFORMSJoC/2021.0151.

Bosshard V, Bünz B, Lubin B, Seuken S (2017) Computing bayes-nash equilibria in combinatorial auctions with continuous value and action spaces. *IJCAI*, 119–127.

Bosshard V, Bünz B, Lubin B, Seuken S (2020) Computing bayes-nash equilibria in combinatorial auctions with verification. *Journal of Artificial Intelligence Research* 69:531–570.

Bosshard V, Seuken S (2021) The cost of simple bidding in combinatorial auctions. *2021 Conference on Economics and Computation* (New York, NY, USA).

Bowling M (2005) Convergence and no-regret in multiagent learning. *Advances in neural information processing systems*, 209–216.

Brown GW (1951) Iterative solution of games by fictitious play. *Activity analysis of production and allocation* 13(1):374–376.

Cai Y, Papadimitriou C (2014) Simultaneous bayesian auctions and computational complexity. *Proceedings of the Fifteenth ACM Conference on Economics and Computation - EC '14*, 895–910 (Palo Alto, California, USA: ACM Press), ISBN 978-1-4503-2565-3.

Carbonell-Nicolau O, McLean RP (2018) On the existence of nash equilibrium in bayesian games. *Mathematics of Operations Research* 43(1):100–129.

Conitzer V, Sandholm T (2008) New complexity results about nash equilibria. *Games and Economic Behavior* 63(2):621–641.

Cournot AA (1838) *Recherches sur les principes mathématiques de la théorie des richesses.* URL `https://gallica.bnf.fr/ark:/12148/bpt6k6117257c`.

Daskalakis C, Goldberg P, Papadimitriou C (2009) The Complexity of Computing a Nash Equilibrium. *SIAM Journal on Computing* 39(1):195–259, ISSN 0097-5397, URL `http://dx.doi.org/10.1137/070699652`.

Etessami K, Yannakakis M (2007) On the complexity of nash equilibria and other fixed points (extended abstract). *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, 113–123, FOCS '07 (USA: IEEE Computer Society), ISBN 0769530109, URL `http://dx.doi.org/10.1109/FOCS.2007.48`.

Ewert M, Heidekrüger S, Bichler M (2022) Approaching the overbidding puzzle in all-pay auctions: Explaining human behavior through bayesian optimization and equilibrium learning. *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, 1586–1588, AAMAS '22 (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), ISBN 9781450392136.

Fudenberg D, Levine DK (2009) Learning and equilibrium. *Annu. Rev. Econ.* 1(1):385–420.

Goeree JK, Lien Y (2016) On the impossibility of core-selecting auctions. *Theoretical Economics* 11(1):41–52, ISSN 1555-7561, URL `http://dx.doi.org/10.3982/TE1198`.

Hartline J, Syrgkanis V, Tardos E (2015) No-Regret Learning in Bayesian Games. Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, eds., *Advances in Neural Information Processing Systems 28*, 3061–3069 (Curran Associates, Inc.), URL `http://papers.nips.cc/paper/6016-no-regret-learning-in-bayesian-games.pdf`.

Hazan E, Agarwal A, Kale S (2007) Logarithmic regret algorithms for online convex optimization. *Machine Learning* 69(2-3):169–192.

Jackson MO, Swinkels JM (2005) Existence of equilibrium in single and double private value auctions 1. *Econometrica* 73(1):93–139.

Kaplan TR, Zamir S (2015) Multiple equilibria in asymmetric first-price auctions. *Economic Theory Bulletin* 3(1):65–77.

Klainerman S (2010) Pde as a unified subject. *Visions in Mathematics*, 279–315 (Springer).

Klambauer G, Unterthiner T, Mayr A, Hochreiter S (2017) Self-normalizing neural networks. *Proceedings of the 31st international conference on neural information processing systems*, 972–981.

Klemperer P (2000) Why every economist should learn some auction theory. *Available at SSRN 241350* .

Kokott GM, Bichler M, Paulsen P (2019) The beauty of Dutch: Ex-post split-award auctions in procurement markets with diseconomies of scale. *European Journal of Operational Research* 278(1):202–210.

Krishna V (2009) *Auction Theory* (Academic press).

Lebrun B (2006) Uniqueness of the equilibrium in first-price auctions. *Games and Economic Behavior* 55(1):131–151.

Letcher A, Balduzzi D, Racanière S, Martens J, Foerster JN, Tuyls K, Graepel T (2019) Differentiable game mechanics. *Journal of Machine Learning Research* 20(84):1–40.

Marshall RC, Meurer MJ, Richard JF, Stromquist W (1994) Numerical analysis of asymmetric first price auctions. *Games and Economic Behavior* 7(2):193–220.

Maskin E, Riley J (2000) Asymmetric auctions. *The Review of Economic Studies* 67(3):413–438.

Mertikopoulos P, Zhou Z (2019) Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming* 173(1-2):465–507.

Milgrom P (2017) *Discovering prices: auction design in markets with complex constraints* (Columbia University Press).

Monderer D, Shapley LS (1996) Potential games. *Games and Economic Behavior* 14(1):124–143.

Nash JF, et al. (1950) Equilibrium points in n-person games. *Proceedings of the national academy of sciences* 36(1):48–49.

Ott M, Beck M (2013) Incentives for overbidding in minimum-revenue core-selecting auctions. Number F16-V3 in Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Auctions and Licensing (Kiel und Hamburg: ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft).

Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, Lin Z, Desmaison A, Antiga L, Lerer A (2017) Automatic differentiation in PyTorch. *NIPS-W*.

Plum M (1992) Characterization and computation of Nash-equilibria for auctions with incomplete information. *International Journal of Game Theory* 20(4):393–418.

Rubinstein A (2016) Settling the complexity of computing approximate two-player Nash equilibria. *arXiv:1606.04550 [cs]* URL `http://arxiv.org/abs/1606.04550`.

Salimans T, Ho J, Chen X, Sidor S, Sutskever I (2017) Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *arXiv:1703.03864 [cs, stat]* .

Sanders JB, Farmer JD, Galla T (2018) The prevalence of chaotic dynamics in games with many players. *Scientific reports* 8(1):1–13.

Schäfer F, Anandkumar A (2019) Competitive gradient descent. *Advances in Neural Information Processing Systems*, 7623–7633.

Singh SP, Kearns MJ, Mansour Y (2000) Nash convergence of gradient dynamics in general-sum games. *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI2000)*, 541–548.

Ui T (2016) Bayesian nash equilibrium and variational inequalities. *Journal of Mathematical Economics* 63:139–146.

Vickrey W (1961) Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance* 16(1):8–37.

Zinkevich M (2003) Online convex programming and generalized infinitesimal gradient ascent. *Proceedings of the 20th International Conference on Machine Learning (icml-03)*, 928–936.

List of Symbols

$\mathcal{A}$ The set of feasible *action profiles* in a Bayesian game, i.e., *bid profiles* in actions. Cross product of individual players' action sets: $\mathcal{A} \equiv \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$

β A joint *strategy profile* in a Bayesian Game.

β^* A strategy profile that constitutes a Bayesian Nash equilibrium.

β_i A feasible *pure strategy* of player i: $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$.

b An action/bid profile. $b \in \mathcal{A}$

$B(\alpha, \beta)$ The Beta-distribution with shape parameters α and β.

b_i An action/bid for player i. $b_i \in \mathcal{A}_i$.

C Efficiency parameter in split-award auction setting. See Subsection 6.4.

d_i The dimension of the parameter vector θ_i of player i's neural network π_i.

ϵ The approximation-bound in an approximate BNE, indicating that each player's incentive to deviate is less than $\epsilon \geq 0$.

$\hat{\epsilon}$ An ex-post estimator for the worst-case ex-interim loss. Does not require access to an analytical BNE. See Section 5.

ε_p The Gaussian noise vector of perturbation p in NPGA gradient computation.

ϕ_p The *fitness* of perturbation $\theta_{i;p}$ of player i's neural network in NPGA gradient computation.

F_v The joint prior distribution over types, marginalized by F_{v_i}.

G A Bayesian Game $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$.

$-i$ Index identifying a *partial* action/bid/strategy profile for all bidders except player i.

$\mathcal{I}$ The set of players in a game. Indexed by i. Total number of players is n.

i Index identifying a particular player.

$\mathcal{K}$ The set of feasible bundles of items, generally the power set of $\mathcal{M}$.

$\hat{\ell}_i$ An estimator of i's ex-ante loss $\tilde{\ell}_i(\beta_i, \beta_{-i})$ computed ex-post from observational data. Does not require access to an analytical BNE. See Section 5.

$\hat{\lambda}_i$ Auxiliary quantity in the computation of $\hat{\ell}$ and $\hat{\epsilon}$. $\hat{\lambda}_i(v_i; b_i, \beta_{-i})$ constitutes an ex-post estimator for the interim utility loss $\bar{\ell}_i$ of playing b_i at valuation v_i. See Equation 5.3.

$\hat{\mathcal{L}}$ An ex-post estimator for the *relative utility loss* $\mathcal{L}(\beta_i)$, when no access to the analytical BNE is available. See Equation 5.4.

$\mathcal{L}$ The *relative utility loss* $\mathcal{L}(\beta_i)$ of strategy β_i compared to an analytical BNE β^*. See Equation 5.1.

$\bar{\ell}_i$ The *interim* utility loss of player i. See Equation 3.3.

$\tilde{\ell}_i$ The *ex-ante* utility loss of player i. See Equation 3.7.

L_2 The L2-loss $L_2(\beta_i)$ of a strategy β_i compared to BNE β^*, i.e., the distance of β_i and β_i^* in strategy space. See Equation 5.2.

$\mathcal{M}$ The set of items sold in an auction. Total number of items is m. Items can be homogenous or heterogeneous.

m Total number of items in an auction.

$\mathcal{N}(\mu, \sigma^2)$ Gaussian Distribution with mean μ and standard deviation σ.

$\mathbb{N}$ The set of natural numbers.

n The total number of players in a game.

n_{batch} Batch size used in sampling opponent behavior when computing ex-post estimators $\hat{\ell}$ and $\hat{\epsilon}$. See Section 5.

n_{grid} Size of the discrete grid of alternative bids evaluated to compute ex-post estimators $\hat{\ell}$ and $\hat{\epsilon}$. See Section 5.

$\mathcal{P}_{\Sigma_i}$ The projection function onto the set Σ_i.

π_i Neural network for player i, implementing i bidding strategy β_i via $\beta_i(v_i) := \pi_i(v_i; \theta_i)$, where $\theta_i \in \Theta_i = \mathbb{R}^{d_i}$ are the network's parameters.

P Hyperparameter in NPGA. The *population size*, or number of perturbations of θ_i considered for each iteration of gradient computation.

p Index used for permutations $1, \ldots, P$ in NPGA gradient computation.

p_i Price paid by bidder i to the auctioneer after receiving bundle x_i.

$\mathbb{R}$ The set of real numbers.

Σ The set of feasible joint strategy profiles.

σ Hyperparameter in NPGA. The standard deviation of the Gaussian noise used in permuting neural network parameters.

Σ_i The set of feasible (pure) *strategies* for player i. Generally an infinite-dimensionally Hilbert space.

θ_i The parameter vector of player i's neural network π_i.

t Time / iteration number.

$\hat{u}_i$ An estimator of ex-ante expected utility $\tilde{u}$, computed ex-post via Monte Carlo integration over a large batch of realizations of $v \sim F_v$.

$\mathcal{U}(l,h)$ Uniform distribution with lower bound l and upper bound h.

$\overline{u}_i$ The expected *interim* utility $\overline{u}_i(v_i, b_i, \beta_{-i})$ of player i. See Equation 3.2.

$\tilde{u}_i$ The expected *ex-ante* utility $\tilde{u}_i(\beta_i, \beta_{-i})$ of player i. See Equation 3.5.

u_i The ex-post utility function $u_i(v_i, b_i, b_{-i})$ of player i. Generally nondifferentiable.

$\mathcal{V}$ The set of possible *valuation profiles*, i.e., generally the support of F_v.

v The *private valuation* or *type* profile, $v \in \mathcal{V}$. Generally used to refer to the Random Variable, sometimes also used to refer to a realization of the RV.

v_i The private valuation of player i. Generally a random vector of length 2^m, indicating i's willingness to pay when allocated a certain bundle. We also write $v_i(x_i)$ for the entry of v_i corresponding to the (scalar) valuation of player i for bundle $x_i \in \mathcal{K}$.

x_i The bundle of items allocated to player i. $x_i \in \mathcal{K}$.

Chapter 6

Learning in Double Auctions

Peer-Reviewed Journal Paper

Title: Learning equilibrium in bilateral bargaining games.

Authors: Martin Bichler, Nils Kohring, Matthias Oberlechner, Fabian R. Pieroth.

In: Elsevier European Journal of Operational Research.

Abstract: Bilateral bargaining of a single good among one buyer and one seller describes the simplest form of trade, yet Bayes-Nash equilibrium strategies are largely unknown. Only for the average mechanism in the standard independent private values model with independent and uniform priors, we know that there is a continuum of equilibria. However, a non-uniform prior distribution already leads to a system of non-linear differential equations for which closed-form bidding strategies cannot be derived. Recent advances in equilibrium learning provide a numerical approach to equilibrium analysis, which can push the boundaries of existing results and allow for the analysis of environments that have been considered intractable so far. We study Neural Pseudogradient Ascent (NPGA) and Simultaneous Online Dual Averaging (SODA), two new equilibrium learning algorithms for Bayesian auction games with continuous type and action spaces. Although the environment is simple to describe, the continuum of equilibria makes it challenging for equilibrium learning algorithms. Empirically, NPGA finds the payoff-maximizing linear equilibrium, while SODA also finds non-differentiable step-function equilibria. Interestingly, the algorithms also find equilibrium with non-uniform priors and risk-averse traders for which we do not know an analytical solution. We show that the game is not globally monotone, but we can prove local convergence for a model with uniform priors and linear bid functions.

Citation: Bichler et al. (2022).

CCC Marketplace

This is a License Agreement between Nils Kohring ("User") and Copyright Clearance Center, Inc. ("CCC") on behalf of the Rightsholder identified in the order details below. The license consists of the order details, the Marketplace Permissions General Terms and Conditions below, and any Rightsholder Terms and Conditions which are included below.

All payments must be made in full to CCC in accordance with the Marketplace Permissions General Terms and Conditions below.

Order Date	11-Apr-2023	**Type of Use**	Republish in a thesis/dissertation
Order License ID	1343824-1	**Publisher**	ELSEVIER BV
ISSN	0377-2217	**Portion**	Chapter/article

LICENSED CONTENT

Publication Title	European journal of operational research	**Language**	Dutch
Article Title	Learning equilibrium in bilateral bargaining games	**Country**	Netherlands
Author/Editor	ASSOCIATION OF EUROPEAN OPERATIONAL RESEARCH SOCIE	**Rightsholder**	Elsevier Science & Technology Journals
Date	01/01/1977	**Publication Type**	Journal

REQUEST DETAILS

Portion Type	Chapter/article	**Rights Requested**	Main product
Page Range(s)	1-19	**Distribution**	Worldwide
Total Number of Pages	19	**Translation**	Original language of publication
Format (select all that apply)	Electronic	**Copies for the Disabled?**	No
Who Will Republish the Content?	Author of requested content	**Minor Editing Privileges?**	No
Duration of Use	Life of current edition	**Incidental Promotional Use?**	No
Lifetime Unit Quantity	Up to 499	**Currency**	EUR

NEW WORK DETAILS

Title	Multi-Agent Reinforcement Learning for the Computation of Market Equilibria	**Institution Name**	Technical University of Munich
Instructor Name	Prof. Dr. Martin Bichler	**Expected Presentation Date**	2023-09-01

ADDITIONAL DETAILS

Order Reference Number	N/A	**The Requesting Person/Organization to Appear on the License**	Nils Kohring

REQUESTED CONTENT DETAILS

Title, Description or Numeric Reference of the Portion(s)	Full article
Editor of Portion(s)	N/A
Volume of Serial or Monograph	N/A
Page or Page Range of Portion	1-19

Title of the Article/Chapter the Portion Is From	Learning equilibrium in bilateral bargaining games
Author of Portion(s)	Bichler, Martin; Kohring, Nils; Oberlechner, Matthias; Pieroth, Fabian
Issue, if Republishing an Article From a Serial	N/A
Publication Date of Portion	2022-12-01

RIGHTSHOLDER TERMS AND CONDITIONS

Elsevier publishes Open Access articles in both its Open Access journals and via its Open Access articles option in subscription journals, for which an author selects a user license permitting certain types of reuse without permission. Before proceeding please check if the article is Open Access on http://www.sciencedirect.com and refer to the user license for the individual article. Any reuse not included in the user license terms will require permission. You must always fully and appropriately credit the author and source. If any part of the material to be used (for example, figures) has appeared in the Elsevier publication for which you are seeking permission, with credit or acknowledgement to another source it is the responsibility of the user to ensure their reuse complies with the terms and conditions determined by the rights holder. Please contact permissions@elsevier.com with any queries.

Marketplace Permissions General Terms and Conditions

The following terms and conditions ("General Terms"), together with any applicable Publisher Terms and Conditions, govern User's use of Works pursuant to the Licenses granted by Copyright Clearance Center, Inc. ("CCC") on behalf of the applicable Rightsholders of such Works through CCC's applicable Marketplace transactional licensing services (each, a "Service").

1) **Definitions.** For purposes of these General Terms, the following definitions apply:

"License" is the licensed use the User obtains via the Marketplace platform in a particular licensing transaction, as set forth in the Order Confirmation.

"Order Confirmation" is the confirmation CCC provides to the User at the conclusion of each Marketplace transaction. "Order Confirmation Terms" are additional terms set forth on specific Order Confirmations not set forth in the General Terms that can include terms applicable to a particular CCC transactional licensing service and/or any Rightsholder-specific terms.

"Rightsholder(s)" are the holders of copyright rights in the Works for which a User obtains licenses via the Marketplace platform, which are displayed on specific Order Confirmations.

"Terms" means the terms and conditions set forth in these General Terms and any additional Order Confirmation Terms collectively.

"User" or "you" is the person or entity making the use granted under the relevant License. Where the person accepting the Terms on behalf of a User is a freelancer or other third party who the User authorized to accept the General Terms on the User's behalf, such person shall be deemed jointly a User for purposes of such Terms.

"Work(s)" are the copyright protected works described in relevant Order Confirmations.

2) **Description of Service.** CCC's Marketplace enables Users to obtain Licenses to use one or more Works in accordance with all relevant Terms. CCC grants Licenses as an agent on behalf of the copyright rightsholder identified in the relevant Order Confirmation.

3) **Applicability of Terms.** The Terms govern User's use of Works in connection with the relevant License. In the event of any conflict between General Terms and Order Confirmation Terms, the latter shall govern. User acknowledges that Rightsholders have complete discretion whether to grant any permission, and whether to place any limitations on any grant, and that CCC has no right to supersede or to modify any such discretionary act by a Rightsholder.

4) **Representations; Acceptance.** By using the Service, User represents and warrants that User has been duly authorized by the User to accept, and hereby does accept, all Terms.

5) **Scope of License; Limitations and Obligations.** All Works and all rights therein, including copyright rights, remain the sole and exclusive property of the Rightsholder. The License provides only those rights expressly set forth in the terms

and conveys no other rights in any Works

6) **General Payment Terms.** User may pay at time of checkout by credit card or choose to be invoiced. If the User chooses to be invoiced, the User shall: (i) remit payments in the manner identified on specific invoices, (ii) unless otherwise specifically stated in an Order Confirmation or separate written agreement, Users shall remit payments upon receipt of the relevant invoice from CCC, either by delivery or notification of availability of the invoice via the Marketplace platform, and (iii) if the User does not pay the invoice within 30 days of receipt, the User may incur a service charge of 1.5% per month or the maximum rate allowed by applicable law, whichever is less. While User may exercise the rights in the License immediately upon receiving the Order Confirmation, the License is automatically revoked and is null and void, as if it had never been issued, if CCC does not receive complete payment on a timely basis.

7) **General Limits on Use.** Unless otherwise provided in the Order Confirmation, any grant of rights to User (i) involves only the rights set forth in the Terms and does not include subsequent or additional uses, (ii) is non-exclusive and non-transferable, and (iii) is subject to any and all limitations and restrictions (such as, but not limited to, limitations on duration of use or circulation) included in the Terms. Upon completion of the licensed use as set forth in the Order Confirmation, User shall either secure a new permission for further use of the Work(s) or immediately cease any new use of the Work(s) and shall render inaccessible (such as by deleting or by removing or severing links or other locators) any further copies of the Work. User may only make alterations to the Work if and as expressly set forth in the Order Confirmation. No Work may be used in any way that is unlawful, including without limitation if such use would violate applicable sanctions laws or regulations, would be defamatory, violate the rights of third parties (including such third parties' rights of copyright, privacy, publicity, or other tangible or intangible property), or is otherwise illegal, sexually explicit, or obscene. In addition, User may not conjoin a Work with any other material that may result in damage to the reputation of the Rightsholder. Any unlawful use will render any licenses hereunder null and void. User agrees to inform CCC if it becomes aware of any infringement of any rights in a Work and to cooperate with any reasonable request of CCC or the Rightsholder in connection therewith.

8) **Third Party Materials.** In the event that the material for which a License is sought includes third party materials (such as photographs, illustrations, graphs, inserts and similar materials) that are identified in such material as having been used by permission (or a similar indicator), User is responsible for identifying, and seeking separate licenses (under this Service, if available, or otherwise) for any of such third party materials; without a separate license, User may not use such third party materials via the License.

9) **Copyright Notice.** Use of proper copyright notice for a Work is required as a condition of any License granted under the Service. Unless otherwise provided in the Order Confirmation, a proper copyright notice will read substantially as follows: "Used with permission of [Rightsholder's name], from [Work's title, author, volume, edition number and year of copyright]; permission conveyed through Copyright Clearance Center, Inc." Such notice must be provided in a reasonably legible font size and must be placed either on a cover page or in another location that any person, upon gaining access to the material which is the subject of a permission, shall see, or in the case of republication Licenses, immediately adjacent to the Work as used (for example, as part of a by-line or footnote) or in the place where substantially all other credits or notices for the new work containing the republished Work are located. Failure to include the required notice results in loss to the Rightsholder and CCC, and the User shall be liable to pay liquidated damages for each such failure equal to twice the use fee specified in the Order Confirmation, in addition to the use fee itself and any other fees and charges specified.

10) **Indemnity.** User hereby indemnifies and agrees to defend the Rightsholder and CCC, and their respective employees and directors, against all claims, liability, damages, costs, and expenses, including legal fees and expenses, arising out of any use of a Work beyond the scope of the rights granted herein and in the Order Confirmation, or any use of a Work which has been altered in any unauthorized way by User, including claims of defamation or infringement of rights of copyright, publicity, privacy, or other tangible or intangible property.

11) **Limitation of Liability.** UNDER NO CIRCUMSTANCES WILL CCC OR THE RIGHTSHOLDER BE LIABLE FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL, OR INCIDENTAL DAMAGES (INCLUDING WITHOUT LIMITATION DAMAGES FOR LOSS OF BUSINESS PROFITS OR INFORMATION, OR FOR BUSINESS INTERRUPTION) ARISING OUT OF THE USE OR INABILITY TO USE A WORK, EVEN IF ONE OR BOTH OF THEM HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. In any event, the total liability of the Rightsholder and CCC (including their respective employees and directors) shall not exceed the total amount actually paid by User for the relevant License. User assumes full liability for the actions and omissions of its principals, employees, agents, affiliates, successors, and assigns.

12) **Limited Warranties.** THE WORK(S) AND RIGHT(S) ARE PROVIDED "AS IS." CCC HAS THE RIGHT TO GRANT TO USER THE RIGHTS GRANTED IN THE ORDER CONFIRMATION DOCUMENT. CCC AND THE RIGHTSHOLDER DISCLAIM ALL OTHER WARRANTIES RELATING TO THE WORK(S) AND RIGHT(S), EITHER EXPRESS OR IMPLIED, INCLUDING WITHOUT LIMITATION IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. ADDITIONAL RIGHTS MAY BE REQUIRED TO USE ILLUSTRATIONS, GRAPHS, PHOTOGRAPHS, ABSTRACTS, INSERTS, OR OTHER PORTIONS OF THE WORK (AS OPPOSED TO THE ENTIRE WORK) IN A MANNER CONTEMPLATED BY USER; USER UNDERSTANDS AND AGREES THAT NEITHER CCC NOR THE RIGHTSHOLDER MAY HAVE SUCH ADDITIONAL RIGHTS TO GRANT.

13) **Effect of Breach.** Any failure by User to pay any amount when due, or any use by User of a Work beyond the scope of the License set forth in the Order Confirmation and/or the Terms, shall be a material breach of such License. Any breach not cured within 10 days of written notice thereof shall result in immediate termination of such License without further notice. Any unauthorized (but licensable) use of a Work that is terminated immediately upon notice thereof may be liquidated by payment of the Rightsholder's ordinary license price therefor; any unauthorized (and unlicensable) use that is not terminated immediately for any reason (including, for example, because materials containing the Work cannot reasonably be recalled) will be subject to all remedies available at law or in equity, but in no event to a payment of less than three times the Rightsholder's ordinary license price for the most closely analogous licensable use plus Rightsholder's and/or CCC's costs and expenses incurred in collecting such payment.

14) **Additional Terms for Specific Products and Services.** If a User is making one of the uses described in this Section 14, the additional terms and conditions apply:

a) ***Print Uses of Academic Course Content and Materials (photocopies for academic coursepacks or classroom handouts).*** For photocopies for academic coursepacks or classroom handouts the following additional terms apply:

i) The copies and anthologies created under this License may be made and assembled by faculty members individually or at their request by on-campus bookstores or copy centers, or by off-campus copy shops and other similar entities.

ii) No License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied) (ii) permit "publishing ventures" where any particular anthology would be systematically marketed at multiple institutions.

iii) Subject to any Publisher Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the academic pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to no more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular anthology, whether photocopied or electronic, at more than one institution of learning;

E) in the case of a photocopy permission, no materials may be entered into electronic memory by User except in order to produce an identical copy of a Work before or during the academic term (or analogous period) as to which any particular permission is granted. In the event that User shall choose to retain materials that are the subject of a photocopy permission in electronic memory for purposes of producing identical copies more than one day after such retention (but still within the scope of any permission granted), User must notify CCC of such fact in the applicable permission request and such retention shall constitute one copy actually sold for purposes of calculating permission fees due; and

F) any permission granted shall expire at the end of the class. No permission granted shall in any way include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied).

iv) Books and Records; Right to Audit. As to each permission granted under the academic pay-per-use Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any photocopies sold or by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this License for any reason.

b) ***Digital Pay-Per-Uses of Academic Course Content and Materials (e-coursepacks, electronic reserves, learning management systems, academic institution intranets).*** For uses in e-coursepacks, posts in electronic reserves, posts in learning management systems, or posts on academic institution intranets, the following additional terms apply:

i) The pay-per-uses subject to this Section 14(b) include:

A) **Posting e-reserves, course management systems, e-coursepacks for text-based content,** which grants authorizations to import requested material in electronic format, and allows electronic access to this material to members of a designated college or university class, under the direction of an instructor designated by the college or university, accessible only under appropriate electronic controls (e.g., password);

B) **Posting e-reserves, course management systems, e-coursepacks for material consisting of photographs or other still images not embedded in text,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorization: to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above, including any necessary resizing, reformatting or modification of the resolution of such requested material (provided that such modification does not alter the underlying editorial content or meaning of the requested material, and provided that the resulting modified content is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms), but not including any other form of manipulation, alteration or editing of the requested material;

C) **Posting e-reserves, course management systems, e-coursepacks or other academic distribution for audiovisual content,** which grants not only the authorizations described in Section 14(b)(i)(A) above, but also the following authorizations: (i) to include the requested material in course materials for use consistent with Section 14(b)(i)(A) above; (ii) to display and perform the requested material to such members of such class in the physical classroom or remotely by means of streaming media or other video formats; and (iii) to "clip" or reformat the requested material for purposes of time or content management or ease of delivery, provided that such “clipping” or reformatting does not alter the underlying editorial content or meaning of the requested material and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular authorization described in the Order Confirmation and the Terms. Unless expressly set forth in the relevant Order Conformation, the License does not authorize any other form of manipulation, alteration or editing of the requested material.

ii) Unless expressly set forth in the relevant Order Confirmation, no License granted shall in any way: (i) include any right by User to create a substantively non-identical copy of the Work or to edit or in any other way modify the Work (except by means of deleting material immediately preceding or following the entire portion of the Work copied or, in the case of Works subject to Sections 14(b)(1)(B) or (C) above, as described in such Sections) (ii) permit "publishing ventures" where any particular course materials would be systematically marketed at multiple institutions.

iii) Subject to any further limitations determined in the Rightsholder Terms (and notwithstanding any apparent contradiction in the Order Confirmation arising from data provided by User), any use authorized under the electronic course content pay-per-use service is limited as follows:

A) any License granted shall apply to only one class (bearing a unique identifier as assigned by the institution, and thereby including all sections or other subparts of the class) at one institution;

B) use is limited to not more than 25% of the text of a book or of the items in a published collection of essays, poems or articles;

C) use is limited to not more than the greater of (a) 25% of the text of an issue of a journal or other periodical or (b) two articles from such an issue;

D) no User may sell or distribute any particular materials, whether photocopied or electronic, at more than one institution of learning;

E) electronic access to material which is the subject of an electronic-use permission must be limited by means of electronic password, student identification or other control permitting access solely to students and instructors in the class;

F) User must ensure (through use of an electronic cover page or other appropriate means) that any person, upon gaining electronic access to the material, which is the subject of a permission, shall see:

- a proper copyright notice, identifying the Rightsholder in whose name CCC has granted permission,
- a statement to the effect that such copy was made pursuant to permission,
- a statement identifying the class to which the material applies and notifying the reader that the material has been made available electronically solely for use in the class, and

- a statement to the effect that the material may not be further distributed to any person outside the class, whether by copying or by transmission and whether electronically or in paper form, and User must also ensure that such cover page or other means will print out in the event that the person accessing the material chooses to print out the material or any part thereof.

G) any permission granted shall expire at the end of the class and, absent some other form of authorization, User is thereupon required to delete the applicable material from any electronic storage or to block electronic access to the applicable material.

iv) Uses of separate portions of a Work, even if they are to be included in the same course material or the same university or college class, require separate permissions under the electronic course content pay-per-use Service. Unless otherwise provided in the Order Confirmation, any grant of rights to User is limited to use completed no later than the end of the academic term (or analogous period) as to which any particular permission is granted.

v) Books and Records; Right to Audit. As to each permission granted under the electronic course content Service, User shall maintain for at least four full calendar years books and records sufficient for CCC to determine the numbers of copies made by User under such permission. CCC and any representatives it may designate shall have the right to audit such books and records at any time during User's ordinary business hours, upon two days' prior notice. If any such audit shall determine that User shall have underpaid for, or underreported, any electronic copies used by three percent (3%) or more, then User shall bear all the costs of any such audit; otherwise, CCC shall bear the costs of any such audit. Any amount determined by such audit to have been underpaid by User shall immediately be paid to CCC by User, together with interest thereon at the rate of 10% per annum from the date such amount was originally due. The provisions of this paragraph shall survive the termination of this license for any reason.

c) ***Pay-Per-Use Permissions for Certain Reproductions (Academic photocopies for library reserves and interlibrary loan reporting) (Non-academic internal/external business uses and commercial document delivery).*** The License expressly excludes the uses listed in Section (c)(i)-(v) below (which must be subject to separate license from the applicable Rightsholder) for: academic photocopies for library reserves and interlibrary loan reporting; and non-academic internal/external business uses and commercial document delivery.

i) electronic storage of any reproduction (whether in plain-text, PDF, or any other format) other than on a transitory basis;

ii) the input of Works or reproductions thereof into any computerized database;

iii) reproduction of an entire Work (cover-to-cover copying) except where the Work is a single article;

iv) reproduction for resale to anyone other than a specific customer of User;

v) republication in any different form. Please obtain authorizations for these uses through other CCC services or directly from the rightsholder.

Any license granted is further limited as set forth in any restrictions included in the Order Confirmation and/or in these Terms.

d) ***Electronic Reproductions in Online Environments (Non-Academic-email, intranet, internet and extranet).*** For "electronic reproductions", which generally includes e-mail use (including instant messaging or other electronic transmission to a defined group of recipients) or posting on an intranet, extranet or Intranet site (including any display or performance incidental thereto), the following additional terms apply:

i) Unless otherwise set forth in the Order Confirmation, the License is limited to use completed within 30 days for any use on the Internet, 60 days for any use on an intranet or extranet and one year for any other use, all as measured from the "republication date" as identified in the Order Confirmation, if any, and otherwise from the date of the Order Confirmation.

ii) User may not make or permit any alterations to the Work, unless expressly set forth in the Order Confirmation (after request by User and approval by Rightsholder); provided, however, that a Work consisting of photographs or other still images not embedded in text may, if necessary, be resized, reformatted or have its resolution modified without additional express permission, and a Work consisting of audiovisual content may, if necessary, be "clipped" or reformatted for purposes of time or content management or ease of delivery (provided that any such resizing, reformatting, resolution modification or “clipping” does not alter the underlying editorial content or meaning of the Work used, and that the resulting material is used solely within the scope of, and in a manner consistent with, the particular License described in the Order Confirmation and the Terms.

15) **Miscellaneous.**

a) User acknowledges that CCC may, from time to time, make changes or additions to the Service or to the Terms, and that Rightsholder may make changes or additions to the Rightsholder Terms. Such updated Terms will replace the prior terms and conditions in the order workflow and shall be effective as to any subsequent Licenses but shall not apply to Licenses already granted and paid for under a prior set of terms.

b) Use of User-related information collected through the Service is governed by CCC's privacy policy, available online at www.copyright.com/about/privacy-policy/.

c) The License is personal to User. Therefore, User may not assign or transfer to any other person (whether a natural person or an organization of any kind) the License or any rights granted thereunder; provided, however, that, where applicable, User may assign such License in its entirety on written notice to CCC in the event of a transfer of all or substantially all of User's rights in any new material which includes the Work(s) licensed under this Service.

d) No amendment or waiver of any Terms is binding unless set forth in writing and signed by the appropriate parties, including, where applicable, the Rightsholder. The Rightsholder and CCC hereby object to any terms contained in any writing prepared by or on behalf of the User or its principals, employees, agents or affiliates and purporting to govern or otherwise relate to the License described in the Order Confirmation, which terms are in any way inconsistent with any Terms set forth in the Order Confirmation, and/or in CCC's standard operating procedures, whether such writing is prepared prior to, simultaneously with or subsequent to the Order Confirmation, and whether such writing appears on a copy of the Order Confirmation or in a separate instrument.

e) The License described in the Order Confirmation shall be governed by and construed under the law of the State of New York, USA, without regard to the principles thereof of conflicts of law. Any case, controversy, suit, action, or proceeding arising out of, in connection with, or related to such License shall be brought, at CCC's sole discretion, in any federal or state court located in the County of New York, State of New York, USA, or in any federal or state court whose geographical jurisdiction covers the location of the Rightsholder set forth in the Order Confirmation. The parties expressly submit to the personal jurisdiction and venue of each such federal or state court.

Last updated October 2022

European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Analytics, Computational Intelligence and Information Management

Learning equilibrium in bilateral bargaining games

Martin Bichler[*], Nils Kohring, Matthias Oberlechner, Fabian R. Pieroth

Department of Computer Science, Technical University of Munich, Garching 85748, Germany

ARTICLE INFO

Article history:
Received 11 May 2022
Accepted 15 December 2022
Available online xxx

Keywords:
Auctions bidding
Game theory
Machine learning

ABSTRACT

Bilateral bargaining of a single good among one buyer and one seller describes the simplest form of trade, yet Bayes–Nash equilibrium strategies are largely unknown. Only for the average mechanism in the standard independent private values model with independent and uniform priors, we know that there is a continuum of equilibria. However, a non-uniform prior distribution already leads to a system of non-linear differential equations for which closed-form bidding strategies cannot be derived. Recent advances in equilibrium learning provide a numerical approach to equilibrium analysis, which can push the boundaries of existing results and allow for the analysis of environments that have been considered intractable so far. We study Neural Pseudogradient Ascent (NPGA) and Simultaneous Online Dual Averaging (SODA), two new equilibrium learning algorithms for Bayesian auction games with continuous type and action spaces. Although the environment is simple to describe, the continuum of equilibria makes it challenging for equilibrium learning algorithms. Empirically, NPGA finds the payoff-maximizing linear equilibrium, while SODA also finds non-differentiable step-function equilibria. Interestingly, the algorithms also find equilibrium with non-uniform priors and risk-averse traders for which we do not know an analytical solution. We show that the game is not globally monotone, but we can prove local convergence for a model with uniform priors and linear bid functions.

1. Introduction

Trade in some of the most important markets for homogenous goods is governed by double auctions. For example, major exchanges use versions of a double auction for trading stocks, bonds, agricultural commodities, metals, and derivative securities (Friedman, 1992). Yet, the game-theoretical analysis of such simple institutions has turned out challenging. Even the simple bilateral trade model with only one buyer, one seller, and one indivisible good has led to several decades of research trying to prove existence and equilibrium bidding strategies under different assumptions. The strategic problem of the traders in this literature is usually modeled as a Bayesian game. In the independent private values model, both buyers know their value ex-interim but only have distributional information about the opponent's value. In a seminal paper, Myerson & Satterthwaite (1983) showed that no mechanism simultaneously satisfies individual rationality, budget balance, incentive-compatibility, and efficiency in bilateral trade.

The Vickrey-Clarke-Groves (VCG) mechanism is individually rational, incentive-compatible, and efficient, but not budget-balanced in such two-sided markets, which provides a reason why it can rarely be found in practice. As a result, the analysis of non-truthful mechanisms has received significant attention. The k-double auction has assumed a central role in the literature (Gresik, 2011; Kadan, 2007; Leininger et al., 1989; Satterthwaite & Williams, 1989; Satterthwaite et al., 2022). It is not incentive-compatible but simple and closer to real-world practices such as a uniform price call market as it is often used on financial markets, where there is a single price at which all trades are cleared. The k-double auction determines the terms of trade when a buyer and a seller negotiate the sale of an item. The buyer submits a bid b, and the seller submits an ask s. Trade occurs if b exceeds s at a price $kb + (1-k)s$. For example, if $k = 0.5$, this is the *average mechanism* or 0.5-double auction. Given that the traders' reports affect the price and the likelihood of trade in the average mechanism, there is an incentive to misrepresent the true value. As a result of this strategic bidding, some trades that could happen do not, which leads to an efficiency loss. The model is so simple to explain that it has become central to the equilibrium analysis of trading mechanisms. Wilson (1985) argues that understanding bilateral bargaining provides a foundation for a theory of large markets.

Yet, even for this simple and central model of trade, we only know equilibrium bidding strategies for very restricted model assumptions. In a seminal contribution, Leininger et al. (1989) analyze the average mechanism with independent private values and

[*] Corresponding author.
E-mail address: bichler@in.tum.de (M. Bichler).

https://doi.org/10.1016/j.ejor.2022.12.022
0377-2217/© 2022 Elsevier B.V. All rights reserved.

Please cite this article as: M. Bichler, N. Kohring, M. Oberlechner et al., Learning equilibrium in bilateral bargaining games, European Journal of Operational Research, https://doi.org/10.1016/j.ejor.2022.12.022

quasi-linear utility functions and find a multitude of equilibria. One family of equilibria has differentiable strategies, another family is composed of (non-differentiable) step-functions with arbitrarily many jumps. In the earlier paper by Chatterjee & Samuelson (1983) also strategies for risk-averse bidders were derived in this setting. While some extensions have been analyzed (e.g., with general k, interdependent private values, or multiple bidders), explicit equilibrium bid functions are unavailable. The equilibrium problem in the k-double auction and many other auction models can often be described as the solution to a system of differential equations. Unless there are simple (uniform) distributional assumptions and simple assumptions about the bidders' utility functions and the goods, we typically do not have a general solution theory. Even setting up the differential equations can be challenging.

1.1. Equilibrium computation

Numerical methods for computing approximate equilibria in Bayesian games with continuous type and action space would be very useful for equilibrium analysis and comparative statics. Actually, there is a long history of thought on equilibrium computation in operational research (Bigi et al., 2013; Jofré et al., 2007). However, while there has been significant research on equilibrium computation in complete-information n-player games with finite actions and players, the computation of Bayes–Nash equilibria (BNE) in games with continuous type- and action-spaces, as they are used to model auctions, has received little attention. The infinite type-space is a key challenge because equilibrium computation algorithms need to find an equilibrium bid function of unknown shape.

Only recently, there have been a number of advances in developing equilibrium learning methods with notable success in single-sided auction games (see Section 2). *Neural Pseudogradient Ascent* (NPGA) (Bichler et al., 2021) and *Simultaneous Online Dual Averaging* (SODA) (Fichtl et al., 2022) have led to breakthroughs providing versatile equilibrium solvers that find equilibrium in a wide variety of single-sided auctions, including single-object, multi-unit, and combinatorial auctions. NPGA and SODA are both based on simultaneous gradient ascent on the expected utility function of each player. Both methods allow for interdependent types and various utility functions, including ones with risk or loss aversion. While NPGA learns approximate pure Bayes–Nash equilibria using self-play and neural networks, SODA learns distributional strategies on a discretized version of the game. Although there is not yet a complete theory of games that are "learnable" and those that are not, we know that if SODA converges to a pure strategy, then it is an equilibrium.

Unfortunately, identifying characteristics of games where gradient-based algorithms converge to a BNE turned out to be a daunting task. Recent results on complete-information normal-form games showed that gradient dynamics either circle, diverge, or are even chaotic (Sanders et al., 2018). Actually, the study of gradient dynamics in games is akin to studying dynamical systems and characterizing environments, where gradient dynamics converge to a Nash equilibrium (if one exists), has been described as arbitrarily complex (Andrade et al., 2021). The study of Bayesian games with continuous action and type space adds a layer of complexity. This is because we not only need to learn an equilibrium bid but a bid function that can take an arbitrary shape. The fact that we do find equilibrium consistently in a wide variety of auction games demands a closer look. The k-double auction with one buyer and one seller is the simplest environment that still captures the main challenges of the equilibrium computation in auction mechanisms and allows us deeper insights into the reasons for convergence in this paper.

1.2. Contributions

The contributions of this article are two-fold: First, we provide a novel convergence result for NPGA for the bilateral bargaining model. Already the convergence analysis of gradient dynamics in this simple model is very challenging. The difficulty arises from the fact that the equilibrium problem is a system of non-linear ordinary differential equations that has the inverse of an unknown bid function as one of its components. There is no analytical solution theory for such differential equations for general priors, and even standard numerical techniques for solving differential equations lead to problems, as we will discuss.

If a game satisfies a payoff monotonicity condition, no-regret learning algorithms are known to converge to an equilibrium in continuous- and finite-action games. This corresponds to monotonicity in variational inequalities, which guarantees convergence of various algorithms. In the bilateral trade environment with uniformly distributed types, we know that there is a linear equilibrium strategy for both traders. Assuming that we know that the equilibrium bid function is linear, we can explore the expected utility function of each player and check for monotonicity. Unfortunately, we can show via an explicit counterexample that the monotonicity condition is not satisfied globally. However, the assumption of a linear bid function allows us to show local convergence of the NPGA equilibrium learning algorithm. More precisely, we prove that in the 0.5-double auction with two quasi-linear traders and linear strategies, the NPGA equilibrium learner will converge locally. Our analysis of this restricted bilateral trade model sheds light on the question why it is so difficult to provide a priori convergence guarantees for gradient dynamics in more general Bayesian games with continuous type and action space.

Second, we provide empirical results of equilibrium computation on bilateral trade and explore equilibria with different prior distributions, different levels of risk-aversion, or different numbers of buyers and sellers and their impact on overall efficiency. So far, no explicit equilibrium bid functions have been known for these environments. In the standard environment with uniform priors for which explicit equilibrium bid functions are known, we reliably find the linear equilibrium with NPGA. Interestingly, with SODA, we find step-function equilibria. This has to do with NPGA only being able to learn continuous equilibrium bid functions. In contrast, the discretization of the type and action space allows SODA also to learn non-differentiable equilibrium bid functions. The multitude of equilibria differs from many single-sided auction models, and it is surprising that equilibrium learning algorithms find one of these equilibria consistently. They do not cycle or end up in disequilibrium with a high utility loss. This way, we push the boundaries of equilibrium analysis to the challenging case of bilateral trade with a continuum of equilibria.

The remainder of this article is structured as follows. The following section will discuss literature on bilateral trade and equilibrium learning. Section 3 introduces the economic model as Bayesian games, whereas in Section 4 the two learning methods will be introduced. Section 5 provides our numerical results before we conclude in Section 6.

2. Related literature

In what follows, we introduce additional related literature on bilateral trade and equilibrium learning.

2.1. Bilateral trade

The famous theorem by Myerson & Satterthwaite (1983) states that in the simple bilateral trade environment, for a single good between one buyer and one seller, no mechanism can be individually rational, budget balanced, incentive-compatible, and efficient. The impossibility result spawned substantial research on bilateral trade. A number of different mechanisms for double auctions with multiple buyers and sellers have been proposed in Gresik & Satterthwaite (1989); McAfee (1992), or Williams (1999). The k-double auction is probably the most popular one as it is deterministic and budget-balanced and, as such, resembles real-world practices. Already Chatterjee & Samuelson (1983) examined BNE and showed that double auctions are asymptotically efficient as the agents become strongly risk-averse. Leininger et al. (1989) analyzed the case of identically distributed costs and benefits of the participants. With a uniform distribution, the sealed-bid game has a continuum of equilibria. Obviously, such equilibrium predictions are weak. One family of equilibria consists of differentiable strategies (including a linear BNE). Another family is composed of step-functions with arbitrarily many jumps. With general independent distributions of benefits and costs the, authors find similar families of equilibria. Radner & Schotter (1989) experimentally analyze the properties of the average mechanism and find linear equilibrium strategies also in the lab. Furthermore, Satterthwaite & Williams (1989) model the environment as a Bayesian game and prove the existence of a multiplicity of equilibria. Their paper focuses on differentiable equilibrium strategies.

Leininger et al. (1989) provide closed-form equilibrium strategies for quasi-linear traders and uniformly distributed priors. For general independent prior distributions, they only show the existence of equilibria. A number of articles analyze different effects on market efficiency under this mechanism. The inefficiency in a k-double auction decreases for increasingly risk-averse agents (Chatterjee & Samuelson, 1983). Additionally, Satterthwaite & Williams (2002) show that the k-double auction reduces the worst-case inefficiency at the fastest possible rate among all interim individually rational and budget-balanced mechanisms. More recent work goes beyond the independent private values model (Kadan, 2007; Satterthwaite et al., 2022), and it explores posted-price (Blumrosen & Dobzinski, 2021) or randomized mechanisms (Garratt & Pycia, 2020).

Overall, this stream of literature spans almost forty years by now, but explicit equilibrium bid functions are unknown except for specific models with uniform distributions, quasi-linear utility functions, and independent private values. Numerical methods that allow us to derive equilibrium predictions for specific models with non-uniform, possibly asymmetric or interdependent, priors or risk-averse traders in minutes rather than years could push the boundaries of equilibrium analysis for bilateral trade with two traders also for larger environments.

2.2. Equilibrium learning algorithms

Let us also discuss related literature on equilibrium learning. As indicated earlier, most of this literature deals with finite games (Fudenberg & Levine, 2009). *Gradient dynamics in games* have been studied in evolutionary game theory and multi-agent learning. While earlier work considered mixed strategies over normal-form games (Bowling, 2005; Bowling & Veloso, 2002; Busoniu et al., 2008; Zinkevich, 2003), more recently, motivated by the emergence of GANs, there has been a focus on (complete-information) continuous games (Bailey & Piliouras, 2018; Balduzzi et al., 2018; Letcher et al., 2019; Mertikopoulos & Zhou, 2019; Schaefer & Anandkumar, 2019). A common result for many settings and algorithms is that gradient-based learning rules do not necessarily converge to Nash equilibria and may exhibit cycling behavior but often achieve no-regret properties and thus converge to weaker Coarse Correlated equilibria (CCE). An analogous result exists for finite-type Bayesian games, where no-regret learners are guaranteed to converge to a Bayesian CCE (Hartline et al., 2015).

Earlier approaches on finding equilibria in auctions were usually setting specific and relied on reformulating the BNE first-order condition of Eq. (9) as a differential equation and then solving this equation analytically (where possible) (Ausubel & Baranov, 2020; Krishna, 2009; Vickrey, 1961). Armantier et al. (2008) introduced a BNE-computation method based on expressing the Bayesian game as the limit of a sequence of complete-information games. They show that the sequence of Nash equilibria in the restricted games converges to a BNE of the original game. While this result holds for any Bayesian game, setting-specific information is required to generate and solve the restricted games. Rabinovich et al. (2013) study best-response dynamics on mixed strategies in auctions with finite action spaces. These articles were focused on single-object auctions. Bosshard et al. (2017, 2020) were the first to compute equilibria for combinatorial auctions. The method explicitly computes point-wise best responses in a fine-grained discretization of the strategy space via sophisticated Monte–Carlo integration.

We focus on NPGA (Bichler et al., 2021) and SODA (Fichtl et al., 2022). These two recent contributions have shown to be very versatile and allowed for the computation of BNE in a large variety of different (single-sided) auction models. Moreover, in contrast to earlier work, both techniques implement gradient dynamics compared to the best-response algorithms mentioned above. They compute approximate equilibria in minutes for standard auction models from the literature. A more detailed explanation will be provided in Section 4.

3. Economic model

We first introduce notation and equilibrium solution concepts used in our analysis. Next, we discuss the k-double auction and equilibrium bidding strategies.

3.1. Preliminaries

A simple two-sided exchange market with unit demand can be modeled as a Bayesian game $\mathcal{G} = (\mathcal{I}, \mathcal{A}, \mathcal{V}, u, F)$. The agents $\mathcal{I}$ consist of n_B buyers and n_S sellers. Each buyer wants to buy one item and each seller wants to sell one item. The action space $\mathcal{A} = \mathcal{A}_1 \times \ldots \times \mathcal{A}_{n_B} \times \mathcal{A}_{n_B+1} \times \ldots \times \mathcal{A}_{n_B+n_S}$ represents the possible bids that buyers and sellers can submit. A buyer's bid denotes the amount he is willing to pay, whereas a seller's bid denotes how much she wants to receive when selling her good. The agents' type space $\mathcal{V} = \mathcal{V}_1 \times \ldots \times \mathcal{V}_{n_B+n_S}$ denotes their possible values for the good. That is, $v_i \in \mathcal{V}_i$ denotes the value agent i places on the good. For a buyer, that is the maximum value he is still willing to pay. For a seller, it might denote the cost that she invested and is the minimum amount she wants to receive when selling the good. We assume the type and action spaces to be non-negative $\mathcal{A}_i = \mathcal{V}_i = \mathbb{R}_0^+$. The joint probability density function $f : \mathcal{V} \to \mathbb{R}_0^+$ describes a prior distribution over the agents' types and is assumed to be common knowledge. The marginal distributions are denoted by f_i, and F_i denotes the associated cumulative distribution function. The vector $u = (u_1, \ldots, u_{n_B+n_S})$ of f-integrable, individual (*ex-post*) utility functions $u_i : \mathcal{V}_i \times \mathcal{A} \to \mathbb{R}$ assigns the game outcome for each possible action and valuation profile. In the game's *interim* stage, an agent knows its valuation but not those of the others, whereas, in the *ex-ante* stage, each agent only knows about the prior distribution f.

In the *ex-ante* stage of the game, each agent is tasked with finding a strategy β_i that maps from each type to an action, i.e., $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$. The strategy profile is denoted by $\beta = (\beta_1, \ldots, \beta_{n_B+n_S}) = (\beta_i, \beta_{-i})$ for every i. An index $-i$ denotes a partial profile for all agents but agent i. We denote the ex-ante action space of agent i

by $\Sigma_i \equiv \mathcal{A}_i^{\mathcal{V}_i}$ and the joint ex-ante action space by $\Sigma \equiv \prod_i \Sigma_i$. Note that the spaces Σ_i are, in general, infinite-dimensional. The equilibrium learning algorithms described in Sections 4.1 and 4.2 transform the infinite-dimensional game with Σ into one with finite-dimensional strategies while maintaining sufficient expressiveness to approximate arbitrary equilibrium strategies.

Fixing a strategy profile β, we can formulate utilities for the game's interim and ex-ante stages. Agent i's interim utility is defined as

$$u_i^{\text{interim}}(v_i, \beta_i(v_i), \beta_{-i}) = \mathbb{E}_{v_{-i}|v_i}[u_i(v_i, \beta_i(v_i), \beta_{-i}(v_{-i}))]. \tag{1}$$

Extending this to the ex-ante stage gives the ex-ante utility of agent i by

$$u_i^{\text{ante}}(\beta_i, \beta_{-i}) = \mathbb{E}_{v_i}\left[u_i^{\text{interim}}(v_i, \beta_i(v_i), \beta_{-i})\right]. \tag{2}$$

An ϵ-Bayes–Nash equilibrium (ϵ-BNE) is given by a strategy profile β^*, such that no agent can increase its utility by more than $\epsilon \geq 0$ by unilaterally deviating from it. That is,

$$u_i^{\text{ante}}(\beta_i, \beta_{-i}^*) - u_i^{\text{ante}}(\beta_i^*, \beta_{-i}^*) \leq \epsilon \quad \text{for all } \beta_i \in \Sigma_i \text{ and } i \in I. \tag{3}$$

The case of $\epsilon = 0$ corresponds to a Bayes–Nash equilibrium (BNE).

The interim stage formulates the individual agent's task when the valuation is already known, reducing the complexity of the strategy space to a single action. In contrast, the ex-ante stage captures the full complexity of the given strategic interaction, which is, e.g., needed to analyze the algorithms' convergence properties (see Section 4).

The game outcomes, i.e., the goods' allocation and the respective prices the buyers need to pay and payments the sellers receive, are determined by a market *mechanism*. The mechanism collects the bids $b \in \mathcal{A}$ of buyers and sellers and outputs an allocation vector $x(b) \in \{0,1\}^{n_B+n_S}$ and a payment vector $p(b) \in \mathbb{R}^{n_B+n_S}$. It holds that a buyer $i \in \{1, \ldots, n_B\}$ gets an item if and only if $x_i(b) = 1$. A seller $j \in \{n_B+1, \ldots, n_B+n_S\}$ sells her item if and only if $x_j(b) = 1$. Tie-breaking rules may be encoded into the allocations x. Agent i's payment satisfies $p_i(b) = 0$ if $x_i(b) = 0$. The baseline utility function is that of a risk-neutral agent with quasi-linear utility. The quasi-linear ex-post utilities for the buyers are given by

$$u_i^{QL}(v_i, b) = \begin{cases} x_i(b) \cdot v_i - p_i(b) & \text{for } i \in \{1, \ldots, n_B\}, \\ 0 & \text{else.} \end{cases} \tag{4}$$

The sellers' ex-post utilities are respectively

$$u_j^{QL}(v_j, b) = \begin{cases} p_j(b) - x_j(b) \cdot v_j & \text{for } j \in \{n_B+1, \ldots, n_B+n_S\}, \\ 0 & \text{else.} \end{cases} \tag{5}$$

We extend this by including risk-aversion into our setting, arguably one of the most studied behavioral effects in single- and double-sided markets. We model this via utilities $u_i^{\text{RA}} = (u_i^{\text{QL}})^\rho$ where $\rho \in (0,1]$ denotes the risk-attitude. The case of $\rho = 1$ corresponds to the risk-neutral traders with quasi-linear utilities. If not stated otherwise, we assume risk-neutral bidders.

3.2. K-double auction

We focus on the *k-double auction*, because, as discussed, it is relevant, strategically complex, and some BNE strategies are known for non-trivial settings.[1] Special cases are the average double auction with $k = 0.5$, the buyer's bid double auction with $k = 1$, and the seller's bid double auction with $k = 0$. Sellers and buyers simultaneously submit asks and bids for one unit each. After collecting the bids $b = (b_1, \ldots, b_{n_B+n_S})$, the mechanism sorts them according to a natural ordering, i.e.,

$$b_1 \geq b_2 \geq \ldots \geq b_{n_B} \quad \text{and} \quad b_{n_B+1} \leq b_{n_B+2} \leq \ldots \leq b_{n_B+n_S}, \tag{6}$$

to form supply and demand curves. The buyers' bids are sorted to be decreasing, whereas the sellers' bids are ordered so that they are increasing. One then determines the break-even index ℓ such that ℓ is the largest index satisfying $b_\ell \geq b_{n_B+\ell}$ and $b_{\ell+1} < b_{n_B+\ell+1}$. This corresponds to the crossing of the supply and demand curves. In the case of ties, a lottery decides the ordering and break-even index. The index ℓ determines the allocations. The first ℓ sellers with the lowest asks pass their goods to the first ℓ buyers with the highest bids, i.e., $x_i(b) = 1$ for $i \leq \ell$ and $n_B + 1 \leq i \leq n_B + \ell$ and 0 otherwise. The market-clearing trade price is derived from b_ℓ and $b_{n_B+\ell}$ and fixed at $P_i(b) = kb_\ell + (1-k)b_{n_B+\ell}$ for agents that trade, $i \leq \ell$ and $n_B + 1 \leq i \leq n_B + \ell$, and 0 otherwise. Unlike in some other mechanisms like the famous VCG auction, having this constant market-clearing price ensures budget balance by definition.

3.3. Equilibrium analysis

This subsection focuses on the bilateral bargaining setting with two traders for the k-double auction mechanism. For this case, we present different classes of equilibrium strategies. However, we start by deriving the first-order conditions for continuous bidding functions, that play a central role in deriving equilibria, as well as for a convergence analysis of NPGA in Section 4.1. We simplify the notation for the case of bilateral bargaining, i.e., a two-sided market with exactly one buyer and seller so that the buyer's variables are indexed by B, and the seller's by S, e.g., the buyer's valuation is denoted by v_B and the seller's by v_S. Let us first introduce some assumptions.

Assumption 1. Let the priors be defined on bounded intervals $\Omega_B = [\underline{v_B}, \overline{v_B}]$ and $\Omega_S = [\underline{v_S}, \overline{v_S}] \subset \mathbb{R}$.[2] We assume that the strategies $\beta_B : \Omega_B \to [\underline{b_B}, \overline{b_B}] =: \hat{\Omega}_B$ and $\beta_S : \Omega_S \to [\underline{b_S}, \overline{b_S}] =: \hat{\Omega}_S$ of buyer and seller respectively, satisfy the following:

1. β_B and β_S are strictly increasing,
2. β_B, β_B^{-1}, β_S and β_S^{-1} are Lipschitz continuous.

These assumptions do not constitute strong restrictions for the setting. It is common to consider strictly increasing bid functions and some additional regularity to derive the first-order conditions (Chatterjee & Samuelson, 1983; Leininger et al., 1989). Independently, they will allow us to prove our convergence result (Proposition 1), which describes a first set of ex-ante criteria for which NPGA finds an equilibrium. Property 1 will be relaxed at other occasions. Here, together with property 2, it ensures that there exist inverse functions β_B^{-1} and β_S^{-1}. Assuming independent prior distributions, the interim utilities of the buyer and seller can now be derived and are given by

$$\begin{aligned} &u_B^{\text{interim}}(v_B, \beta_B(v_B), \beta_S) \\ &= \mathbb{1}_{\{\beta_B(v_B) \geq \underline{b_S}\}} \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} (v_B - P(\beta_B(v_B), y)) f_S(\beta_S^{-1}(y))(\beta_S^{-1})'(y)dy \end{aligned} \tag{7}$$

[1] Other common mechanisms for two-sided markets are not as strategically complex Blumrosen & Dobzinski (2021); Hagerty & Rogerson (1987) or applicable McAfee (1992).

[2] Note that allowing unbounded intervals for the prior distributions leads to an additional (but well-behaved) error term for the seller's interim utility. Therefore, we omit this special case for clarity.

and

$$u_S^{\text{interim}}(v_S, \beta_B, \beta_S(v_S)) = \mathbb{1}_{\{\overline{b_B} \geq \beta_S(v_S)\}} \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} (P(x, \beta_S(v_S)) - v_S) f_B(\beta_B^{-1}(x))(\beta_B^{-1})'(x) dx, \tag{8}$$

where $\mathbb{1}$ denotes the indicator function of wether or not trade takes place. The detailed derivations can be found in Appendix C. The first-order conditions to optimize the interim utilities can now be summarized in the following system of non-linear ordinary differential equations (ODE):

$$A(v_B, v_S, \beta_B, \beta_S) := \begin{pmatrix} \frac{d}{d\beta_B(v_B)} u_B^{\text{interim}}(v_B, \beta_B(v_B), \beta_S) \\ \frac{d}{d\beta_S(v_S)} u_S^{\text{interim}}(v_S, \beta_B, \beta_S(v_S)) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{9}$$

How to solve such systems to determine strategies β_B and β_S, which are non-trivial (i.e., such that trade occurs over a set of non-zero measure) is an open problem. In general, there is no principled method to derive closed-form solutions for systems of non-linear ODEs, and also numerical techniques turned out challenging.

A few articles discuss the related equilibrium problem in the asymmetric independent private values model of one-sided auctions, which also results in a system of non-linear ODEs (Hubbard & Paarsch, 2014). Because the Lipschitz condition is not satisfied for the system, much of the theory concerning systems of ODEs no longer applies and numerical methods for differential equations such as the class of Runge–Kutta methods (Butcher, 2008) have been explored. Fibich & Gavish (2011) discuss the inherent numerical instability of such shooting methods. Importantly, the derived solutions might not constitute inverses of valid bidding strategies. That is due to the solution's dependence on the initial value and boundary conditions, which do not guarantee that Assumption 1 holds for the derived strategies. Additionally, the system's complexity increases tremendously with more types of bidders or by allowing interdependent prior distributions, which holds true for asymmetric auctions and bilateral trade. For general interdependent priors, an agent i needs access to the conditional distribution $F_{v_{-i}|v_i}$ to find its optimal action. Thus, one cannot even state the ODEs because they require explicit knowledge of the conditional distributions for which there is no general analytical framework (Hormann, 2013). Moreover, such numerical techniques to solve asymmetric independent private values auctions lack convergence guarantees (Hubbard & Paarsch, 2014).

Only when making further assumptions on the system of ODEs, such as a specific payment rule and prior, can one derive analytical solutions by finding the inverse bid functions for well-chosen initial values and then using the implicit function theorem to find the optimal bid function. Linear equilibrium bid strategies satisfy Eq. (9) in a model with independent uniform priors under the k-DA pricing rule (see Satterthwaite & Williams, 1989):

$$\beta_B(v_B; k) = \begin{cases} \frac{1}{1+k} v_B + \frac{k(1-k)}{2(1+k)}, & \text{if } v_B \in \left[\frac{1-k}{2}, 1\right], \\ h_B(v_B), & \text{else,} \end{cases} \tag{10}$$

$$\beta_S(v_S; k) = \begin{cases} \frac{1}{2-k} v_S + \frac{1-k}{2}, & \text{if } v_S \in \left[0, \frac{2-k}{2}\right], \\ h_S(v_S), & \text{else.} \end{cases} \tag{11}$$

The functions h_B and h_S can be arbitrary as long as they do not lead to more trade, i.e., $h_B < \frac{1-k}{2}$ and $h_S > \frac{2-k}{2}$. We refer to the whole class and any strategy from this class of equilibrium strategies as *linear equilibrium*. The linear equilibrium is of special interest as it has the highest expected gains from trade of any equilibrium (Myerson & Satterthwaite, 1983).

For the special case of the average double auction ($k = 0.5$) with uniform distributions, one can derive a broader continuum of equilibrium strategies (see Chatterjee & Samuelson, 1983; Leininger et al., 1989). For example, if we set h_B and h_S to be the continuation of the corresponding linear functions in the linear equilibrium, one obtains an equilibrium strategy that belongs to the class of *symmetric equilibria*. This class has been derived by using the symmetry condition

$$\beta_B(v_B) = 1 - \beta_S(1 - v_B), \tag{12}$$

which means that the curve of β_S is obtained from β_B by a rotation of π. In a symmetric equilibrium, the buyer underbids, when his valuation is v_B, by the same amount that the seller overbids when her valuation is $v_S = 1 - v_B$. It turns out that a symmetric equilibrium is uniquely determined by choosing a value $g_{\text{sym}} \in (0, 1/2)$ at the symmetry point $1/2$, which constitutes a unique equilibrium strategy for each value of g_{sym}. See Fig. 1(a) for some exemplary strategies from this class. The linear equilibrium is attained for $g_{\text{sym}} = 3/8$ and is the only value where a closed-form solution is known (Leininger et al., 1989). This class of equilibria has several notable properties. It consists of infinitely many different equilibria and the efficiency obtained in equilibrium, and the resulting gains from trade range from zero to second-best.

The third class of equilibria consists of strategies where bidders only submit a *finite* number of different bids. That means buyer and seller may post identical bids for different valuations.[3] This class has particular relevance for real-world situations where it is usually required to submit bids in, e.g., full dollars. We denote this set as the class of *step function equilibria*. Leininger et al. (1989) provide properties and explicit equilibria for the case of the average mechanism and the case of buyer and seller using strategies with an equal amount of steps. They show that all step function equilibria with exactly n steps are of the following form:

$$\beta_S(v_S) = \begin{cases} a_1, & 0 \leqslant v_S \leqslant x_1, \\ a_2, & x_1 < v_S \leqslant x_2, \\ \vdots & \\ a_n, & x_{n-1} < v_S \leqslant x_n, \\ 1, & x_n < v_S \leqslant 1, \end{cases} \qquad \beta_B(v_B) = \begin{cases} 0, & 0 \leqslant v_B < z_1, \\ a_1, & z_1 \leqslant v_B < z_2, \\ a_2, & z_2 \leqslant v_B < z_3, \\ \vdots & \\ a_n, & z_n \leqslant v_B \leqslant 1, \end{cases} \tag{13}$$

where

$$0 < a_1 < a_2 < \ldots < a_n < 1,$$

$$z_1 = a_1, \qquad z_i = a_i + \frac{x_{i-1}}{(x_i - x_{i-1})} \frac{(a_i - a_{i-1})}{2} \text{ for } i = 2, \ldots, n,$$

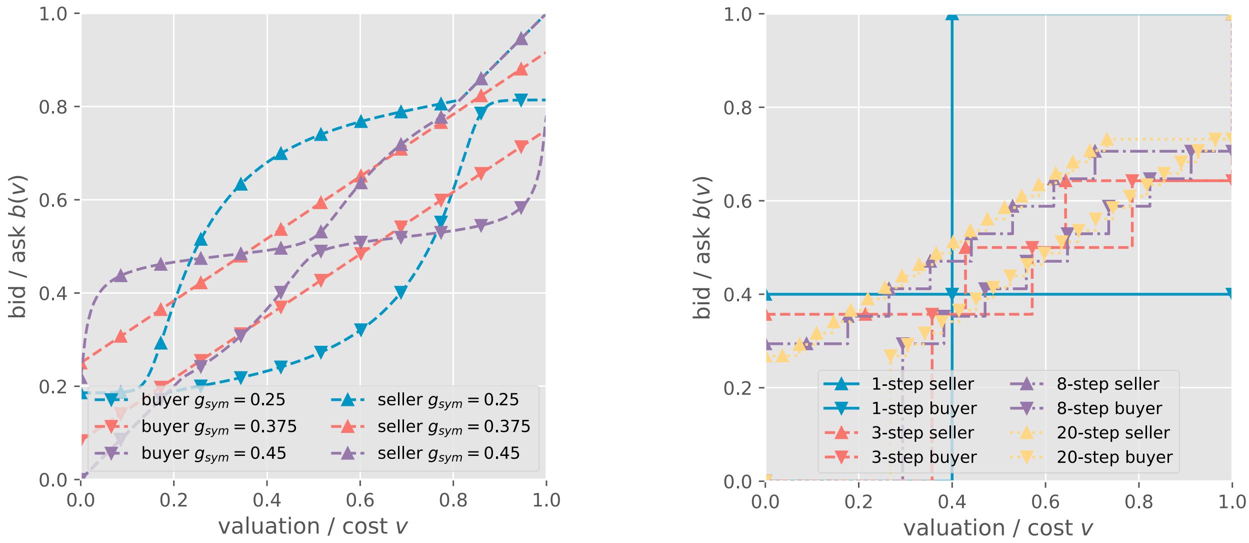
$$x_i = a_i - \frac{(1 - z_{i+1})}{(z_{i+1} - z_i)} \frac{(a_{i+1} - a_i)}{2} \text{ for } i = 1, \ldots, n-1, \qquad x_n = a_n.$$

Note that this is only a necessary condition and does not guarantee functions of the form of Eq. (13) to be an equilibrium for all $a \in [0, 1]^n$ such that $0 < a_1 < \ldots < a_n < 1$. We denote this subset of step function equilibria as the class of n-step equilibria. Some of their notable properties are that,

1. the buyer's lowest bid has to be zero, whereas the seller's highest bid has to be one,
2. every non-marginal bid (non-zero for the buyer and unequal one for the seller) of one bidder lies in the set of potential bids of the other,
3. the supports of non-marginal bids for both bidders coincide.

Furthermore, Leininger et al. (1989) provide several explicit examples of n-step equilibria that in part constitute continua of equilibria on their own. Fig. 1(b) shows some strategies for a different number of steps. However, these are not determined by the number of steps alone. For example, for a single step, $a_1 = a$ (see Eq. (13)) constitutes an equilibrium for any $a \in (0, 1)$. For more details on this class of equilibria, we refer to Leininger et al. (1989).

[3] Note that these equilibrium strategies do not satisfy Assumption 1.



(a) Equilibrium strategies from the class of symmetric equilibria for different values of g_{sym}. Including the special case of linear equilibrium strategies.

(b) Equilibrium strategies from the class of n-step equilibria for a different number of steps.

Fig. 1. Exemplary equilibrium strategies for symmetric and step function equilibria classes.

Another important property of every n-step equilibrium is its robustness to small perturbations (see Proposition 3.6 in their work), which indicates that these equilibria are likely to be attracting under local search algorithms. Even though their results only regard n-step equilibria, we observe similar properties for general step function equilibria in our experiments in Section 5.3.

For the special case of an average mechanism, Chatterjee & Samuelson (1983) also derived another linear BNE under risk-averse traders. With a risk parameter of ρ, the equilibrium profile is given as:

$$\beta_B(v_B) = \left(\frac{1-\frac{1}{2c}}{4c^2-1}\right) + \left(1-\frac{1}{2c}\right)v_B, \tag{14}$$

$$\beta_S(v_S) = \left(\frac{c-\frac{1}{2}}{2c^2-\frac{1}{2}}\right) + \left(1-\frac{1}{2c}\right)v_S, \tag{15}$$

for $c = 2^{1/\rho} - \frac{1}{2}$. This also covers the special case of risk-neutral traders in the linear BNE from Eq. (10). Intuitively, the higher the risk aversion, the lower the marginal utility of misreporting one's valuation compared to the possible loss under no trade. This leads to risk-averse traders asymptotically biding truthfully for increasing risk aversion.

So, given these different assumptions on the market and possibly multiple classes of equilibria, bidders face a substantial coordination problem. Moreover, it is unclear which equilibria will be found by equilibrium learning algorithms or if such algorithms even find an equilibrium.

3.4. Expected utility with linear strategies

The analysis of gradient dynamics and the types of equilibria emerging in a game requires a thorough understanding of the participants' utility functions. For example, Rosen (1965) showed that games admit a unique Nash equilibrium when the participants' utility functions satisfy the strict monotonicity. More recently, Mertikopoulos & Zhou (2019) showed conditions of the utility functions for which no-regret learning algorithms result in a Nash equilibrium if they converge to a pure equilibrium. In the following, we derive an a priori result for local convergence by drawing on a recent result by Chasnov et al. (2020). Among other things, one needs to show specific properties, such as Lipschitz continuity, of the ex-ante utility functions and negative definiteness of the game Jacobian in equilibrium. However, without knowing the parametric form of the bid function, it is impossible to study the properties of the expected utility functions. Therefore, we now choose a specific parametrization of linear bid functions, allowing us to derive an analytical equilibrium. The procedure works for other parametrizations as well. We provide results for this highly restricted setting which already turns out to be difficult, thereby illustrating how limited the current approach in solving the resulting system of differential equations is.

We focus on bilateral bargaining with one buyer and one seller, independent and uniform prior distributions $F_B(x) = F_S(x)$ on $[0, 1]$, and assume linear strategies, which are known to include a BNE in the unrestricted game, as we have seen in the previous subsection. This means, there exist $m_B, m_S, t_B, t_S \in \mathbb{R}$ such that the strategies are given by

$$\beta_B(v_B) = m_B v_B + t_B, \qquad \beta_S(v_S) = m_S v_S + t_S. \tag{16}$$

Based on Assumption 1, we can define the feasible set for all possible linear strategies for this setting.

1. $m_B, m_S > 0$;
2. $\Omega_B = \Omega_S = [0, 1]$ and $\hat{\Omega}_B = [t_B, m_B + t_B]$, $\hat{\Omega}_S = [t_S, m_S + t_S]$;
3. $\beta_B^{-1}(y) = \frac{1}{m_B} \cdot (y - t_B)$;
4. $\beta_S^{-1}(y) = \frac{1}{m_S} \cdot (y - t_S)$.

Besides, we need to make the following assumption to restrict the slope of the linear strategies so that they cannot be arbitrarily flat, ensure that the intersects t_B and t_S are bounded, and restrict ourselves to situations where demand is not strictly exceeding supply.

Assumption 2. In the restricted setting of linear strategies, we make the following additional assumptions:

1. There exists an $\epsilon_0 > 0$ such that $m_B, m_S \geq \epsilon_0 > 0$,
2. there exists a $K > 0$ such that $|t_B|, |t_S| \leq K < \infty$,

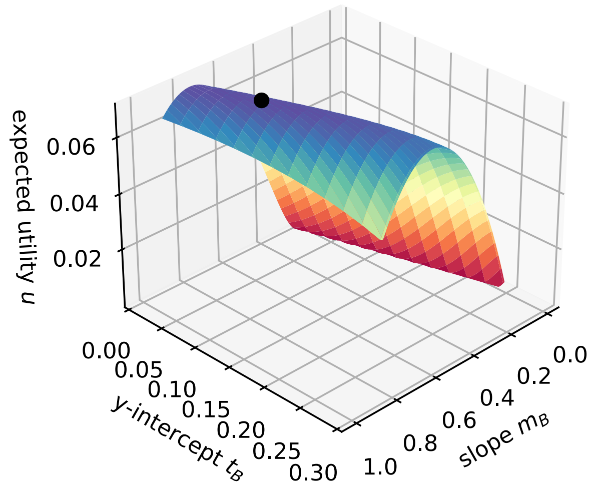


Fig. 2. Ex-ante utility of the buyer under an opposing seller that plays according to the linear BNE strategy. The maximal utility on the feasible action set (black dot) is achieved by also playing the BNE strategy $\beta_B(v) = \frac{2}{3}v + \frac{1}{12}$. Note that the points are restricted to the feasible set according to Assumption 1.

3. $m_B x + t_B \leq m_S + t_S$ for all $x \in [\underline{v_B}, \overline{v_B}]$, i.e., the highest ask price of the seller is at least as high as any bid of the buyer,
4. $m_S y + t_S \geq t_B$ for all $y \in [\underline{v_S}, \overline{v_S}]$, i.e., the lowest bid price of the buyer is less or equal to any ask of the seller.

The first two properties guarantee Lipschitz continuity of the ex-ante utilities later on. Properties three and four considerably simplify calculations by restricting the setting to competitive market scenarios. Note that this simplification is not restrictive, in the sense that the resulting feasible set includes the equilibrium. We can now derive the ex-ante utility of the buyer and seller for the general k-double auction (see Appendix E for details):

$$u_B^{\text{ante}}(m_B, t_B, m_S, t_S, k) \tag{17}$$
$$= -\frac{1}{6{m_B}^2 m_S}(m_B + t_B - t_S)^2$$
$$\cdot (t_B - t_S + m_B(m_B + t_B + 2t_S - 2) + m_B k(m_B + t_B - t_S)).$$

Similarly, the seller's ex-ante utility is

$$u_S^{\text{ante}}(m_B, t_B, m_S, t_S, k) \tag{18}$$
$$= -\frac{1}{6 m_B {m_S}^2}(m_B + t_B - t_S)^3$$
$$+ \frac{1}{6 m_B {m_S}^2} m_S (m_B + t_B - t_S)^2 (m_B + t_B + 2t_S + k m_B + k t_B - k t_S).$$

Fig. 2 depicts the utility landscape (based on m_B and t_B) from the buyer's perspective when faced with a seller playing the linear BNE in the average double auction. This resulting utility surface is concave in large parts, which gives some rationale why gradient-based learners converge in this environment. Following Rosen (1965), we demonstrate that global monotonicity of the game is not satisfied (see Appendix F). Even in this restricted game with linear strategies, the game is only locally monotone, e.g., in a neighborhood of the equilibrium. This is a strong indication that global monotonicity is not satisfied for more complex parametrizations as well.

Additional visualizations of the expected utility landscape assuming arbitrary linear, concave, or convex bid functions can be found in Appendix D. These figures plot utility as a function of value and bid submitted. Interestingly, all of them are concave in large regions, as well.

4. Learning algorithms

Let us briefly introduce NPGA and SODA, the two learning algorithms we will use in our numerical experiments, and discuss important properties. On a high level, both methods rely on an approximation of the original problem. NPGA uses neural networks to approximate pure strategies with a finite-dimensional parameter space and learns Bayes–Nash equilibria through self-play. Individual agents submit bids, observe the ex-post utility of their bids in a large batch of auctions, and then go a step in the direction of their utility gradient. The fact that the ex-post utility is discontinuous describes a key technical challenge, which is solved using smoothing techniques. In contrast, SODA solves a discretized version of the game with discrete type and action space. While this leads to an additional error term in the original game, the utility gradient is available exactly and does not need to be estimated from the smoothed utility function. The method uses the dual averaging method and learns distributional strategies, an extension of mixed strategies for Bayesian games. We also know that if SODA converges, then it has to converge to an equilibrium. While SODA is very fast for small environments with only a few participants and strategies, it suffers from a curse of dimensionality for larger markets with many players and strategies. Let us now introduce these algorithms in more detail.

4.1. NPGA

NPGA follows the gradient dynamics of a game via simultaneous gradient ascent of all bidders. Conceptually, players observe a gradient-oracle $\nabla_{\beta_i} u_i^{\text{ante}}(\beta_i, \beta_{-i})$ with respect to the current strategy profile β^t in each iteration. Then the rule proposes that players perform a gradient update:

$$\beta_i^t \equiv \beta_i^{t-1} + \Delta_i^t \quad \text{with} \quad \Delta_i^t \propto \nabla_{\beta_i} u_i^{\text{ante}}(\beta_i, \beta_{-i}), \tag{19}$$

Note that in this high-level description, we refer to the gradient dynamics of the ex-ante utility u^{ante}. Consequently, $\beta_i \in \Sigma_i$ are functions in an infinite-dimensional function space, so the gradient $\nabla_{\beta_i} u_i^{\text{ante}}$ is itself a *functional* derivative such as a Gateaux derivative[4] over the Hilbert space Σ_i. To compute the gradient estimate in practice, NPGA represents each bidder's strategy by a neural network $\beta_i(v_i) \equiv \pi_i(v_i; \theta_i)$ and a corresponding parameter vector $\theta_i \in \mathbb{R}^{d_i}$. $d_i \in \mathbb{N}$ is finite and we thus transform the problem of choosing an infinite-dimensional strategy into choosing a finite-dimensional parameter vector θ_i.

Due to the discrete nature of the allocations x, the ex-post utilities $u_i(v_i, b_i, b_{-i})$ are usually discontinuous, and thus the gradient provides wrong signals. Therefore, NPGA estimates the gradient using evolutionary strategies (ES) as it was used by Salimans et al. (2017). To calculate $\nabla_\theta u^{\text{ante}}$, we perturb the parameter vector P times, $\theta_{i;p} \equiv \theta_i + \varepsilon_p$, using zero-mean Gaussian noise $\varepsilon_p \sim \mathcal{N}(0, \sigma^2)$ for $p \in \{1, \ldots, P\}$, where P and σ are hyperparameters. NPGA then calculates each perturbation's fitness, $\varphi_p \equiv u_i^{\text{ante}}(\pi_i(v_i; \theta_{i;p}), \beta_{-i})$, via Monte–Carlo integration, and estimates the gradients as the fitness-weighted perturbation noise $\nabla_\theta^{ES} \equiv \frac{1}{\sigma^2 P} \sum_p \varphi_p \varepsilon_p$. The technique gives an asymptotically unbiased estimator of $\nabla_\theta u^{\text{ante}}$. The pseudo-code of NPGA is given in Algorithm 1. Note that the original paper by Bichler et al. (2021) focuses on symmetric auctions, where all bidder valuations are drawn from the same prior distribution, and all bidders share the same equilibrium bid function. Therefore, only a single neural network needs to be trained in such one-sided auctions. The bilateral bargaining model that we analyze in this paper is inherently

[4] Gateaux derivatives are a generalization of directional derivatives in Euclidean spaces to Banach spaces (of which Hilbert spaces are a subset of).

Algorithm 1: Neural pseudogradient ascent using evolutionary strategies.

```
Input: Initial policy, ES population size $P$, ES noise variance,
       learning rate, batch size
for $t = 1, 2, \ldots$ do
    Sample a batch of valuation profiles;
    Calculate joint utility of current strategy profile;
    for each agent $i \in \mathcal{I}$ do
        for each $p \in \{1, \ldots, P\}$ do
            Perturb agent $i$'s current policy;
            Evaluate fitness of perturbation $p$ by playing
            against current opponents;
        end
        Calculate ES pseudogradient as fitness-weighted
        perturbation noise;
        Perform a gradient ascent update step on the current
        policy;
    end
end
```

asymmetric, and we train two neural networks, one for the buyer and one for the seller. In larger environments with more participants, symmetry among some or all of the bidders on one side is a widespread assumption. Therefore, we only need to train a single neural network for bidders in a symmetry class, which makes the implementation of larger markets much more efficient.

Given a vectorized implementation of the joint ex-post utility u, estimating u^{ante} via Monte–Carlo integration over $\mathcal{V}$ is feasible due to parallel execution on hardware accelerators such as GPUs. In our experiments, we use custom vectorized implementations of the double auction mechanisms considered using the PyTorch framework (Paszke et al., 2017). This effectively allows us to simulate hundreds of thousands of games in parallel to get more precise approximations for the gradients and utilities on consumer-level hardware.

The action space is usually restricted, e.g., for auctions, the bids and asks must be non-negative. This can be achieved, e.g., by equipping the neural networks' last layer with a ReLU activation function so that negative values are mapped to be zero. If not stated otherwise, we pretrain the neural networks for 500 iterations to submit truthful bids, similar to the original paper by Bichler et al. (2021). This makes the experiments easier to compare, prevents numerical instabilities (see Section 5.2 for details) and prevents the so-called dead-ReLU problem.

It is interesting to understand when NPGA converges to an equilibrium. Unfortunately, the analysis of gradient dynamics, in general, can be arbitrarily complex (Andrade et al., 2021). Learning dynamics do not generally obtain a Nash equilibrium (Benaim & Hirsch, 1999). A number of recent results on matrix games showed that gradient dynamics may circle, diverge, or are even chaotic (Sanders et al., 2018). However, for bilateral bargaining with uniform priors, we can show that the linear equilibrium is locally attracting for NPGA in the space of linear strategies. That means, if one initializes the algorithm close enough, it is ensured to converge to the equilibrium. In other words, assuming that NPGA receives exact gradient feedback, the learning rate is small enough, and the starting point is in the region of attraction, NPGA converges to the linear BNE strategy:

Proposition 1. *Consider the bilateral bargaining model with two quasi-linear traders and independent uniform priors under the average double auction ($k = 1/2$) satisfying Assumptions 1 and 2. Suppose agents learn with NPGA under exact gradient feedback, neural networks consisting of a single neuron, and a learning rate s.t. $0 < \gamma < \tilde{\gamma}$, where $\tilde{\gamma} = \arg\min_{h>0} \max_j |1 - h\lambda_j(J(\theta^*))| = 1$ and $\lambda_j(J(\theta^*))$ denotes the j'th eigenvalue of the game Jacobian $J(\theta^*)$. Then, NPGA converges to the linear BNE from Eq. (10) when initialized in the region of attraction, $\theta_0 \in \mathcal{R}(\theta^*)$: $\theta_k \to \theta^*$ exponentially.*

The detailed proof with the corresponding derivations can be found in Appendix G. We draw on a recent result by Chasnov et al. (2020) on local convergence of gradient-based learners. Note that even without a priori convergence guarantees, we can certify an approximate BNE ex-post (see Section 4.3).

4.2. SODA

Instead of approximating pure strategies $\beta : \mathcal{V} \to \mathcal{A}$, *simultaneous online dual averaging* (SODA) (Fichtl et al., 2022) aims for distributional strategies in a discretized version of the auction game. Distributional strategies (Milgrom & Weber, 1985) are a form of mixed strategies for Bayesian games and are modeled as probability measures over $\mathcal{V}_i \times \mathcal{A}_i$. By discretizing the type spaces $\mathcal{V}_i$ and action spaces $\mathcal{A}_i$, we get discrete versions of the distributional strategies. In this setting, the set of feasible discrete distributional strategies $\mathcal{S}_i$ is a compact and convex subset of the probability simplex $\Delta^{N \cdot M}$, where N is the number of discretization points of the type space and M of the action space. Learning discrete distributional strategies means learning an $N \times M$ matrix, where each coefficient denotes the probability of the respective discrete type-action pair. The discretized auction game can be interpreted as a complete information game, where the set of feasible strategies $\mathcal{S}_i$ corresponds to a compact, convex action set, and the expected utility function corresponds to the respective utility function that is linear in the bidders' own actions.

This discretized formulation allows us to compute the gradient exactly, which implement well-known gradient-based learning methods for complete information games such as dual averaging. Dual Averaging (Nesterov, 2009) is based on two steps: (1) Given the current strategies of all traders, bidder i computes the individual gradient of the expected utility and performs a gradient ascent step in the dual space. (2) The updated dual variable is mirrored back to the feasible set in the primal space using a link function which leads to an updated strategy. This step is performed simultaneously by all bidders. It can be shown that if this procedure converges to a pure strategy for all bidders, then this profile is a Bayes–Nash Equilibrium for the discretized auction game (Fichtl et al., 2022, Corollary 1). Therefore, SODA provides an ex-post certificate. Moreover, for some single-object auction formats such as first or second-price sealed bid and all-pay auctions, it is shown that if SODA finds an approximate equilibrium of the discretized game, this is also an approximate equilibrium of the continuous auction game (Fichtl et al., 2022, Theorem 1).

To evaluate the computed strategies in the settings we consider, bids are sampled from the discrete distributional strategy. Given an observed valuation in the original continuous setting, the nearest discrete valuation is identified and a bid is sampled from the induced conditional probability distribution over the discrete bids.

4.3. Empirical certification

While global a priori convergence guarantees might be out of reach, we can verify the quality of a solution ex-post. Our primary evaluation metric will be the *relative efficiency* in terms of the gains from trade achieved in an equilibrium, which allows us to compare different environments. Besides, we will report metrics about the quality of the learned strategy profile β learned with NPGA and SODA.

Whenever we know the analytical equilibrium β^*, we use it for direct comparison. In this case, we sample the *BNE utility* of each player, $\hat{u}_i(\beta^*) = \frac{1}{n_{\text{batch}}} \sum_v u_i(v_i, \beta_i(v_i), \beta_{-i}(v_{-i})) \approx u_i^{\text{ante}}(\beta_i^*, \beta_{-i}^*)$, as well as the utility β_i played against the BNE, $\hat{u}_i(\beta_i, \beta_{-i}^*) \approx u_i^{\text{ante}}(\beta_i, \beta_{-i}^*)$, with a sample size of $n_{\text{batch}} = 2^{22}$ valuations from $\mathcal{V}$. Then, we report the resulting *relative ex-ante utility loss*:

$$\mathcal{L}_i(\beta_i) \;=\; 1 - \frac{\hat{u}_i(\beta_i, \beta_{-i}^*)}{\hat{u}_i(\beta_i^*, \beta_{-i}^*)}. \tag{20}$$

Besides, we report the probability-weighted root mean squared error of β_i and β_i^* in the action space, which approximates the L_2 distance $\|\beta_i - \beta_i^*\|_{\Sigma_i}$ of these two functions:

$$L_2(\beta_i) \;=\; \left(\frac{1}{n_{\text{batch}}} \sum_{v_i} \big(\beta_i(v_i) - \beta_i^*(v_i)\big)^2\right)^{\frac{1}{2}}. \tag{21}$$

This metric circumvents the drawback of $\mathcal{L}_i$ that even a strategy with a loss very close to zero could be arbitrarily far from the actual BNE in strategy space.

When no analytical BNE is available, we compute the ex-ante utility loss

$$\ell_i^{\text{ante}}(\beta_i, \beta_{-i}) = \sup_{\beta_i' \in \Sigma_i} u_i^{\text{ante}}(\beta_i', \beta_{-i}) - u_i^{\text{ante}}(\beta_i, \beta_{-i}). \tag{22}$$

Our estimator $\hat{\ell}_i$ of ℓ_i^{ante} relies on finding approximate interim best-responses. For this, we place an equidistant grid indexed with $w = 1, \dots, n_{\text{grid}}$ over the action space $\mathcal{A}_i$ ranging from zero to the maximum valuation. For a value v_i and each of the alternative bids b_w we evaluate the interim utility, $u_i^{\text{interim}}(v_i, b_w, \beta_{-i})$, against the current opponent strategy profile. In the case of independent private values, this is easily done by keeping v_i fixed and drawing a batch of samples from the opponents' valuations v_{-i}. For n_{batch} samples of v_i and n_{batch} samples of $v_{-i}|v_i$ for each of the v_i's, we then have

$$\hat{\ell}_i(\beta) = \frac{1}{n_{\text{batch}}} \sum_{v_i} \max_w \lambda_i(v_i, b_w, \beta) \tag{23}$$

with λ_i being the estimated expected utility gain by deviating from playing according to β_i to playing action b':

$$\lambda_i(v_i, b', \beta) = \frac{1}{n_{\text{batch}}} \cdot \sum_{v_{-i}|v_i} \Big(u_i\big(v_i, b', \beta_{-i}(v_{-i})\big) - u_i\big(v_i, \beta_i(v_i), \beta_{-i}(v_{-i})\big)\Big). \tag{24}$$

For an increasing number of samples and alternative actions, we have $\hat{\ell}_i \to \ell_i^{\text{ante}}$. Our estimate for ϵ in an ex-ante ϵ-BNE is then $\epsilon \equiv \max_i \hat{\ell}_i$. Based on these estimates, we can compute an *approximate relative ex-ante utility loss* without access to an analytical BNE:

$$\hat{\mathcal{L}}_i(\beta) \;=\; 1 - \frac{\hat{u}_i(\beta)}{\hat{u}_i(\beta) + \hat{\ell}_i(\beta)}. \tag{25}$$

This metric is the average loss incurred by not playing a best-response but instead playing the strategy learned via NPGA. For SODA we achieve a similar approximation of the utility loss by increasing the discretization to n_{grid}. The computed strategy is translated to the higher level of discretization by assigning the probability weights for a given valuation action pair to the nearest discrete action of the new discretization and distributing it among the closest valuations such that we get a feasible strategy. We can then compute the best-response and hence the relative utility loss $\hat{\mathcal{L}}$.

Hyperparameters that were used throughout our experiments for both algorithms can be found in Appendix A.

5. Results

This section summarizes the experimental results using NPGA and SODA. We analyze the few environments for which we have a closed-form equilibrium strategy and others for which this is not the case. Sometimes, we use the VCG mechanism as a baseline, for which we know that bidders have a dominant strategy to bid truthfully.

Table 1
Mean and standard deviation for different initialization procedures for the 1/2-double and VCG auction for NPGA over ten different seeds. The selective random initialization is a random initialization excluding those runs where one starts with non-trading strategies. The training period was 2000 iterations for all runs.

Auction	Initialization	Bidder	L_2	$\mathcal{L}$
0.5-DA	truthful	buyer	0.0081 (0.0042)	0.0028 (0.0004)
		seller	0.0076 (0.0031)	0.0004 (0.0003)
VCG	selective	buyer	0.0090 (0.0040)	0.0009 (0.0002)
	rand.	seller	0.0089 (0.0039)	0.0003 (0.0003)

Further experimental results for multiple buyers and sellers can be found in Appendix B.

5.1. Two quasi-linear traders with uniform priors

First, let us analyze the average mechanism ($k = 0.5$) with two quasi-linear traders and a uniform prior distribution for which closed-form solutions are available. We first report the results using NPGA and then those achieved with SODA. We show that NPGA reliably finds the welfare-maximizing linear BNE from Fig. 1(a), whereas SODA converges to different step function equilibria depending on the initialization.

5.2. NPGA

The first experiment is meant to validate that NPGA finds an equilibrium strategy and, if so, which one. The strategies are initially pretrained to be truthful to make them more comparable (see Section 4.1). The agents are subsequently trained for 2000 iterations. The results for ten different seeds are presented in the first two rows of Table 1. The relative utility loss $\mathcal{L}$ is close to zero, i.e., each bidder plays close to a best-response given that the opponent plays the linear BNE strategy. The L_2 loss is also low, which means the learned strategies are close to the linear BNE strategy in the L_2-norm. These results indicate that NGPA finds the linear equilibrium reliably for the truthful initialization, bypassing any suboptimal equilibrium from the class of differentiable equilibria (see Fig. 1(a)).

5.3. SODA

With SODA the results look different. In general the algorithm finds step function equilibria that show similar properties as the n-step equilibria mentioned in Section 3.3. One might argue that due to the discretization of the valuation and action space the computed strategies always resemble step function, but our experiments show that there are significant differences. For example, if we initialize the strategy near the welfare-maximizing linear equilibrium, the algorithm converges to a strategy that resembles a step function but closely approximates this equilibrium, which indicates that the equilibrium is at least locally attracting for SODA. In Table 2 we can see that the approximated L_2 distance to the linear equilibrium has almost the same accuracy as NPGA.

On the other hand, if we start with random initializations, we can observe that SODA consistently finds step function equilibria that might look different depending on the initialization or even the step size used in the algorithm. In this case, the computed strategies approximate step functions with very few steps (Fig. 3). Note that for low valuations of the buyer or high costs of the seller where no trade takes place, no strategy is learned and the bids are more or less at random in the respective interval. For the VCG

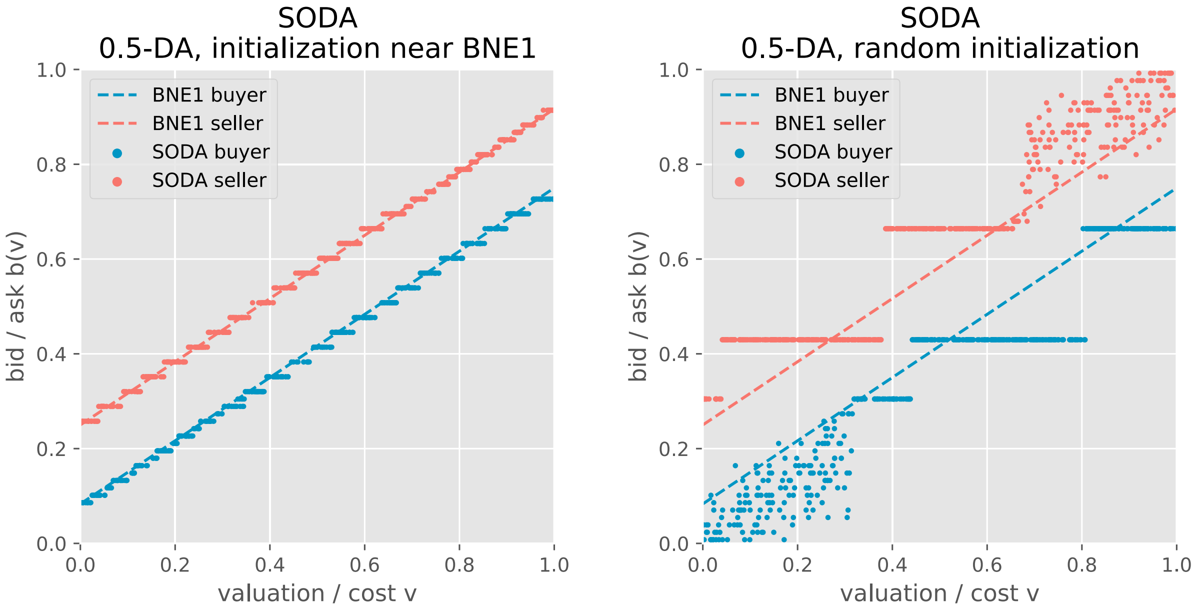


Fig. 3. 500 bids sampled from the strategies computed with SODA after initialization near the linear equilibrium BNE1 (left) and after random initilization (right) for the average mechanism with uniform prior.

Table 2
Mean and standard deviation over ten runs of SODA for the two most common mechanisms in the bilateral bargaining setup. For the average double auction, we only compare the learned strategies to the payoff dominant equilibrium strategies.

Auction	Initialization	Bidder	L_2	$\mathcal{L}$
0.5-DA	near equil.	buyer	0.0103 (0.0000)	0.0014 (0.0012)
		seller	0.0081 (0.0000)	0.0009 (0.0011)
	random	buyer	0.0734 (0.0063)	0.0398 (0.0151)
		seller	0.0725 (0.0064)	0.0386 (0.0150)
VCG	random	buyer	0.0140 (0.0003)	0.0006 (0.0000)
		seller	0.0139 (0.0004)	0.0006 (0.0000)

Table 3
Mean and standard deviation over ten runs of 2000 iterations with NPGA and SODA of the learning metrics for the two most common mechanisms in the bilateral bargaining setup for a Gaussian prior with Mean 15 and standard deviation 5. The NPGA strategies were pretrained on the truthful strategy for 500 iterations whereas SODA was initialized with random strategies.

Auction	Bidder	NPGA $\hat{\mathcal{L}}$	SODA $\hat{\mathcal{L}}$
0.5-DA	buyer	0.030 (0.002)	0.001 (0.002)
	seller	0.034 (0.006)	0.001 (0.001)
VCG	buyer	0.024 (0.000)	0.001 (0.000)
	seller	0.024 (0.000)	0.001 (0.000)

mechanism, the bids derived from the learned distributional strategy closely match the analytical equilibrium regardless of different initializations.

5.4. Two quasi-linear traders with Gaussian priors

The uniform distribution makes the analytical treatment much easier, but often one is interested in predictions for non-uniform priors. Below, we report SODA and NPGA for scenarios with a Gaussian prior for which no closed-form equilibrium is known. Table 3 shows the results for the VCG and average auction for Gaussian priors with a mean 15 and a standard deviation of 5 when running NPGA and SODA. The results are comparable to the uniform case in the sense that the learned strategies reach similar low levels of utility loss and SODA ends up in different step-function equilibria depending on the initialization in the average auction.

5.5. Two risk-averse traders

It is well-known that risk aversion among bidders mitigates the efficiency loss in double auctions and dates back to work by Chatterjee & Samuelson (1983). For the specific case of uniform priors and equal risk attitudes of the traders, we again can compare our results to the analytical equilibrium from Eq. (14). Fig. 4 compares the efficiency loss of the average double auction and the VCG double auction as predicted analytically and when learning with NPGA and SODA. Here, we measure the gains from trade in the strategy profile at hand compared to the gains from trade if the agents were truthful. As expected, the VCG mechanism is efficient throughout.

5.5.1. NPGA

For the average double auction, efficiency increases for higher levels of risk-aversion from about 84% under risk neutrality to above 99% for high levels of risk-aversion. One observes higher deviations from the predicted levels of efficiency for stronger risk aversion. This is explained by the fact that a decreasing exponent in $(u_i^{\mathrm{QL}})^\rho$ leads to its convergence to 1 for all values of u_i^{QL}, effectively squishing the learning signals of NPGA that only has a fixed absolute precision. This is also measured in the relative utility loss of NPGA (see Table 4), where we observe a correlation between low-risk attitudes (larger values of ρ) towards better performance. Overall, the relative utility loss decreases consistently below 1.4%.

5.5.2. SODA

When learning with SODA, the increasing efficiency with higher levels of risk-aversion can also be observed for the step-function equilibria, albeit at a lower level. It is surprising that despite the different outcomes in the computed strategies regarding the number and position of the steps, a consistent level of efficiency with a standard deviation of less than 1% is achieved for fixed risk parameters. In general, we see that as risk aversion increases, the number of steps in the approximated strategies increases and the strategies continue to converge to the linear equilibria (see Table 4).

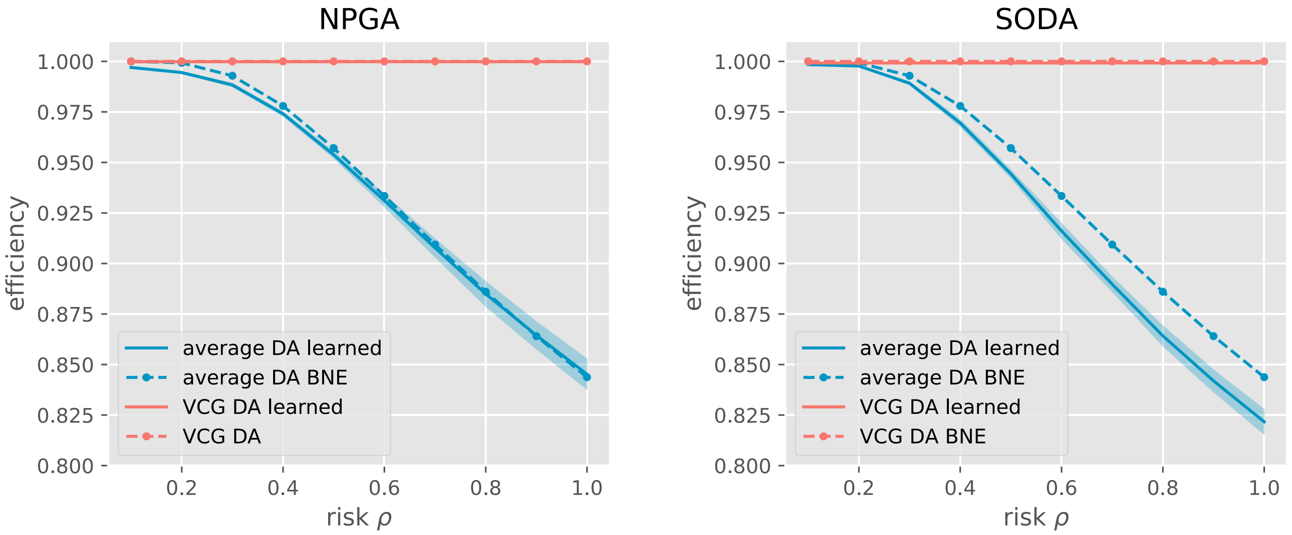


Fig. 4. Mean and standard deviation of efficiency for NPGA (left) and SODA (right) applied to the average and VCG double auction with different risk parameters. Dashed lines depict efficiency in the linear BNE.

Table 4
Evaluation of the algorithms for multiple levels of risk aversion in the average double auction. Results are averaged over five runs each.

Risk ρ	Bidder	NPGA L_2	NPGA $\mathcal{L}$	SODA L_2	SODA $\mathcal{L}$
0.1	buyer	0.015 (0.002)	0.011 (0.002)	0.016 (0.000)	0.014 (0.000)
	seller	0.016 (0.002)	0.014 (0.003)	0.016 (0.000)	0.014 (0.000)
0.5	buyer	0.007 (0.001)	0.001 (0.000)	0.044 (0.003)	0.018 (0.003)
	seller	0.007 (0.003)	0.001 (0.000)	0.043 (0.004)	0.018 (0.004)
0.9	buyer	0.007 (0.002)	0.002 (0.000)	0.066 (0.005)	0.033 (0.010)
	seller	0.007 (0.002)	0.000 (0.000)	0.065 (0.006)	0.031 (0.011)

6. Conclusions

Bilateral trade is an interesting environment to study. First, it is as simple as possible, with only a single participant on each side and a single object. With independent and uniform prior distributions and possibly risk-averse bidders, we even have a simple linear equilibrium bidding strategy. The environment, nonetheless, is very challenging because there is a continuum of equilibria, so it is unclear whether equilibrium computation would converge in this setting. Under strong assumptions, such as linear bid functions, one can study the expected utility landscape in much more detail than would be possible in richer environments. In this case, one can even derive analytical solutions to the equilibrium problem, but the necessary assumptions also illustrate how restrictive the analytical approach is to provide a priori convergence guarantees.

The equilibrium learning algorithms analyzed in this paper allow for equilibrium analysis in far more general environments. An open question concern ex-ante properties for convergence to equilibrium. The concavity of the utility functions and payoff monotonicity of the game are known properties of convergence. However, they are difficult to check in games with continuous type- and action spaces. We show that the utility functions are concave in large domains in equilibrium. However, we can also show that the game is not globally monotone, and we cannot rely on convergence results for variational inequalities. Nevertheless, we can prove local convergence of NPGA in this specific bilateral trade model. Further, we use both techniques to find equilibrium in a variety of bilateral trade environments for which no explicit equilibrium bid function has been known so far. This includes bilateral bargaining with Gaussian priors or risk-averse traders. We report experiments with multiple buyers, multiple sellers, or both in the appendix. This way, the paper pushes the boundaries of equilibrium computation and contributes to understanding equilibrium learning in the simplest and arguably most well-known model of trade.

Acknowledgment

We gratefully acknowledge financial support from the German National Science Foundation through Grant DFG BI 1057/9-1.

Appendix A. Reproducibility and hyperparameters

All our experiments are run with the following learning parameters, if not specified otherwise.

A1. NPGA

We use common hyperparameters across almost all settings (except where noted otherwise): Fully connected neural networks with two hidden layers of ten nodes each with SeLU activations on the inner nodes (Klambauer et al., 2017), as well as ReLU activations in the output layer. The parameters θ_i are then given by the weights and biases of these networks. All experiments were performed on a single Nvidia GeForce 2080Ti with 11 gigabyte of RAM and batch sizes in Monte–Carlo sampling were chosen to maximize GPU-RAM utilization: A learning batch size of $n_{\text{batch}} = 2^{18}$; primary evaluation batch size (for $\mathcal{L}$ and L_2) of 2^{22}; and secondary evaluation batch size 2^{13} and grid size $n_{\text{grid}} = 2^{10}$ (for $\hat{\ell}$ and $\hat{\epsilon}$). The code will be available at blinded for review. Run times for the markets with a single seller and a single buyer are around 0.36 seconds per iteration. The more extensive experiments with up to eight agents took about 0.95 seconds per iteration. The middle column of

Table A1
Mean runtime per iteration for NPGA and SODA with a different number of agents. The average is over all iterations and experiments with a uniform prior distribution.

Num agents	Time/iter [s] NPGA	Time/iter [s] SODA
2	0.363	0.001
3	0.506	0.035
4	0.561	2.281
5	0.624	–
6	0.823	–
8	0.949	–

Table A1 shows the average time per iteration for a different number of agents per experiment. We found that it made no difference for the runtime whether we have more buyers or sellers but only the total number of agents. We averaged over all seeds and runs with the same total number of agents using a uniform distribution to make the results comparable. The results show that the runtime increases sublinearly in the number of agents, demonstrating the efficiency of running the whole learning process on GPU.

A2. SODA

To discretize the problem we split the valuation and action space in $n_{\text{discr}} = 64$ equally sized intervals and take the respective midpoints as discretization points. If the valuation space is unbounded we only consider a suitable compact interval, e.g., $[0, 30]$ for the Gaussian prior $\mathcal{N}(15, 5)$. Further, we assume that the action space is equal to the valuation space. For the update step in the dual space we use a decreasing step size of the form $\eta_t = \eta_0 / t^{\beta}$ where t is the current iteration, $\beta = 0.05$ and $\eta_0 = 200$ for uniform priors, and $\eta_0 = 20$ for the gaussian prior. The algorithms either stops after 2000 iterations or when the relative utility loss within the discretized setting is less than 10^{-4}. All experiments where performed on a single Intel Core i7-8565U CPU @ 1.80 gigahertz and 16 gigabyte of RAM. The way the game is discretized limits the applicability of SODA due to the curse of dimensionality. To compute the gradient or the utility, given a strategy profile, one must take the weighted sum over all possible valuation and action profile combinations. The number of such possible combinations increases exponentially in the number of agents, i.e., $n_{\text{discr}}^{n_B+n_S+1}$. This has significant impact on the running time as we can observe in Table A1 and on the amount of storage required. For this reason, we could not, for instance, calculate the utility loss $\hat{\mathcal{L}}$ for three or four agents, or even compute the respective strategies for larger settings on our current hardware with SODA. Therefore, we only report the results for NPGA in Appendix B.

Appendix B. Experiments for multiple buyers and sellers

Up to this point, we considered bilateral bargaining with one buyer and one seller only. Next, we study markets with multiple buyers or sellers. For the k-double auction, already for one seller and two buyers (or vice versa), there is no closed-form BNE. From the view of a single buyer (seller), the task is symmetric in the sense that each buyer (seller) has the same utility function and faces opponents from the same market side with the same prior distributions. We conducted experiments allowing different strategies for all agents on both sides of the market, thus, allowing for the discovery of asymmetric equilibria. We found that there was no significant difference and, therefore, restrict our presentation to symmetric strategies for each market side for clarity in the presentation. This slightly reduces memory consumption and the variance in learning.

We are going to place a special emphasis on the market efficiency in analyzing equilibria in markets using the k-double auction. That is due to a number of articles that analyze the implications of increasing the level of competition market efficiency Rustichini et al. (1994); Wilson (1985). Overall, the inefficiency in a k-double auction decreases for symmetrically growing markets (Cripps & Swinkels, 2006). However, this increase in efficiency does not happen if the market is growing asymmetrically, e.g., if the number of buyers grows faster than the number of sellers.

Table B1
Mean and standard deviation over ten runs of 2000 iterations with NPGA for the 0.5-DA mechanism with several buyers and one seller for a uniform prior.

Auction	Bidder	$\hat{\mathcal{L}}$	L_2^{truthful}
2b1s	buyers	0.065 (0.004)	0.060 (0.004)
	seller	0.051 (0.001)	0.200 (0.004)
3b1s	buyers	0.065 (0.004)	0.039 (0.007)
	seller	0.043 (0.001)	0.248 (0.004)
4b1s	buyers	0.070 (0.004)	0.035 (0.011)
	seller	0.038 (0.002)	0.281 (0.003)

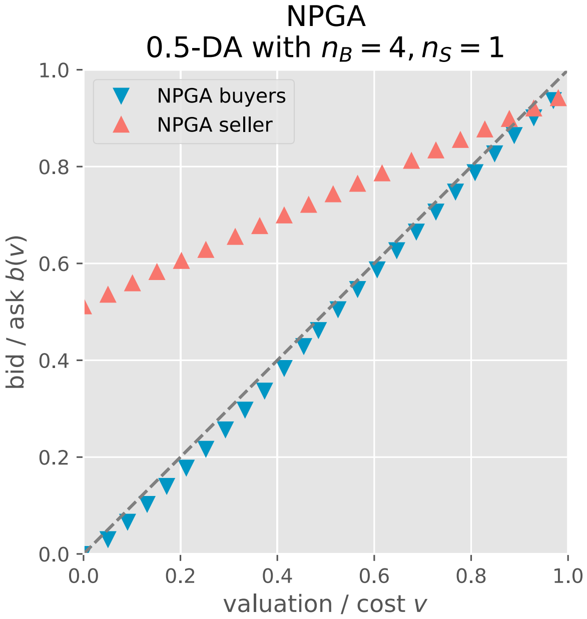


Fig. B1. The strategies of four buyers and one seller after 2000 iterations with NPGA for a uniform prior in the average mechanism.

B1. Asymmetrically growing markets

Let us first analyze asymmetric markets with multiple buyers and one seller (or vice versa). Imagine a case with n_B buyers and one seller, where the buyers' priors are independent. For a drawn valuation v_S of the seller, denote the probability that the valuation v_{B_i} of one of the buyers is below v_S by $P(v_{B_i} < v_S)$. Then the probability that all buyers' valuations are below v_S is given by $\prod_{i=1}^{n}(1 - P(v_{B-i} < v_S))$. This means, for more buyers, it becomes more likely that at least one buyer's valuation is above that of the seller. A seller can leverage this asymmetry for his strategy, which is something that we can observe in our experiments.

Table B1 shows the approximate relative utility loss of the traders and the distance to the truthful strategies for 2, 3, and 4 buyers (2b-4b) and one seller (1s). Whereas the buyers' strategies tend towards the truthful strategy the more buyers participate in the market, the single seller's strategy deviates more from it. Fig. B1 illustrates this observation for the case of four buyers and one seller. The buyers' strategy is very close to being truthful

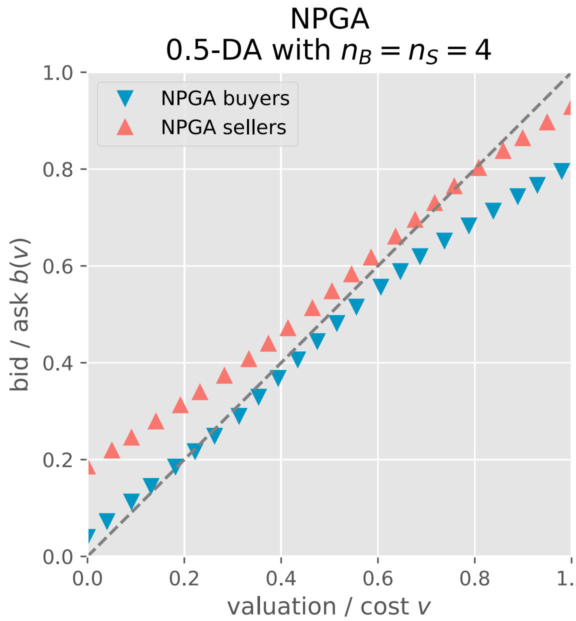


Fig. B2. The strategies of buyers and seller after 4,000 iterations with NPGA for a uniform prior with four buyers and sellers in the average-auction.

Table B2
Mean and standard deviation over ten runs of 4,000 NPGA iterations of the learning metrics for the 0.5-DA mechanism in a double auction setup with several buyers and sellers for a uniform prior.

Auction	Bidder	$\hat{\mathcal{L}}$	L_2^{truthful}
2b2s	buyers	0.046 (0.001)	0.104 (0.005)
	sellers	0.046 (0.001)	0.107 (0.002)
3b3s	buyers	0.039 (0.001)	0.089 (0.004)
	sellers	0.039 (0.002)	0.093 (0.002)
4b4s	buyers	0.036 (0.001)	0.083 (0.003)
	sellers	0.036 (0.001)	0.085 (0.003)

(blue downward-pointing triangles), whereas the seller's strategy is to bid significantly higher for lower costs (red upward-pointing triangles). One gets qualitatively similar results for the reversed scenario with multiple sellers and one buyer.

B2. Symmetrically growing markets

Theoretical results suggest that a symmetric market with more buyers and sellers should become more and more efficient with growing size (Cripps & Swinkels, 2006). That is, for the number of buyers and sellers going to infinity, all non-trivial BNE strategies for buyers and sellers are converging towards the truthful strategy.

Fig. B2 shows the learned strategies in a scenario with four buyers and sellers with NPGA after 4,000 iterations. We can see that the learned strategies are closer to the truthful strategy (which is also depicted as reference). This observation is supported by Table B2. The distance to truthful strategies is decreasing with an increasing number of buyers and sellers.

Appendix C. First-order conditions in bilateral bargaining

For drawn valuations $v_B \sim f_B$ and $v_S \sim f_S$ of buyer and seller, respectively, let the buyer's bid be $b_B = \beta_B(v_B)$ and the seller's ask be $b_S = \beta_S(v_S)$. Then, the ex-post utility of the buyer is given by

$$u_B(v_B, b_B, b_S) = \mathbb{1}_{\{b_B \geq b_S\}} \cdot (v_B - P(b_B, b_S)),$$

where P denotes the price function that the buyer has to pay and the seller receives. For some other mechanisms, one may also want to differentiate between the payments. The seller's corresponding ex-post utility is given by

$$u_S(v_S, b_B, b_S) = \mathbb{1}_{\{b_B \geq b_S\}} \cdot (P(b_B, b_S) - v_S).$$

If the buyer's bid b_B is smaller than the lowest ask price $\underline{b_S}$, the buyer's interim utility is zero. This describes a case where the buyer bids so little that there is no trade for any valuation of the seller. Reversely, the same holds for the seller's interim utility if the seller's ask price b_S is higher than the highest bid of the buyer $\overline{b_B}$. We derive the interim utilities for all other cases next. We will start with the buyer's assuming that $\beta_B(v_B) \geq \underline{b_S}$:

$$\begin{aligned}
&\mathbb{E}_{v_S \sim f_S}[u_B(v_B, b_B, \beta_S(v_S))] \\
&= \int_{\Omega_S} u_B(v_B, \beta_B(v_B), \beta_S(v_S)) \cdot f_S(v_S) dv_S \\
&= \int_{\beta_S^{-1}(\hat{\Omega}_S)} u_B(v_B, \beta_B(v_B), \beta_S(v_S)) \cdot f_S(v_S) dv_S \\
&\overset{(*_1)}{=} \int_{\hat{\Omega}_S} u_B(v_b, \beta_B(v_B), y) \cdot f_S(\beta_S^{-1}(y)) \cdot |(\beta_S^{-1})'(y)| dy \\
&\overset{\text{prop. 1}}{=} \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} (v_B - P(\beta_B(v_B), y)) \cdot f_S(\beta_S^{-1}(y)) \cdot (\beta_S^{-1})'(y) dy \\
&\overset{\text{PI}}{=} \left[(v_B - P(\beta_B(v_B), y)) \cdot F_S(\beta_S^{-1}(y))\right]_{y=\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} \\
&+ \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} \frac{d}{dy} P(\beta_B(v_B), y) \cdot F_S(\beta_S^{-1}(y)) dy \\
&\overset{(*_2)}{=} (v_B - P(\beta_B(v_B), \beta_B(v_B))) \cdot F_S\left(\beta_S^{-1}\left(\min\{\beta_B(v_B), \overline{b_S}\}\right)\right) \\
&+ \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} \frac{d}{dy} P(\beta_B(v_B), y) \cdot F_S(\beta_S^{-1}(y)) dy.
\end{aligned}$$

Note that we used substitution in multivariate integrals for bi-Lipschitz functions (Federer, 1996) in step $(*_1)$ which uses both conditions of Assumption 1. In step $(*_2)$, one can see that $\beta_S^{-1}(\underline{b_S}) = \underline{v_S}$, again due to Assumption 1. This results in $F_S(\beta_S^{-1}(\underline{b_S})) = F_S(\underline{v_S}) = 0$, as F_S is the CDF of f_S on $[\underline{v_S}, \overline{v_S}] = \Omega_S$.

Analog derivations for the seller's interim utility give the following under the assumption that $\beta_S(v_S) \leq \overline{b_B}$.

$$\begin{aligned}
&\mathbb{E}_{v_B \sim f_B}[u_S(v_S, \beta_B(v_B), b_S)] \\
&= \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} (P(x, \beta_S(v_S)) - v_S) \cdot f_B(\beta_B^{-1}(x)) \cdot (\beta_B^{-1})'(x) dx \\
&\overset{\text{PI}}{=} \left[(P(x, \beta_S(v_S)) - v_S) \cdot F_B(\beta_B^{-1}(x))\right]_{x=\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} \\
&- \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} \frac{d}{dx} P(x, \beta_S(v_S)) \cdot F_B(\beta_B^{-1}(x)) dx \\
&= \left(P(\overline{b_B}, \beta_S(v_S)) - v_S\right) \\
&- \left(P\left(\max\{\beta_S(v_S), \underline{b_B}\}, \beta_S(v_S)\right) - v_S\right) \cdot F_B\left(\beta_B^{-1}\left(\max\{\beta_S(v_S), \underline{b_B}\}\right)\right) \\
&- \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} \frac{d}{dx} P(x, \beta_S(v_S)) \cdot F_B(\beta_B^{-1}(x)) dx.
\end{aligned}$$

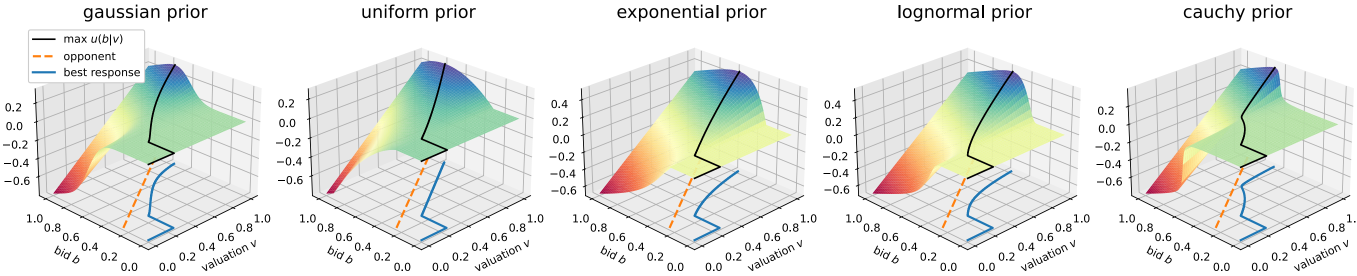


Fig. D1. Utility of the buyer under an opposing seller that plays according to the linear strategy that constitutes a BNE under uniform priors.

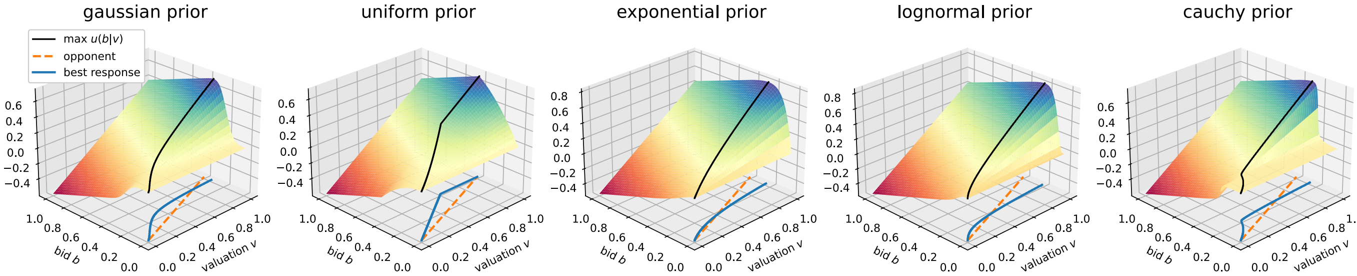


Fig. D2. Utility of the buyer under an opposing seller that plays according to some different linear strategy: $\beta(v) = \frac{2}{5}v$.

Note that the term including the maximal buyer's bid does not equal zero, which was the case for the minimal seller's ask price in the derivations for the buyer's interim utility. With the definition of the allocation as above and for the case of the k-double auction, the buyer's interim utility is given by

$$u_B^{\text{interim}}(v_B, \beta_B(v_B), \beta_S)$$
$$= \mathbb{1}_{\{\beta_B(v_B)\geq \underline{b_S}\}} \cdot \Bigg((v_B - \beta_B(v_B)) \cdot F_S(\beta_S^{-1}(\min\{\beta_B(v_B), \overline{b_S}\}))$$
$$+ (1-k) \cdot \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} F_S(\beta_S^{-1}(y))dy \Bigg), \tag{C.1}$$

and the seller's interim utility by

$$u_S^{\text{interim}}(v_S, \beta_B, \beta_S(v_S))$$
$$= \mathbb{1}_{\{\overline{b_B}\geq \beta_S(v_S)\}} \cdot \Bigg(\Big(k \cdot \overline{b_B} + (1-k) \cdot \beta_S(v_S) - v_S\Big) \cdot F_B(\beta_B^{-1}(\overline{b_B}))$$
$$- \Big(k \max\big(\beta_S(v_S), \underline{b_B}\big) + (1-k)\beta_S(v_S) - v_S\Big) \cdot F_B(\beta_B^{-1}(\max\{\beta_S(v_S), \underline{b_B}\}))$$
$$- k \cdot \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} F_B(\beta_B^{-1}(x))dx \Bigg). \tag{C.2}$$

The first-order conditions are then given by the following system of non-linear ODEs:

$$A(v_B, v_S, \beta_B, \beta_S) := \begin{pmatrix} \frac{d}{d\beta_B(v_B)} u_B^{\text{interim}}(v_B, \beta_B(v_B), \beta_S) \\ \frac{d}{d\beta_S(v_S)} u_S^{\text{interim}}(v_S, \beta_B, \beta_S(v_S)) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{C.3}$$

Appendix D. Utility landscape

In this section, we will take an empirical look at the utility landscape that the buyer faces under different circumstances in the bilateral trade setting. Figs. D1–D4 show the buyer's utility for all his possible valuations and actions against different sellers. Assuming specific priors and specific strategies of the sellers, the utility can be derived analytically. All resulting utility functions of the buyer are concave in large ranges, which might explain why gradient-based methods consistently converge in all settings considered.

Appendix E. Expected utility with linear strategies

In what follows, we derive the expected utility of buyer and seller with independent uniform priors and quasi-linear utility in the average double auction. Furthermore, we assume that the strategies are linear. That is, there exist $m_B, t_B, m_S, t_S \in \mathbb{R}$ such that $\beta_B(v_B) = m_Bv_B + t_B$ and $\beta_S(v_S) = m_Sv_S + t_S$. Finally, we restrict the feasible set as described in Section 3.4 according to Assumptions 1 and 2.

Using Eqs. (2) and (7), the buyer's ex-ante utility is given by

$$u_B^{\text{ante}}(\beta_B, \beta_S, k) = u_B^{\text{ante}}(m_B, t_B, m_S, t_S, k) \tag{E.1}$$

$$= \mathbb{E}_{v_B\sim f_B}\left[u_B^{\text{interim}}(v_B, m_Bv_B + t_B, (m_S, t_S), k)\right] \tag{E.2}$$

$$= \int_{\frac{1}{m_B}(t_S - t_B)}^{1} u_B^{\text{interim}}(v_B, m_Bv_B + t_B, (m_S, t_S), k)dv_B. \tag{E.3}$$

Note that here we used that the PDF of the uniform distribution is constant on the unit interval, $f_B(v_B) = 1$, and that the integral's lower bound comes from the buyer's interim utility being zero if the bid is below the lowest ask price of the seller. That is $\beta_B(v_B) = m_Bv_B + t_B < \underline{b_S} = t_S$. As the strategies are strictly increasing, we get for all valuations $v_B < \frac{1}{m_B}(t_S - t_B)$ that the inner term in the integral is zero. That means we can calculate the inner term first and then take the integral afterward. For the case of $\beta_B(v_B) \geq t_S$, the inner term is given by

$$u_B^{\text{interim}}(v_B, m_Bv_B + t_B, (m_S, t_S), k)$$
$$= (v_B - \beta_B(v_B)) \cdot F_S(\beta_S^{-1}(\beta_B(v_B))) + (1-k) \cdot \int_{\underline{b_S}}^{\beta_B(v_B)} F_S(\beta_S^{-1}(y))dy \tag{E.4}$$

$$= (v_B - m_Bv_B - t_B) \cdot F_S(\beta_S^{-1}(m_Bv_B + t_B)) \tag{E.5}$$

$$+ (1-k) \int_{t_S}^{m_Bv_B+t_B} F_S\Big(\frac{1}{m_S}(y - t_S)\Big)dy$$
$$= (v_B - m_Bv_B - t_B) \cdot F_S\Big(\frac{1}{m_S}(m_Bv_B + t_B - t_S))\Big) \tag{E.6}$$

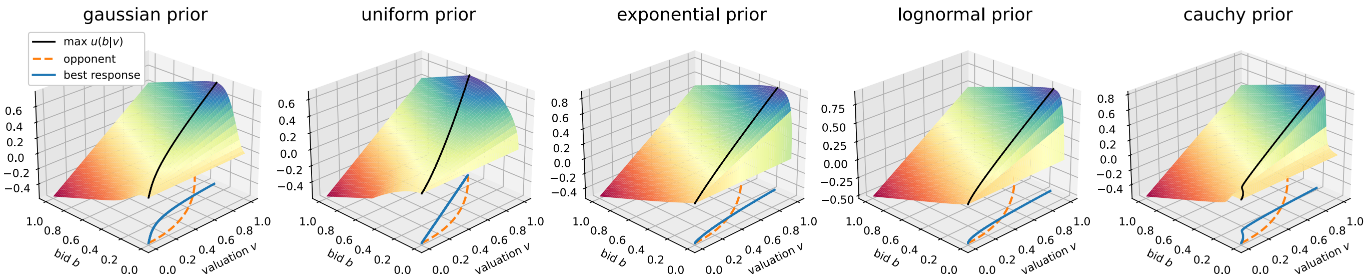


Fig. D3. Utility of the buyer under an opposing seller that plays according to some convex strategy: $\beta(v) = \frac{1}{2}v^2$.

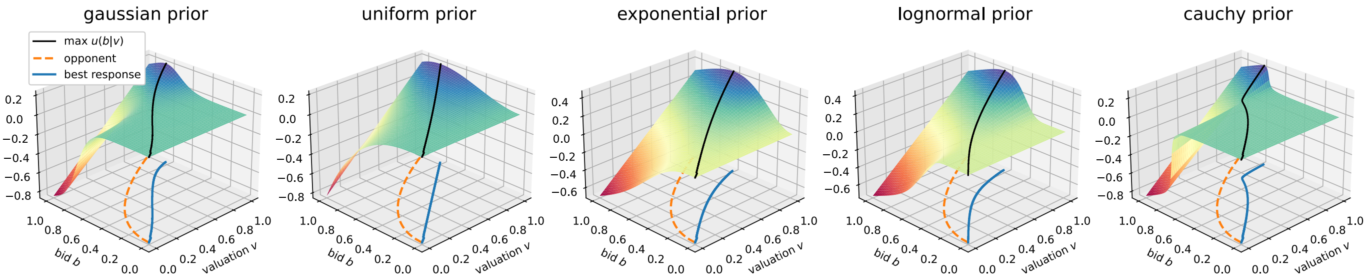


Fig. D4. Utility of the buyer under an opposing seller that plays according to some concave strategy: $\beta(v) = \sqrt{v}$.

$$+ (1-k)\int_{t_S}^{m_B v_B + t_B} F_S\Big(\frac{1}{m_S}(y - t_S)\Big)dy$$

$$\overset{(*_1)}{=} (v_B - m_B v_B - t_B)\cdot\frac{1}{m_S}(m_B v_B + t_B - t_S) \tag{E.7}$$

$$+ (1-k)\int_{t_S}^{m_B v_B + t_B} \frac{1}{m_S}(y - t_S)dy$$
$$= \frac{1}{m_S}\cdot\Bigg[(v_B - m_B v_B - t_B)(m_B v_B + t_B - t_S) + (1-k)\Big[\frac{1}{2}y^2 - t_S y\Big]_{y=t_S}^{m_B v_B + t_B}\Bigg] \tag{E.8}$$

$$= \frac{1}{m_S}\big[-m_B^2 v_B^2 - 2m_B t_B v_B + m_B v_B^2 + t_S m_B v_B - t_B^2 + t_B v_B + t_S t_B - t_S v_B\big]$$
$$+ \frac{1-k}{2m_S}\big[m_B^2 v_B^2 + 2m_B t_B v_B - 2m_B t_S v_B + t_B^2 - 2t_B t_S + t_S^2\big] \tag{E.9}$$

$$= \frac{1}{m_S}\big[\big(-m_B^2 + m_B\big)v_B^2 + (-2m_B t_B + t_S m_B + t_B - t_S)v_B - t_B^2 + t_S t_B\big]$$
$$+ \frac{1-k}{2m_S}\big[m_B^2 v_B^2 + (2m_B t_B - 2m_B t_S)v_B + (t_B - t_S)^2\big]. \tag{E.10}$$

In step $(*_1)$, we get that the argument of F_S always comes from the unit interval. The lower bound of the integral t_S evaluates to an argument of zero, whereas the upper bound gives $\frac{1}{m_S}((m_B v_B + t_B) - t_S) \leq \frac{1}{m_S}((m_S + t_S) - t_S) = 1$ by property 3 of Assumption 2.

We proceed by calculating the integral from Eq. (E.3). This gives us

$$u_B^{\text{ante}}(m_B, t_B, m_S, t_S, k)$$
$$= \left(\frac{m_B - m_B^2}{m_S} + \frac{1-k}{2m_S}m_B^2\right)\cdot\int_{\frac{1}{m_B}(t_S - t_B)}^{1} v_B^2 dv_B$$
$$+ \left(\frac{t_S m_B - 2m_B t_B + t_B - t_S}{m_S} + \frac{1-k}{2m_S}(2m_B t_B - 2m_B t_S)\right)\cdot\int_{\frac{1}{m_B}(t_S - t_B)}^{1} v_B dv_B$$
$$+ \left(\frac{t_S t_B - t_B^2}{m_S} + \frac{(1-k)(t_B - t_S)^2}{2m_S}\right)\left(1 - \frac{t_S - t_B}{m_B}\right) \tag{E.11}$$

$$= \left(-\frac{m_B(m_B + k m_B - 2)}{2m_S}\right)\left(\frac{(t_B - t_S)^3}{3m_B^3} + \frac{1}{3}\right)$$
$$+ \left(-\frac{t_S - t_B + m_B t_B + k m_B t_B - k m_B t_S}{m_S}\right)\left(\frac{1}{2} - \frac{(t_S - t_B)^2}{2m_B^2}\right)$$
$$+ \left(\frac{t_S t_B - t_B^2}{m_S} + \frac{(1-k)(t_B - t_S)^2}{2m_S}\right)\left(1 - \frac{t_S - t_B}{m_B}\right). \tag{E.12}$$

We expand each of the three terms first, before collapsing it back into a function of m_B, m_S, t_B, t_S, and k.

$$= \left(-\frac{1}{2m_S}\right)\cdot\left(\frac{1}{3m_B^2}\right)(m_B + k m_B - 2)\big((t_B - t_S)^3 + m_B^3\big)$$
$$+ \left(-\frac{1}{2m_S}\right)\cdot\left(\frac{1}{3m_B^2}\right)\cdot 3(t_S - t_B + m_B t_B + k m_B t_B - k m_B t_S)\big(m_B^2 - (t_S - t_B)^2\big)$$
$$+ \left(-\frac{1}{2m_S}\right)\cdot\left(\frac{1}{3m_B^2}\right)\cdot 3m_B\big(2t_S t_B - 2t_B^2 + (1-k)(t_B - t_S)^2\big)(t_S - t_B - m_B) \tag{E.13}$$

$$= -\frac{1}{6m_S m_B^2}\big(k m_B^4 + m_B t_B^3 - m_B t_S^3 - 6t_B t_S^2 + 6t_B^2 t_S - 2m_B^3 + m_B^4 - 2t_B^3 + 2t_S^3\big)$$
$$- \frac{1}{6m_S m_B^2}\big(k m_B t_B^3 - k m_B t_S^3 + 3m_B t_B t_S^2 - 3m_B t_B^2 t_S + 3k m_B t_B t_S^2 - 3k m_B t_B^2 t_S\big)$$
$$- \frac{1}{6m_S m_B^2}\big(3m_B^3 t_B - 3m_B t_B^3 - 3m_B^2 t_B + 3m_B^2 t_S + 9t_B t_S^2 - 9t_B^2 t_S + 3t_B^3$$
$$-3t_S^3 - 3k m_B t_B^3\big)$$
$$- \frac{1}{6m_S m_B^2}\big(3k m_B^3 t_B + 3k m_B t_S^3 - 3k m_B^3 t_S - 3m_B t_B t_S^2 + 6m_B t_B^2 t_S - 9k m_B t_B t_S^2$$
$$+9k m_B t_B^2 t_S\big)$$
$$- \frac{1}{6m_S m_B^2}\big(3m_B^2 t_B^2 - 3m_B^2 t_S^2 + 3m_B t_B^3 + 3m_B t_S^3 + 3k m_B t_B^3 - 3k m_B t_S^3 - 3m_B t_B t_S^2\big)$$
$$- \frac{1}{6m_S m_B^2}\big(-3m_B t_B^2 t_S + 3k m_B^2 t_B^2 + 3k m_B^2 t_S^2 + 9k m_B t_B t_S^2 - 9k m_B t_B^2 t_S - 6k m_B^2 t_B t_S\big) \tag{E.14}$$

$$= -\frac{(m_B + t_B - t_S)^2\big(t_B - t_S + m_B(m_B + t_B + 2t_S - 2) + m_B k(m_B + t_B - t_S)\big)}{6m_B^2 m_S}. \tag{E.15}$$

Inserting the seller's linear equilibrium strategy, $\beta_S(v_S) = \frac{2}{3}v_S + \frac{1}{4}$, into the equation of the buyer's ex-ante utility indeed verifies that it has a local maximum at the buyer's corresponding equilibrium strategy, $\beta_B(v_B) = \frac{2}{3}v_B + \frac{1}{12}$. (See Section 3.3 for more details on the equilibrium strategies.) The buyer's ex-ante utility landscape in this scenario is depicted in Fig. 2, which shows the local maximum at that point.

We repeat this process for the seller's ex-ante utility.

$$u_S^{\text{ante}}(m_B, t_B, m_S, t_S, k)$$
$$= \mathbb{E}_{v_S \sim f_S}\left[u_S^{\text{interim}}(v_S, (m_B, t_B), m_S v_S + t_S, k)\right] \tag{E.16}$$

$$= \int_0^{\frac{1}{m_S}(m_B+t_B-t_S)} u_S^{\text{interim}}(v_S, (m_B, t_B), m_S v_S + t_S, k) dv_S \tag{E.17}$$

$$= -\frac{(m_B + t_B - t_S)^3 - m_S(m_B + t_B - t_S)^2(m_B + t_B + 2t_S + km_B + kt_B - kt_S)}{6m_B {m_S}^2}. \tag{E.18}$$

Appendix F. Monotonicity of the parametrized game

Rosen (1965) introduced the notion of (strict) monotonicity[5] in games, which has been established as central concept to show convergence of learning algorithms in games (Guo et al., 2021; Mertikopoulos & Zhou, 2019). One can formulate the ex-ante game as variational inequality over the infinite dimensional action space Σ. As we know that the game has more than one equilibrium in Σ, one can already derive that the game is not strictly monotone (Cavazzuti et al., 2002). In this section, we demonstrate that this negative result extends to discretizations of the strategy space, as is done with NPGA and SODA. NPGA considers the parameter space of a neural network, whereas SODA discretizies the type and action spaces themselves. Monotonicity, by itself thus cannot explain the positive convergence results we observed in practice.

Let us start by considering NPGA. For simplicity, we formulate the monotonicity condition for two players and refer to Rosen (1965) for additional details. Consider a game between two players $i \in \{1, 2\}$, with action spaces $E_i \subset \mathbb{R}^{m_i}$, $m_i \in \mathbb{N}$, and continuously differentiable utility functions $U_1, U_2 : E \to \mathbb{R}$ for $E = E_1 \times E_2$. Denote the payoff gradients by $v_i = \nabla_{y_i} U_i(y_1, y_2)$ for $i \in \{1, 2\}$ and $v = [v_1, v_2]^T$.

Definition 1. Such a game is called *strictly monotone* if

$$\langle v(y') - v(y), y' - y\rangle \leq 0 \quad \text{for all } y, y' \in E, \tag{F.1}$$

where equality holds if and only if $y \neq y'$.

The NPGA algorithm's setting in general double auctions can be identified with the game above (see Section 4.1). We consider the setting with linear strategies introduced in Section 3.4 for the average double auction (i.e., $k = 0.5$). Using the derivations for the ex-ante utilities from Eqs. (E.15), Eq. (18) in Appendix E, we can derive the payoff gradients

$$v_{\text{ls}}(m_B, t_B, m_S, t_S) = \begin{pmatrix} \frac{d}{dm_B} u_B^{\text{ante}} \\ \frac{d}{dt_B} u_B^{\text{ante}} \\ \frac{d}{dm_S} u_S^{\text{ante}} \\ \frac{d}{dt_S} u_S^{\text{ante}} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{(t_B-t_S)^3}{3\,{m_B}^3\,m_S} - \frac{6\,m_B+9\,t_B-3\,t_S-4}{12\,m_S} + \frac{(t_B+t_S)\,(t_B-t_S)^2}{4\,{m_B}^2\,m_S} \\ -\frac{(m_B+t_B-t_S)\left(t_B-m_B-t_S+\frac{3\,m_B\,t_B}{2}+\frac{m_B\,t_S}{2}+\frac{3\,{m_B}^2}{2}\right)}{2\,{m_B}^2\,m_S} \\ \frac{(m_B+t_B-t_S)^3}{3\,m_B\,{m_S}^3} - \frac{(m_B+t_B-t_S)^2\,(m_B+t_B+t_S)}{4\,m_B\,{m_S}^2} \\ \frac{(m_B+t_B-t_S)^2-m_S\,(m_B+t_B-t_S)\left(\frac{m_B}{2}+\frac{t_B}{2}+\frac{3\,t_S}{2}\right)}{2m_B\,{m_S}^2} \end{pmatrix}.$$

Consider the following two points

$$y' = \begin{pmatrix} 0.080 \\ 0.171 \\ 0.250 \\ 0.200 \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} 0.080 \\ 0.171 \\ 0.260 \\ 0.199 \end{pmatrix}.$$

Then one can directly verify that these points do satisfy Assumptions 1 and 2. However, plugging these points into Eq. (F.1) gives

$$\langle v_{\text{ls}}(y') - v_{\text{ls}}(y), y' - y\rangle = \frac{3{,}631}{3{,}000{,}000{,}000} > 0.$$

Therefore, the monotonicity condition does not hold. This is a strong indication that using monotonicity to derive global convergence guarantees, also for more complex parametrizations, is impossible without further restrictions.

For the discretized game from SODA, the experimental results already show that the monotonicity does not hold. From Rosen (1965, Theorem 2) we know that monotonicity implies uniqueness of the equilibrium point. Since we can observe that SODA converges to different equilibrium points, uniqueness and hence monotonicity cannot be satisfied. Moreover, we can check the monotonicity condition directly. The set of discrete distributional strategies together with the expected utilities of the discretized game (see Section 4.2) define a game as defined above. Analogous to NPGA, we then checked the inequality Eq. (F.1) for different strategies and could verify numerically that the condition does not hold. This was done for different numbers of discretization points of the game.

Appendix G. Local convergence of NPGA assuming linear strategies

This section presents the proof of Proposition 1. For this, we use a result of Chasnov et al. (2020), which is stated first. Then, we draw on the formulas for the interim utilities derived in Section Appendix C, where we derive the buyer's and seller's ex-ante utilities assuming linear equilibrium bid functions. With these, we formulate the ex-ante game explicitly and successively show all needed properties for the result to hold.

G1. Convergence of gradient-based learning

Consider a set of $\mathcal{I} = \{1, \ldots, n\}$ agents, an action space $\mathbb{R}^d = \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_n}$ (or possibly subsets thereof). Let $f_i : \mathbb{R}^d \to \mathbb{R}$ denote agent i's cost function. This corresponds to the negative utilities for participants in bilateral bargaining. Then, the collection of costs $(f_1, \ldots, f_n)$ on the action space $\mathbb{R}^d$ defines a continuous game. Let $D_i f_i$ and $D_i^2 f_i$ denote the first and second partial derivative of f_i with respect to θ_i and $D_{ji} f_i$ denote the partial derivative of $D_i f_i$ with respect to θ_j. Define the *game gradient* as

$$\omega(\theta) = (D_1 f_1(\theta), \ldots, D_n f_n(\theta)), \tag{G.1}$$

and the *game Jacobian*, i.e., the Jacobian of ω, by

$$J(\theta) = \begin{bmatrix} D_1^2 f_1(\theta) & \ldots & D_{1n} f_1(\theta) \\ \vdots & \ddots & \vdots \\ D_{n1} f_n(\theta) & \ldots & D_n^2 f_n(\theta) \end{bmatrix}. \tag{G.2}$$

We make the following assumption so that the game gradient and Jacobian exist and are well-defined.

Assumption 3. For each $i \in \mathcal{I}$, $f_i \in C^q(\mathbb{R}^d, \mathbb{R})$ for $q \geq 2$ and $\omega(\theta)$ is L-Lipschitz.

The following two definitions characterize local properties of a Nash equilibrium strategy $\theta^* \in \mathbb{R}^d$.

[5] Rosen originally referred to strict monotonicity as diagonal strict concavity.

Definition 2 (Definition 3 of Ratliff et al. (2016)). A strategy $\theta^* \in \mathbb{R}^d$ is a *differential Nash equilibrium* if $\omega(\theta^*) = 0$ and $D_i^2 f_i(\theta^*) > 0$ for each $i \in \mathcal{I}$.

Definition 3. Let $\theta^* \in \mathbb{R}^d$ be a differential Nash equilibrium. If the game Jacobian $J(\theta^*)$ is non-degenerate, i.e., $\det J(\theta^*) \neq 0$, and the spectrum of $J(\theta^*)$ is strictly in the right half-plane, i.e., $\operatorname{spec}(J(\theta^*)) \subset \mathbb{C}_+^o$), then we call θ^* a *stable differential Nash equilibrium*.

Now, we state a special case of Proposition 2 of Chasnov et al. (2020), which gives conditions on convergence to a Nash equilibrium assuming exact gradient feedback and a constant learning rate.

Proposition 2. *Consider an* n*-player game* $\mathcal{G} = (f_1, \ldots, f_n)$ *satisfying Assumption 3. Let* $\theta^* \in \mathbb{R}^d$ *be a stable differentiable Nash equilibrium with* $\mathcal{R}(\theta^*)$ *being its region of attraction. Suppose agents use the gradient-based learning rule* $\theta_{k+1} = \theta_k - \Gamma\omega(\theta_k)$ *with* $\Gamma = \gamma \cdot I_m$ *s.t.* $0 < \gamma < \tilde{\gamma}$, *where* $\tilde{\gamma} = \arg\min_{h>0} \max_j |1 - h\lambda_j(J(\theta^*))| = 1$ *and* $\lambda_j(A)$ *denotes the j'th eigenvalue of matrix A. Then, for* $\theta_0 \in \mathcal{R}(\theta^*)$, $\theta_k \to \theta^*$ *exponentially.*

G2. Proof of Proposition 1

Combining the findings up to this point, we can state the proof of Proposition 1.

Proof. We aim to use Proposition 2 to show the final result. For this, we check that Assumption 3 holds and the linear equilibrium needs to be a stable differentiable NE.

We start by showing that Assumption 3 holds. Note that the ex-ante utilities of buyer and seller from Eqs. (E.15) to (E.18) are rational functions in m_B and m_S and polynomials in t_B and t_S, where the poles are not in the feasible set as $m_B, m_S > 0$ according to Assumption 1. Therefore, these are in $\mathcal{C}^\infty$. The *game gradient* is given by

$$\omega(m_B, t_B, m_S, t_S) = \begin{pmatrix} \frac{\partial}{\partial m_B} u_B^{\text{ante}} \\ \frac{\partial}{\partial t_B} u_B^{\text{ante}} \\ \frac{\partial}{\partial m_S} u_S^{\text{ante}} \\ \frac{\partial}{\partial t_S} u_S^{\text{ante}} \end{pmatrix} \tag{G.3}$$

$$= \begin{pmatrix} \frac{3m_B(t_B-t_S)^2(t_B+t_S)-m_B^3(6m_B+9t_B-3t_S-4)+4(t_B-t_S)^3}{12m_B^3m_S} \\ -\frac{(m_B+t_B-t_S)\left(t_B-m_B-t_S+\frac{3}{2}m_B(t_B+t_S+m_B)\right)}{2m_B^2m_S} \\ -\frac{(m_B+t_B-t_S)^2\left(2(t_S-t_B-m_B)\frac{3}{2}m_S(m_B+t_B+t_S)\right)}{6m_Bm_S^3} \\ -\frac{(m_B+t_B-t_S)\left(t_S-t_B-m_B+\frac{1}{2}m_S(m_B+t_B+3t_S)\right)}{2m_Bm_S^2} \end{pmatrix}. \tag{G.4}$$

The game gradient ω is Lipschitz continuous if its derivative is bounded. Therefore, we proceed by verifying that every entry of the game Jacobian is bounded under Assumptions 1 and 2. For this, we derive the game Jacobian next, which is given by

$$J(m_B, t_B, m_S, t_S) = \begin{bmatrix} D_B^2 u_B^{\text{ante}}(m_B, t_B, m_S, t_S) & D_{B,S} u_B^{\text{ante}}(m_B, t_B, m_S, t_S) \\ D_{S,B} u_S^{\text{ante}}(m_B, t_B, m_S, t_S) & D_S^2 u_S^{\text{ante}}(m_B, t_B, m_S, t_S) \end{bmatrix}. \tag{G.5}$$

All terms are 4×4 matrices, which are given by

$$D_B^2 u_B^{\text{ante}}(m_B, t_B, m_S, t_S) = \begin{pmatrix} d_{B,B}^{1,1} & d_{B,B}^{1,2} \\ d_{B,B}^{2,1} & d_{B,B}^{2,2} \end{pmatrix},$$

where

$$d_{B,B}^{1,1} = -\frac{m_B\left(m_B^3 + (t_B - t_S)^2(t_B + t_S)\right) + 2(t_B - t_S)^3}{2m_B^4 m_S},$$

$$d_{B,B}^{1,2} = -\frac{m_B\left(3m_B^2 + (t_S - t_B)(3t_B + t_S)\right) - 4(t_B - t_S)^2}{4m_B^3 m_S},$$

$$d_{B,B}^{2,1} = -\frac{m_B\left(3m_B^2 + (t_S - t_B)(3t_B + t_S)\right) - 4(t_B - t_S)^2}{4m_B^3 m_S},$$

$$d_{B,B}^{2,2} = -\frac{t_B - t_S + \frac{3}{2}m_B(t_B - t_S + m_B)}{m_B^2 m_S}.$$

Further,

$$D_{B,S} u_B^{\text{ante}}(m_B, t_B, m_S, t_S) = \begin{pmatrix} d_{B,S}^{1,1} & d_{B,S}^{1,2} \\ d_{B,S}^{2,1} & d_{B,S}^{2,2} \end{pmatrix},$$

where

$$d_{B,S}^{1,1} = \frac{m_B\left(m_B^2(6m_B + 9t_B - 3t_S - 4) - 3(t_B + t_S)(t_B - t_S)^2\right) - 4(t_B - t_S)^3}{12m_B^3 m_S^2},$$

$$d_{B,S}^{1,2} = -\frac{-{m_B}^3 + m_B{t_B}^2 + 2m_Bt_Bt_S - 3m_B{t_S}^2 + 4{t_B}^2 - 8t_Bt_S + 4{t_S}^2}{4{m_B}^3 m_S},$$

$$d_{B,S}^{2,1} = \frac{(m_B + t_B - t_S)\left(t_B - m_B - t_S + \frac{3}{2}m_B(t_B + t_S + m_B)\right)}{2{m_B}^2{m_S}^2},$$

$$d_{B,S}^{2,2} = \frac{2t_B - 2t_S + m_Bt_B + m_Bt_S + {m_B}^2}{2{m_B}^2 m_S}.$$

Further,

$$D_{S,B} u_S^{\text{ante}}(m_B, t_B, m_S, t_S) = \begin{pmatrix} d_{S,B}^{1,1} & d_{S,B}^{1,2} \\ d_{S,B}^{2,1} & d_{S,B}^{2,2} \end{pmatrix},$$

where

$$d_{S,B}^{1,1} = \frac{4(m_B + t_B - t_S)^2(2m_B - t_B + t_S) - 3m_S(m_B + t_B - t_S)\left(2{m_B}^2 + m_Bt_B + m_Bt_S - {t_B}^2 + {t_S}^2\right)}{12m_B^2 m_S^3},$$

$$d_{S,B}^{1,2} = \frac{4(m_B + t_B - t_S)^2 - m_S(m_B + t_B - t_S)(3m_B + 3t_B + t_S)}{4m_B{m_S}^3},$$

$$d_{S,B}^{2,1} = \frac{m_B^2(2 - m_S) - 2(t_B - t_S)^2 + m_S(t_B - t_S)(t_B + 3t_S)}{4m_B^2 m_S^2},$$

$$d_{S,B}^{2,2} = \frac{m_B(2 - m_S) + 2t_B - 2t_S - m_S(t_B + t_S)}{2m_B m_S^2}.$$

Lastly,

$$D_S^2 u_S^{\text{ante}}(m_B, t_B, m_S, t_S) = \begin{pmatrix} d_{S,S}^{1,1} & d_{S,S}^{1,2} \\ d_{S,S}^{2,1} & d_{S,S}^{2,2} \end{pmatrix},$$

where

$$d_{S,S}^{1,1} = -\frac{(m_B + t_B - t_S)^3 - \frac{1}{2}m_S(m_B + t_B - t_S)^2(m_B + t_B + t_S)}{m_B{m_S}^4},$$

$$d_{S,S}^{1,2} = \frac{m_S(m_B + t_B - t_S)(m_B + t_B + 3t_S) - 4(m_B + t_B - t_S)^2}{4m_B{m_S}^3},$$

$$d_{S,S}^{2,1} = \frac{m_S(m_B + t_B - t_S)(m_B + t_B + 3t_S) - 4(m_B + t_B - t_S)^2}{4m_B m_S{}^3},$$
$$d_{S,S}^{2,2} = \frac{-m_B(m_S + 2) - 2t_B + 2t_S - m_S(t_B - 3t_S)}{2m_B m_S^2}.$$

As each entry of J is bounded under Assumption 2, we get that ω is Lipschitz continuous. Therefore, Assumption 3 is satisfied in bilateral bargaining with linear strategies.

It remains to show that $\theta^* = \left(\frac{2}{3}, \frac{1}{12}, \frac{2}{3}, \frac{1}{4}\right)$ is a stable differential Nash Equilibrium. One can readily check that

$$\omega\left(\frac{2}{3}, \frac{1}{12}, \frac{2}{3}, \frac{1}{4}\right) = 0.$$

Furthermore, the matrices

$$D_B^2 u_B^{\text{ante}}\left(\frac{2}{3}, \frac{1}{12}, \frac{2}{3}, \frac{1}{4}\right) = \begin{pmatrix} -\frac{189}{256} & -\frac{135}{128} \\ -\frac{135}{128} & -\frac{27}{16} \end{pmatrix},$$
$$D_S^2 u_S^{\text{ante}}\left(\frac{2}{3}, \frac{1}{12}, \frac{2}{3}, \frac{1}{4}\right) = \begin{pmatrix} -\frac{81}{256} & -\frac{81}{128} \\ -\frac{81}{128} & -\frac{27}{16} \end{pmatrix}$$

are negative definite. One easily verifies this using the principal minor criterion. Note that the matrices need to be negative definite instead of positive definite, as we are maximizing utilities instead of minimizing cost functions. Therefore, θ^* is a differential Nash Equilibrium. The Jacobian's determinant at θ^* satisfies

$$\det\left(J\left(\frac{2}{3}, \frac{1}{12}, \frac{2}{3}, \frac{1}{4}\right)\right) = \frac{531441}{33554432} \neq 0.$$

Finally, using a computer program (Matlab, 2020), we calculate the eigenvalues of $J(\theta^*)$, which are given by

$$\lambda(J(\theta^*)) = \begin{pmatrix} \lambda_1(J(\theta^*)) \\ \lambda_2(J(\theta^*)) \\ \lambda_3(J(\theta^*)) \\ \lambda_4(J(\theta^*)) \end{pmatrix} = \begin{pmatrix} \frac{\sigma_3 \sqrt{214}66411919^{3/4} \sigma_2 \sigma_1 \left(\sigma_{11}{}^{3/4} + \sigma_5 + \sigma_6\right)\sigma_7}{69036339115606897664} \\ \frac{\sigma_3 \sqrt{214}66411919^{3/4} \sigma_2 \sigma_1 \left(\sigma_{11}{}^{3/4} - \sigma_5 + \sigma_6\right)\sigma_7}{69036339115606897664} \\ \frac{\sigma_3 \sqrt{214}66411919^{3/4} \sigma_2 \sigma_1 \left(-\sigma_{11}{}^{3/4} + \sigma_4 + \sigma_6\right)\sigma_7}{69036339115606897664} \\ -\frac{\sigma_3 \sqrt{214}66411919^{3/4} \sigma_2 \sigma_1 \left(\sigma_{11}{}^{3/4} + \sigma_4 - \sigma_6\right)\sigma_7}{69036339115606897664} \end{pmatrix},$$

where

$$\sigma_1 = \left(29221932781 - 6048\sqrt{1402682838}\mathrm{i}\right)^{1/6},$$
$$\sigma_2 = \left(-29221932781 + 6048\sqrt{1402682838}\mathrm{i}\right)^{1/4},$$
$$\sigma_3 = (-1)^{3/4},$$
$$\sigma_4 = \sqrt{-\sigma_8 + \sigma_9 - \sigma_{10} - 9487417\sqrt{\sigma_{11}}},$$
$$\sigma_5 = \sqrt{\sigma_8 + \sigma_9 - \sigma_{10} - 9487417\sqrt{\sigma_{11}}},$$
$$\sigma_6 = 63\sqrt{3}\sigma_{12}{}^{1/6}\sigma_{11}{}^{1/4},$$
$$\sigma_7 = \left(\sigma_{12}{}^{1/3}\left(-153710549 + 72\sqrt{1402682838}\mathrm{i}\right) + 121846\sigma_{12}{}^{2/3} - 473375263673 - 319752\sqrt{1402682838}\mathrm{i}\right)^{1/4}$$
$$\sigma_8 = 161784\sqrt{87665798343 + 18144\sqrt{1402682838}\mathrm{i}},$$
$$\sigma_9 = 8882\sigma_{12}{}^{1/3}\sqrt{\sigma_{11}},$$
$$\sigma_{10} = \sigma_{12}{}^{2/3}\sqrt{\sigma_{11}},$$
$$\sigma_{11} = 4441\sigma_{12}{}^{1/3} + \sigma_{12}{}^{2/3} + 9487417,$$
$$\sigma_{12} = 29221932781 + 6048\sqrt{1402682838}\mathrm{i}.$$

One can numerically verify that all eigenvalues have a strictly negative real part. Therefore, it holds that $\operatorname{spec}(J(\theta^*)) \subset \mathbb{C}_-^o$.

That means we can use Proposition 2 to show that, if we use a sufficiently small learning rate, gradient-based algorithms, in particular, NPGA with exact gradient feedback, indeed converges to the linear equilibrium strategies, which finishes the proof. □

Remark 1. Note that the proof is conducted for the special case of $k = 1/2$ as stated in the proposition, but essentially works for any k. However, in the final step, we rely on a computer program to calculate the eigenvalues of the game Jacobian matrix, because there is no obvious way of doing so for general k. Nonetheless, we successfully conducted the proof for $k \in \{\frac{0}{10}, \frac{1}{10}, \dots, \frac{10}{10}\}$.

References

Andrade, G. P., Frongillo, R., & Piliouras, G. (2021). Learning in Matrix Games can be Arbitrarily Complex. *Proceedings of Thirty Fourth Conference on Learning Theory, 134*, 159–185. https://proceedings.mlr.press/v134/andrade21a/andrade21a.pdf.

Armantier, O., Florens, J.-P., & Richard, J.-F. (2008). Approximation of Nash equilibria in Bayesian games. *Journal of Applied Econometrics, 23*(7), 965–981. https://doi.org/10.1002/jae.1040.

Ausubel, L. M., & Baranov, O. (2020). Core-selecting auctions with incomplete information. *International Journal of Game Theory, 49*, 251–273. https://doi.org/10.1007/s00182-019-00691-3.

Bailey, J. P., & Piliouras, G. (2018). Multiplicative weights update in zero-sum games. In *Proceedings of the 2018 ACM conference on economics and computation* (pp. 321–338). ACM.

Balduzzi, D., Racaniere, S., Martens, J., Foerster, J., Tuyls, K., & Graepel, T. (2018). The mechanics of *n*-player differentiable games. In J. Dy, & A. Krause (Eds.), *Proceedings of the 35th international conference on machine learning*. In *Proceedings of machine learning research: vol. 80* (pp. 354–363). PMLR. https://proceedings.mlr.press/v80/balduzzi18a.html

Benaim, M., & Hirsch, M. W. (1999). Mixed equilibria and dynamical systems arising from fictitious play in perturbed games. *Games and Economic Behavior, 29*, 36–72. https://escholarship.org/uc/item/4qj1335f

Bichler, M., Fichtl, M., Heidekrüger, S., Kohring, N., & Sutterer, P. (2021). Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence, 3*, 687–695. https://doi.org/10.1038/s42256-021-00365-4.

Bigi, G., Castellani, M., Pappalardo, M., & Passacantando, M. (2013). Existence and solution methods for equilibria. *European Journal of Operational Research, 227*(1), 1–11.

Blumrosen, L., & Dobzinski, S. (2021). (Almost) efficient mechanisms for bilateral trading. *Games and Economic Behavior, 130*, 369–383. https://doi.org/10.1016/j.geb.2021.08.011.

Bosshard, V., Bünz, B., Lubin, B., & Seuken, S. (2017). Computing Bayes–Nash equilibria in combinatorial auctions with continuous value and action spaces. In *Ijcai* (pp. 119–127).

Bosshard, V., Bünz, B., Lubin, B., & Seuken, S. (2020). Computing Bayes–Nash equilibria in combinatorial auctions with verification. *Journal of Artificial Intelligence Research, 69*, 531–570.

Bowling, M. (2005). Convergence and no-regret in multiagent learning. In L. K. Saul, Y. Weiss, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems: vol. 17* (pp. 209–216). MIT Press.

Bowling, M., & Veloso, M. (2002). Multiagent learning using a variable learning rate. *Artificial Intelligence, 136*(2), 215–250. https://doi.org/10.1016/S0004-3702(02)00121-2.

Busoniu, L., Babuska, R., & De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 38*(2), 156–172. https://doi.org/10.1109/TSMCC.2007.913919.

Butcher, J. C. (2008). Runge–Kutta methods. In *Numerical methods for ordinary differential equations* (pp. 137–316). John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470753767.ch3.

Cavazzuti, E., Pappalardo, M., & Passacantando, M. (2002). Nash equilibria, variational inequalities, and dynamical systems. *Journal of Optimization Theory and Applications, 114*(3), 491–506. https://doi.org/10.1023/A:1016056327692.

Chasnov, B., Ratliff, L., Mazumdar, E., & Burden, S. (2020). Convergence analysis of gradient-based learning in continuous games. In *Uncertainty in artificial intelligence* (pp. 935–944). PMLR.

Chatterjee, K., & Samuelson, W. (1983). Bargaining under incomplete information. *Operations Research, 31*(5), 835–851.

Cripps, M. W., & Swinkels, J. M. (2006). Efficiency of large double auctions. *Econometrica, 74*(1), 47–92.

Federer, H. (1996). Geometric measure theory. *Classics in mathematics*. Springer Berlin Heidelberg.

Fibich, G., & Gavish, N. (2011). Numerical simulations of asymmetric first-price auctions. *Games and Economic Behavior, 73*(2), 479–495. https://doi.org/10.1016/j.geb.2011.02.010.

Fichtl, M., Oberlechner, M., & Bichler, M. (2022). Computing Bayes Nash equilibrium strategies in auction games via simultaneous online dual averaging. arXiv preprint arXiv:2208.02036

Friedman, D. (1992). *The double auction market: Institutions, theories, and evidence*. Routledge.

Fudenberg, D., & Levine, D. K. (2009). Learning and equilibrium. *Annual Review of Economics*. https://doi.org/10.1146/annurev.economics.050708.142930.

Garratt, R., & Pycia, M. (2020). Efficient bilateral trade. *Unpublished paper*. SSRN.

Gresik, T. A. (2011). The effects of statistically dependent values on equilibrium strategies of bilateral *k*-double auctions. *Games and Economic Behavior, 72*(1), 139–148.

Gresik, T. A., & Satterthwaite, M. A. (1989). The rate at which a simple market converges to efficiency as the number of traders increases: An asymptotic result for optimal trading mechanisms. *Journal of Economic Theory, 48*(1), 304–332. https://doi.org/10.1016/0022-0531(89)90128-2.

Guo, W., Jordan, M. I., & Lin, T. (2021). A variational inequality approach to Bayesian regression games. In *60th IEEE conference on decision and control (CDC)* (pp. 795–802). IEEE Press. https://doi.org/10.1109/CDC45484.2021.9683562.

Hagerty, K. M., & Rogerson, W. P. (1987). *Robust trading mechanisms* (42, pp. 94–107). Elsevier. https://econpapers.repec.org/article/eeejetheo/v_3a42_3ay_3a1987_3ai_3a1_3ap_3a94-107.htm

Hartline, J., Syrgkanis, V., & Tardos, E. (2015). No-regret learning in bayesian games. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems: vol. 28* (pp. 3061–3069). Curran Associates, Inc..

Hormann, W. (2013). *Automatic nonuniform random variate generation*. Springer.

Hubbard, T. P., & Paarsch, H. J. (2014). Chapter 2—On the numerical solution of equilibria in auction models with asymmetries within the private-values paradigm. In K. Schmedders, & K. L. Judd (Eds.), *Handbook of computational economics*. In *Handbook of computational economics: vol. 3* (pp. 37–115). Elsevier. https://doi.org/10.1016/B978-0-444-52980-0.00002-5.

Jofré, A., Rockafellar, R. T., & Wets, R. J. B. (2007). Variational inequalities and economic equilibrium. *Mathematics of Operations Research, 32*(1), 32–50.

Kadan, O. (2007). Equilibrium in the two-player, *k*-double auction with affiliated private values. *Journal of Economic Theory, 135*(1), 495–513. https://doi.org/10.1016/j.jet.2006.06.004.

Klambauer, G., Unterthiner, T., Mayr, A., & Hochreiter, S. (2017). Self-normalizing neural networks. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems: vol. 30*. Curran Associates, Inc.. https://proceedings.neurips.cc/paper/2017/file/5d44ee6f2c3f71b73125876103c8f6c4-Paper.pdf

Krishna, V. (2009). *Auction theory*. Academic press.

Leininger, W., Linhart, P. B., & Radner, R. (1989). Equilibria of the sealed-bid mechanism for bargaining with incomplete information. *Journal of Economic Theory, 48*(1), 63–106. https://doi.org/10.1016/0022-0531(89)90120-8.

Letcher, A., Balduzzi, D., Racanière, S., Martens, J., Foerster, J., Tuyls, K., & Graepel, T. (2019). Differentiable game mechanics. *Journal of Machine Learning Research, 20*(84), 1–40. http://jmlr.org/papers/v20/19-008.html

Matlab (2020). *Version 9.11.0.1809720 (R2021b)*. Natick, Massachusetts: The MathWorks Inc..

McAfee, R. P. (1992). A dominant strategy double auction. *Journal of Economic Theory, 56*(2), 434–450. https://doi.org/10.1016/0022-0531(92)90091-U.

Mertikopoulos, P., & Zhou, Z. (2019). Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming, 173*(1-2), 465–507.

Milgrom, P. R., & Weber, R. J. (1985). Distributional strategies for games with incomplete information. *Mathematics of Operations Research, 10*(4), 619–632.

Myerson, R. B., & Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory, 29*(2), 265–281. https://doi.org/10.1016/0022-0531(83)90048-0.

Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems. *Mathematical Programming, 120*(1), 221–259.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in pytorch.

Rabinovich, Z., Naroditskiy, V., Gerding, E. H., & Jennings, N. R. (2013). Computing pure Bayesian–Nash equilibria in games with finite actions and continuous types. *Artificial Intelligence, 195*, 106–139. https://doi.org/10.1016/j.artint.2012.09.007.

Radner, R., & Schotter, A. (1989). The sealed-bid mechanism: An experimental study. *Journal of Economic Theory, 48*(1), 179–220. https://doi.org/10.1016/0022-0531(89)90124-5.

Ratliff, L. J., Burden, S. A., & Sastry, S. S. (2016). On the characterization of local Nash equilibria in continuous games. *IEEE Transactions on Automatic Control, 61*(8), 2301–2307. https://doi.org/10.1109/TAC.2016.2583518.

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave *N*-person games. *Econometrica, 33*(3), 520–534. https://doi.org/10.2307/1911749.

Rustichini, A., Satterthwaite, M. A., & Williams, S. R. (1994). Convergence to efficiency in a simple market with incomplete information. *Econometrica: Journal of the Econometric Society*, 1041–1063.

Salimans, T., Ho, J., Chen, X., Sidor, S., & Sutskever, I. (2017). Evolution strategies as a scalable alternative to reinforcement learning. arXiv:1703.03864[cs, stat], http://arxiv.org/abs/1703.03864.

Sanders, J. B. T., Farmer, J. D., & Galla, T. (2018). The prevalence of chaotic dynamics in games with many players. *Scientific Reports, 8*(1), 1–13.

Satterthwaite, M. A., & Williams, S. R. (1989). Bilateral trade with the sealed bid *k*-double auction: Existence and efficiency. *Journal of Economic Theory, 48*(1), 107–133. https://doi.org/10.1016/0022-0531(89)90121-X.

Satterthwaite, M. A., & Williams, S. R. (2002). The optimality of a simple market mechanism. *Econometrica, 70*(5), 1841–1863.

Satterthwaite, M. A., Williams, S. R., & Zachariadis, K. E. (2022). Price discovery using a double auction. *Games and Economic Behavior, 131*, 57–83.

Schaefer, F., & Anandkumar, A. (2019). Competitive gradient descent. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, & R. Garnett (Eds.), *Advances in neural information processing systems: vol. 32*. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2019/file/56c51a39a7c77d8084838cc920585bd0-Paper.pdf

Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance, 16*(1), 8–37.

Williams, S. R. (1999). A characterization of efficient, Bayesian incentive compatible mechanisms. *Economic Theory, 14*(1), 155–180.

Wilson, R. (1985). Incentive efficiency of double auctions. *Econometrica, 53*(5), 1101–1115. http://www.jstor.org/stable/1911013

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the twentieth international conference on international conference on machine learning* (pp. 928–935). Washington, DC, USA: AAAI Press.

Chapter 7

First-Order Learning in Markets

Peer-Reviewed Conference Paper

Title: Enabling first-order gradient-based learning for equilibrium computation in markets.

Authors: Nils Kohring, Fabian R. Pieroth, Martin Bichler.

In: Proceedings of the 40th International Conference on Machine Learning (ICML), to appear.

Abstract: Understanding and analyzing markets is crucial, yet analytical equilibrium solutions remain largely infeasible. Recent breakthroughs in equilibrium computation rely on zeroth-order policy gradient estimation. These approaches commonly suffer from high variance and are computationally expensive. The use of fully differentiable simulators would enable more efficient gradient estimation. However, the discrete allocation of goods in economic simulations is a non-differentiable operation. This renders the first-order Monte Carlo gradient estimator inapplicable and the learning feedback systematically misleading. We propose a novel smoothing technique that creates a surrogate market game, in which first-order methods can be applied. We provide theoretical bounds on the resulting bias which justifies solving the smoothed game instead. These bounds also allow choosing the smoothing strength a priori such that the resulting estimate has low variance. Furthermore, we validate our approach via numerous empirical experiments. Our method theoretically and empirically outperforms zeroth-order methods in approximation quality and computational efficiency.

Citation: Kohring et al. (2023).

Enabling First-Order Gradient-Based Learning for Equilibrium Computation in Markets

Nils Kohring [1] **Fabian R. Pieroth** [1] **Martin Bichler** [1]

Abstract

Understanding and analyzing markets is crucial, yet analytical equilibrium solutions remain largely infeasible. Recent breakthroughs in equilibrium computation rely on zeroth-order policy gradient estimation. These approaches commonly suffer from high variance and are computationally expensive. The use of fully differentiable simulators would enable more efficient gradient estimation. However, the discrete allocation of goods in economic simulations is a non-differentiable operation. This renders the first-order Monte Carlo gradient estimator inapplicable and the learning feedback systematically misleading. We propose a novel smoothing technique that creates a surrogate market game, in which first-order methods can be applied. We provide theoretical bounds on the resulting bias which justifies solving the smoothed game instead. These bounds also allow choosing the smoothing strength a priori such that the resulting estimate has low variance. Furthermore, we validate our approach via numerous empirical experiments. Our method theoretically and empirically outperforms zeroth-order methods in approximation quality and computational efficiency.

1. Introduction

Auctions are at the center of modern economic theory. Given some private valuation of goods available for purchase, participants must place bids on the market that maximize their expected payoff while remaining unaware of the other participants' valuations. In the seminal paper (Vickrey, 1961) the foundation for most auction theory results of today was laid. It is crucial to understand the strategic behavior in various auction applications, ranging from treasury and industrial procurement auctions to spectrum sales. Depending on the circumstances and behavioral assumptions, optimal strategies may differ drastically, starting from strategies, such as understating demand (bid-shading) (Krishna, 2009) and overstating demand (overbidding) (Ott & Beck, 2013), or much more convoluted strategies. However, computing such equilibria and approximations a priori remains challenging. Analytical equilibria can only be derived under strong assumptions such as in single-item auctions or the independent private values model.

A recent approach based on policy optimization uses randomized finite difference approximations of the gradient (Bichler et al., 2021). They proposed an algorithm called *neural pseudogradient ascent* (NPGA), which parametrizes the bidding strategies using neural networks and follows the approximate gradient dynamics of the game via simultaneous gradient ascent of all agents. The gradients are computed via *evolution strategies* (ES) (Salimans et al., 2017), which smoothen the objective by adding noise in the parameter space, thereby treating the environment as a black box. Compared with the well-known REINFORCE algorithm, where the actions are perturbed, this also results in zeroth-order gradient estimates with better precision and lower variance but much higher computational cost.

Under the differentiable programming paradigm, there is a growing interest in computing gradients for numerous reinforcement learning applications that allow for first-order gradient estimates. It is possible to create a full computational graph for applications with a certain amount of structure. First-order methods have the advantage of much lower variance, which leads to faster convergence rates to local minima of non-convex objective functions (Mohamed et al., 2020). However, there are two common problems in employing first-order methods. First, most reinforcement learning environments are provided only as black boxes. This implies that there is no explicit access to the underlying state transition function and the gradient can only be estimated by repeatedly evaluating the reward function. The wide applicability of zeroth-order policy optimization, like REINFORCE and more advanced actor-critic techniques (Schulman et al., 2017), contributes to their popularity. Sec-

[1] School of Computation, Information and Technology, Technical University of Munich. Correspondence to: Nils Kohring <nils.kohring@tum.de>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

ond, in some applications, such as the training of variational autoencoders, the computational path of the derivate is blocked (i.e., repeatably applying the chain rule to calculate the gradient of the reward with respect to the parameters of the policy) because it consists of sampling a random variable, which is a non-differentiable operation (Bangaru et al., 2021).

The situation is similar in auction games. The allocation of indivisible goods causes biased gradients of first-order methods. Example 1.1 showcases this observation. It was observed that the first-order Monte Carlo gradient estimate does not converge to equilibrium and quickly causes consistent zero-bidding (Bichler et al., 2021). From a mathematical standpoint, the single-sample (ex post) utility has a discontinuity. Thus, the sample mean of its exact gradients is an inadequate estimate for its true (ex ante) utility gradient (the expected utility over all possible valuations).

Example 1.1. Consider a first-price sealed-bid (FPSB) single-item auction. Two bidders compete for a single good, where the winner pays his or her bid. The derivative of the utility with respect to the bid is zero for losing bids and minus one for winning bids after a point of discontinuity. Either the bidder loses and receives no feedback or wins and could have won with an even smaller bid.

In this study, we propose transforming multi-agent auction games such that their utility functions are sufficiently regular for applying efficient first-order gradient methods while keeping the overall gradient dynamics close to the original game. In contrast to the original allocations of indivisible items, we use *soft* allocations instead. We effectively treat the items as divisible and allocate the proportional fraction of an item to the bidders based on their reported bids. An additional adaption to the pricing rule eliminates the discontinuity at the threshold of winning and losing an object. However, this comes at the expense of introducing a bias in the utility function. For example, a losing bidder has zero utility the original auction. However, in the smoothed auction, this bidder receives a small fraction of the good (and pays a correspondingly small price), such that the gradient indicates that a higher bid would have resulted in higher utility. The feedback to bid lower when winning remains of similar magnitude. Thus, there is always appropriate feedback on the current bidding strategy in the smoothed game. Figure 1 shows the utility function and its relaxed version.

This approach is applicable widely to economic models and general auction formats, such as sequential or simultaneous sales of multiple goods, as in combinatorial auctions with item bidding. It is further independent of the number of bidders, payment rule, risk preferences of the bidders, or correlations among the bidders' valuations. We demonstrate that the choice of a smoothing parameter follows a natural trade-off. Importantly, computing equilibria in multi-agent games is not straightforward and many negative results are known (Chasnov et al., 2020; Mazumdar et al., 2020; Letcher, 2020). Therefore, changes to the game dynamics must be implemented with great caution, and we can prove that an approximate equilibrium in the smoothed game still constitutes an approximate equilibrium in the original game.

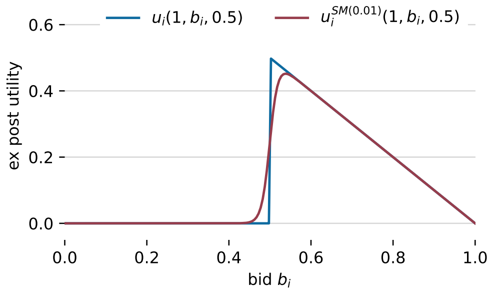


Figure 1. Ex post utility in the original FPSB single-item auction (blue) and its smoothed version (red) for a temperature of $\lambda = 0.01$ and a highest opponent bid of 0.5. The utility in the original auction is zero for losing bids and decreases linearly for winning bids.

Computational-wise, the smoothing only comes with the cost of tracking the gradients of the individual operations, upon which the game dynamics are built. Compared with NPGA, learning in the *smooth market* (SM) via the first-order estimator is more than ten times faster while yielding better results. For example, an iteration of NPGA in a small single-item auction with the default hyperparameters from (Bichler et al., 2021) takes approximately 0.16 s, whereas first-order policy gradients applied to the SM take an average of 0.01 s.

Our contribution can be summarized as follows: We introduce the SM and show that first-order methods provide an unbiased estimator of the utility gradient of the SM game. Furthermore, we provide theoretical guarantees showing that policy improvements in the SM result in improvements in the original game, and we provide theoretical and empirical insights showing that the empirical variance can be controlled. Finally, we demonstrate a substantial improvement to previous methods in performance and computational speed via multiple experiments.

2. Related Work

The theory of learning in games largely considers complete-information finite games, hence, traditional techniques rely on discretization. However, it is unclear how well a discretized strategy performs in the original continuous game in general (Waugh et al., 2009) and it suffers from the curse of dimensionality. The first attempts to compute equilibria in imperfect-information auction games followed such an approach (Athey, 2001) or expressed the game as a limit of a sequence of complete-information games (Armantier et al., 2008). In larger combinatorial auctions equilibria were first computed with an algorithm that computes pointwise best responses in a discretization of the strategy space via Monte Carlo integration (Bosshard et al., 2020). Besides the aforementioned NPGA, an approach that similarly learns continuous-action strategies was proposed (Li & Wellman, 2021). Both algorithms learn bid functions via zeroth-order gradient estimates that are used during simultaneous gradient ascent in self-play. Our method considers a continuous surrogate game and enables the use of first-order gradient methods.

The idea of analytically smoothing markets is conceptually similar to that of differentiable physics simulations. Smooth approximations of the underlying dynamics were used in these simulations (Huang et al., 2021). Zeroth- and first-order methods were compared and the pros and cons of both when available were discussed (Suh et al., 2022). Furthermore, they demonstrated that the presence of discontinuities in the objective causes the first-order estimator to be biased, whereas the zeroth-order estimator remains unbiased. Smooth markets transfer these ideas to auctions.

3. Preliminaries

We restrict the formulations to the case of single-item auctions for brevity in the presentation. The extension to *auctions of multiple independent items* is straightforward and we present some experimental results for both cases.

3.1. Auctions as Bayesian Games

A *Bayesian auction game* is defined as a quintuple $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$. $\mathcal{I} = \{1, \dots, n\}$ describes the set of bidders participating in the game. The set of possible bid profiles is given as $\mathcal{A} = \mathcal{A}_1 \times \dots \times \mathcal{A}_n$, where $\mathcal{A}_i$ is the set of bids available to agent $i \in \mathcal{I}$. Whereas $\mathcal{V} = \mathcal{V}_1 \times \dots \times \mathcal{V}_n$ is the set of *valuation profiles*. $F \colon \mathcal{V} \rightarrow [0, 1]$ defines the joint prior probability distribution over valuation profiles, which is assumed to be common knowledge among all agents and atomless. F_i denotes agent i's marginal distribution of valuations. In this study, the index *-i* denotes a profile of valuations, bids, or strategies for all bidders, except bidder i.

At the beginning of the game, nature draws a valuation profile $v \sim F$, and each agent i is informed of his or her valuation $v_i \in \mathcal{V}_i$. We denote by F_i the marginal distribution of bidder i and by $F_{\text{-}i|i}$ the conditional distribution of the opponents given v_i. Based on the drawn valuation v_i, each agent submits a bid b_i according to the *strategy*, *policy*, or *bid function* $\beta_i \colon \mathcal{V}_i \rightarrow \mathcal{A}_i$. We denote the resulting strategy space of bidder i as $\Sigma_i \subseteq \mathcal{A}_i^{\mathcal{V}_i}$ and the space of possible joint strategies as $\Sigma = \prod_i \Sigma_i$.

As part of the environment, the auctioneer collects these bids and applies an *auction mechanism* that determines allocations $x_i \in \{0, 1\}$ for each bidder i, such that the item is allocated to at most one bidder. Also, it determines payments $p(b) \in \mathbb{R}^n_{\geq 0}$ according to a payment rule p, which the agents must pay to the auctioneer. We will consider bidders with risk-neutral utility functions given by $u_i \colon \mathcal{V}_i \times \mathcal{A} \rightarrow \mathbb{R}$,

$$u_i(v_i, b) = v_i \, x_i(b) - p_i(b) \tag{1}$$

$$= \begin{cases} v_i - p_i(b) & b_i > \max b_{\text{-}i}, \\ 0 & \text{else,} \end{cases} \tag{2}$$

i.e., the players' utility is given by how much they value the good allocated to them minus the price to be paid. We will also write $u_i(v_i, b_i, b_{\text{-}i}) = u_i(v_i, b)$ with a slight abuse of notation. Thus, the bidders' utilities depend on all bidders' actions but only on their own valuations. They aim to maximize their utility u_i. We omit bidders with risk aversion or other forms of utility and valuation correlations for brevity. Notwithstanding, our treatment of equilibrium computation also extends to these settings. We will differentiate between the *ex ante* state of the game, where bidders know only the prior F, the *interim* state, where bidders additionally know their valuation $v_i \sim F_i$, and the *ex post* state, where all bids have been submitted; thus, $u_i(v_i, b)$ can be evaluated.

3.2. Equilibria

Nash equilibria (NE) are often regarded as the central solution concept in game theory. Informally, given the equilibrium strategy of the opponents in an NE, no agent has an incentive to unilaterally deviate. *Bayesian Nash equilibria* (BNE) extend this concept to games of incomplete information. Here, the expected utility over the distribution of opponent valuations is calculated instead. For a private valuation $v_i \in \mathcal{V}_i$, bid $b_i \in \mathcal{A}_i$, and opponent strategies $\beta_{\text{-}i} \in \Sigma_{\text{-}i}$, we denote the *interim utility* of bidder i as

$$\overline{u}_i(v_i, b_i, \beta_{\text{-}i}) \;=\; \mathbb{E}_{v_{\text{-}i}|v_i}[u_i(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))], \tag{3}$$

where $v_{\text{-}i}|v_i$ denotes the expectation over the opponent's conditional prior distribution given the valuation v_i. We also denote the *interim utility loss* of bid b_i incurred by not playing a best response, given v_i and $\beta_{\text{-}i}$ by:

$$\overline{\ell}_i(v_i, b_i, \beta_{\text{-}i}) \;=\; \sup_{b_i' \in \mathcal{A}_i} \overline{u}_i(v_i, b_i', \beta_{\text{-}i}) - \overline{u}_i(v_i, b_i, \beta_{\text{-}i}). \tag{4}$$

An ε-*Bayes Nash equilibrium* (ε-BNE) with $\varepsilon \geq 0$ is a strategy profile $\beta^* = (\beta_1^*, \dots, \beta_n^*) \in \Sigma$, such that no bidder can improve his or her interim expected utility more than ε

by deviating. Therefore, in an ε-BNE for all $i \in \mathcal{I}$, it holds that

$$\sup_{v_i \in \mathcal{V}_i} \bar{\ell}_i(v_i, \beta_i^*(v_i), \beta_{\text{-}i}^*) \;\leq\; \varepsilon. \tag{5}$$

A 0-BNE is simply called a BNE. In a BNE, every bidder's strategy maximizes his or her expected interim utility across his or her valuation space, given the opponents' strategies. While BNEs are often defined at the *interim* stage of the game, we also consider *ex ante* equilibria as strategy profiles that concurrently maximize each bidder's *ex ante* utility

$$\tilde{u}_i(\beta_i, \beta_{\text{-}i}) \;=\; \mathbb{E}_{v_i}[\bar{u}_i(v_i, \beta_i(v_i), \beta_{\text{-}i})]. \tag{6}$$

To estimate the worst-case interim utility loss $\bar{\ell}_{\text{max}}$, we choose an equidistant grid of n_{grid} alternative actions ranging from zero to the maximum valuation for all dimensions and calculate approximate best responses based on the average utility over a sample of n_{batch} prior distributions. Taking the maximum over all valuations and bidders then gives an estimate of $\bar{\ell}_{\text{max}}$, bounding ε for the ex ante case from above.

As a second metric, we additionally report the probability-weighted root mean squared error of the learned strategy β_i to the exact BNE strategy β_i^* for those settings where an analytical BNE is known. For a sample from the prior valuation of size n_{batch}, this approximates the L_2 distance $\|\beta_i - \beta_i^*\|_{\Sigma_i}$ of these two functions as

$$L_2(\beta_i) \;=\; \left(\frac{1}{n_{\text{batch}}} \sum_{v_i} \left(\beta_i(v_i) - \beta_i^*(v_i)\right)^2 \right)^{1/2}. \tag{7}$$

Unlike $\bar{\ell}_{\text{max}}$, this metric is much easier to compute and does not suffer the drawback that a strategy with a negatable small loss may still be arbitrarily distant from the actual BNE. However, it is only computable when an analytical BNE is available and may need multiple evaluations when there are multiple BNE.

3.3. Gradient Optimization Methods

Policy gradient methods are concerned with learning a parameterized policy β_{θ_i} that selects actions based on the current observations (Sutton & Barto, 2018). To maximize utility, bidder i updates the parameters θ_i according to gradient ascent. This process is intended to compute approximate ex ante BNEs, that is, to find mutual best responses of the bidders for all possible valuations. The exact gradient update for valuation v_i in iteration t is

$$\theta_i^t \;=\; \theta_i^{t-1} + \eta \cdot \nabla_{\theta_i^{t-1}} \bar{u}_i(v_i, \beta_{\theta_i^{t-1}}(v_i), \beta_{\theta_{\text{-}i}^{t-1}}). \tag{8}$$

This must be approximated in practice. Two common methods are zeroth- and first-order gradient approximations. The former solely relies on evaluating the objective function u_i, whereas the gradient $\nabla_{\theta_i} u_i$ can be evaluated in the latter.

As stated in the introduction, the discontinuous nature of the ex post utility function stems from the sampling of the opponents' priors and their corresponding actions. We encounter u_i from Equation 2 and its derivative (in general) is discontinuous in b_i. The observation of this inapplicability persists for all pricing regimes and behavioral assumptions that are commonly considered in auctions. Thus, an unbiased gradient estimate of the interim utility function *cannot* be derived by sampling the ex post gradient. Specifically, interchanging taking an expectation and differentiating is invalid:

$$\nabla_{\theta_i} \mathbb{E}_{v_{\text{-}i}|v_i}[u_i] \;\neq\; \mathbb{E}_{v_{\text{-}i}|v_i}[\nabla_{\theta_i} u_i]. \tag{9}$$

We supply the mathematical details in Appendix A. Therefore, the naive application of backpropagating the accumulated exact ex post gradients may not be expected to provide a meaningful estimate of the ex ante gradient. This study establishes a path towards valid first-order gradient estimates in auction games.

3.4. Zeroth-Order Approximation Methods

(Bichler et al., 2021) employed ES to circumvent the interchange of differentiation and integration. ES rely on a randomized finite difference approximation of the gradient based on perturbations in the parameter space of the neural networks which can be computed after averaging over the priors (Salimans et al., 2017). This is an alternative zeroth-order method to the REINFORCE algorithm. Unlike ES, REINFORCE relies on perturbations in the action space by using mixed strategies (typically Gaussian distributions) such that the gradient of the action probability density can be approximated. (Salimans et al., 2017) compared these estimates for RL applications and argued that the variance of the ES estimate can be significantly lower. We overload the notation for the ease of readability and write $u_i(\theta_i, v_{\text{-}i}) = u_i(v_i, \beta_{\theta_i}(v_i), \beta_{\theta_{\text{-}i}}(v_{\text{-}i}))$. For a hyperparameter $\sigma > 0$, the ES estimator can be derived from

$$\begin{aligned} &\nabla_{\theta_i} \mathbb{E}_{v_{\text{-}i}|v_i}\left[u_i(\theta_i, v_{\text{-}i})\right] \\ &\quad\approx \nabla_{\theta_i} \mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)} \mathbb{E}_{v_{\text{-}i}|v_i}\left[u_i(\theta_i + \sigma\epsilon, v_{\text{-}i})\right] && (10) \\ &\quad= \mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)} \mathbb{E}_{v_{\text{-}i}|v_i}\left[\frac{\epsilon}{\sigma} u_i(\theta_i + \sigma\epsilon, v_{\text{-}i})\right]. && (11) \end{aligned}$$

The last term can now be approximated via sampling. However, the ES gradient estimate comes at massive computational costs. It requires a large number of additional environment evaluations for the sampled population values of ϵ. Parallelization is essentially unavailable, because it would reduce the number of samples from the prior when considering a fixed amount of memory. Latter of which is the main limiting factor in getting precise estimates of the expected utility in auction games. Thus, (Bichler et al., 2021) kept a large batch size and computed the ES sequentially using a default population size of 64. Based on the variance of the estimate, (Salimans et al., 2017) argued that ES are an attractive choice if the number of episodes is large, which is not the case for single-round auctions.

4. Smoothing Single-Item Auctions

This section proposes the market-specific approach.

4.1. Allocation and Price Smoothing

The allocation of indivisible objects in auction games is typically modeled as a binary vector, with a one indicating that the item is allocated to the corresponding buyer. The set of legitimate allocations is defined as

$$\mathcal{X} = \left\{ x \in \{0,1\}^n \,\middle|\, \sum_{i=1}^{n} x_i \leq 1 \right\}. \tag{12}$$

For all commonly considered auctions, the allocations label the bids as winning or losing to maximize the auctioneer's revenue. They are calculated according to

$$x(b) = \underset{x' \in \mathcal{X}}{\arg\max} \sum_{i=1}^{n} b_i x_i'. \tag{13}$$

Typical auction mechanisms only differ in their payment rules. Two noteworthy examples are the first-price mechanism, where bidders pay what they bid and the celebrated VCG mechanism (second-price), where they pay for the harm they cause others by competing (Krishna, 2009).

These allocations result in the utilities not being continuous. Therefore, we propose relaxing the calculation of the allocations using the softmax function as a surrogate for the argmax operation:

$$x_i^{\text{SM}(\lambda)}(b) = \frac{\exp\left(\frac{b_i}{\lambda}\right)}{\sum_{j=1}^{n} \exp\left(\frac{b_j}{\lambda}\right)}, \quad i = 1, \dots n. \tag{14}$$

The temperature $\lambda > 0$ denotes the smoothing strength. This can be interpreted as dividing the item among all bidders according to their proportional bid magnitudes, where $\sum_i x_i^{\text{SM}(\lambda)}(b) = 1$ remains valid. The softmax asymptotically recovers the true argmax as λ approaches zero. As we are interested in a continuous utility surface, the discontinuity in the prices (only the winners pay) must also be considered. An obvious choice is to calculate the original prices of the good and then distribute the price according to the fractional allocations $x^{\text{SM}(\lambda)}$:

$$p^{\text{SM}}(b) = \sum_{j=1}^{n} p_j(b). \tag{15}$$

Hence, the ex post utility in the relaxed game takes the form

$$u_i^{\text{SM}(\lambda)}(v_i, b) = \left(v_i - p^{\text{SM}}(b)\right) x_i^{\text{SM}(\lambda)}(b). \tag{16}$$

By definition, we have almost everywhere (a.e.) pointwise convergence of $x_i^{\text{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}(\,\cdot\,))$ to $x_i(v_i, b_i, \beta_{\text{-}i}(\,\cdot\,))$ as functions of $v_{\text{-}i}$, except at $b_i = \max b_{\text{-}i}$. Furthermore, the fractional prices $p^{\text{SM}(\lambda)}(b_i, \beta_{\text{-}i}(\,\cdot\,))$ also converge a.e. pointwise to $p_i(b_i, \beta_{\text{-}i}(\,\cdot\,))$. Thus, the ex post utilities are recovered (a.e.) for ever smaller temperature. The resulting utilities are visualized for the special case of an FPSB auction (Figure 1). Throughout the rest of the article, we make the following regularity assumptions.

Assumption 4.1. Consider a Bayesian auction game G and assume:

1. The action $\mathcal{A}_i$ and valuation spaces $\mathcal{V}_i$ are compact intervals.
2. F is an atomless prior.
3. The bidding and pricing functions are measurable.

We regain continuity of the ex post utility and its gradient by this smoothing of allocations and payments. Specifically, we have the following theorem:

Theorem 4.2. *Let the conditions of Assumption 4.1 hold and assume the pricing function p^{SM}, the marginal density functions $\{f_{\text{-}i|i}\}_{v_i \in \mathcal{V}_i, i \in \mathcal{I}}$, and strategies $\{\beta_i\}_{i \in \mathcal{I}}$ to be Lipschitz continuous. Then, the estimator on the smooth interim utility's gradient by sampling from the smoothed ex post utilities' gradients is unbiased, i.e.,*

$$\nabla_{\theta_i} \overline{u}_i^{SM}(v_i, b_i) = \mathbb{E}_{v_{\text{-}i}|v_i}[\nabla_{\theta_i} u_i^{SM}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))], \tag{17}$$

for all $i \in \mathcal{I}$, $v_i \in \mathcal{V}_i$, and $b_i \in \mathcal{A}_i$.

We refer to Appendix A for the proof. Importantly, this relaxation technique is applicable to general markets with different payment rules, utility functions, or correlated priors. Compared with the ES gradient estimate, where the parameter space is perturbed, the SM gradient estimate perturbs the utility function. Thus, the origin of bias is different and can be controlled by σ for ES and by λ for SM.

4.2. Approximation Quality

We check the validity of the smoothing intervention by ensuring that the error to the original game dynamics can be controlled by choosing a sufficiently small value of λ. This ensures that conducting policy optimization in the smoothed game can be expected to result in policy improvements in the original game. Furthermore, this will clarify the question of an optimal choice of the temperature value.

Generally, analytically computing equilibria of the SM game is infeasible. Instead, we focus on comparing the expected interim and ex ante utilities in the original and SM game. A small error implies similar utility surfaces and gradient dynamics. Note that the ex post utilities can be quite different.

Suppose multiple bidders compete for a single commodity and bidder i has approximately the same bid magnitude as the strongest opponent. The smoothed allocation is close to one-half, whereas the true allocation is either zero or one. This would result in a significant difference in the ex post utility driven by the magnitude of the utility discontinuity in the original auction. The probability of such large errors decreases with smaller smoothing factors; however, this event cannot be completely ruled out. We verify in the following theorem, that the error in expected interim and ex ante utility approaches zero under mild assumptions on the auction format.

Theorem 4.3. *Let the conditions of Assumption 4.1 hold and suppose the payment rule p is bounded. Then, for bidder i, we have convergence in interim and ex ante utility:*

1. *Let $v_i \in \mathcal{V}_i$ and $b_i \in \mathcal{A}_i$, then*
$$\lim_{\lambda \to 0} \overline{u}_i^{SM(\lambda)}(v_i, b_i, \beta_{\text{-}i}) \;=\; \overline{u}_i(v_i, b_i, \beta_{\text{-}i}). \qquad (18)$$

2. *Further assume β_i to be measurable. Then,*
$$\lim_{\lambda \to 0} \tilde{u}_i^{SM(\lambda)}(\beta_i, \beta_{\text{-}i}) \;=\; \tilde{u}_i(\beta_i, \beta_{\text{-}i}). \qquad (19)$$

The proof is delegated to Appendix B. Theorem 4.3 ensures that for ever smaller λ, the bias in the expected utilities vanishes compared with the utilities in the original game. This implies that the smoothed gradients converge, thus justifying gradient-based learning in the perturbed game. Although Theorem 4.3 ensures convergence, it does not state how fast the error approaches zero. However, this information is crucial for practical applications. Therefore, we make the following additional assumptions on the auction format.

Assumption 4.4. For all $i \in \mathcal{I}$ assume:

1. β_i is strictly increasing and Lipschitz continuous.
2. β_i^{-1} is Lipschitz continuous.
3. There exists a uniform bound for all marginal conditional prior density functions $f_{i|\cdot}$.
4. p_i is bounded.

Note that assuming Lipschitz continuous strategies is satisfied by common function approximations, e.g., neural networks. With these stronger assumptions, we can present a worst-case convergence rate of the interim and ex ante utility errors.

Proposition 4.5. *Consider an auction with n bidders that satisfies Assumptions 4.1 and 4.4. Then, the absolute interim and ex ante utility errors are of order $\mathcal{O}(\lambda)$.*

Proof Sketch. Use substitution on the opponents' bidding strategies, followed by iterated use of Hölder's inequality. The details of the proof can be found in Appendix C. □

Note that Restrictions 1 and 4 in Assumption 4.4 are standard in the literature (Krishna, 2009). Restriction 2 is slightly stronger by demanding that strategy β_i cannot become infinitely flat (e.g., a saddle-point would not be allowed). However, this restriction can be somewhat lifted resulting in a worse convergence rate. Details on this can be found in Appendix C. Finally, Restriction 3 holds for all commonly used prior distributions, however, it rules out perfect correlation. Based on the previous result, we can characterize how a learned ε-BNE of the SM game translates to an approximate BNE the original game:

Theorem 4.6. *In an auction with n bidders that satisfies Assumptions 4.1 and 4.4, let β^* be an ex ante ε-BNE in the smoothed game with smoothing parameter λ. Then β^* is an ex ante $\varepsilon + \mathcal{O}(\lambda)$-BNE of the original game.*

The proof can be found in Appendix D. The derived bounds in the previous results consider worst-case scenarios. However, we observed that the error may be significantly lower in practice. To rationalize this observation, we compare the worst-case bound to the exact error in a restricted setting. Consider an FPSB auction with two bidders, independent uniform priors, and a linear bidding function of the second bidder, $\beta_2(v_2) = sv_2 + t$. Then, the bound derived in Proposition 4.5 translates to

$$\left|\tilde{u}_1^{\mathrm{SM}(\lambda)}(\beta_1, \beta_2) - \tilde{u}_1(\beta_1, \beta_2)\right| \;\leq\; \frac{\ln(2) + 1}{s}\lambda. \qquad (20)$$

In Figure 2, we compare this bound (for bidder 2's BNE strategy with $s = 0.5$ and $t = 0$) to the exact interim utility error, which can be derived for this restricted setting (see Appendix E). The convergence rate of the interim utilities depends on the specific prior sample v_1 and bid b_1. The ex ante utilities converge more rapidly than predicted by the worst-case bound. We conjecture that this often holds in practice, resulting in better learning behavior than suggested by Proposition 4.5.

4.3. Choosing the Smoothing Temperature

Let us consider the question of an optimal smoothing strength. There is an incentive to keep temperature values as low as possible, such that the original game dynamics are distorted as little as possible. On the other hand, one does not want to decrease λ too low, as this causes numerical problems. The magnitude of the gradient goes towards infinity at the former discontinuity as λ decreases. Therefore, with finite sample size, the first-order gradient estimate might have a high empirical variance (Suh et al., 2022).

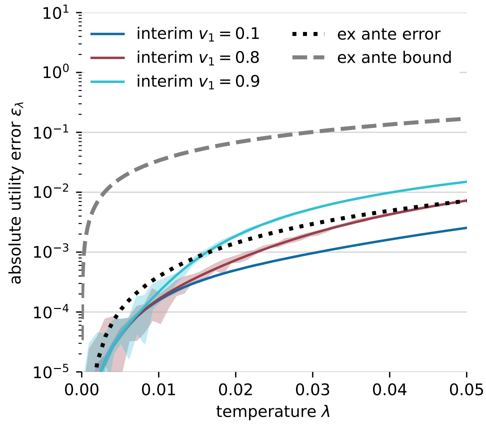


Figure 2. Comparison of the absolute utility errors. (i) The linear ex ante bound that holds for all valuations (gray dashed line) from Equation 20. (ii) Exact interim utility errors when both bidders act according to the BNE for some exemplary valuations (colorized lines) and their sampled mean values ± standard deviation (shaded areas). (iii) The approximate ex ante error (black dotted line).

We propose to use the utility sampling precision as a natural way to choose the temperature. For the special case presented in Figure 2 and the default batch size of 2^{18}, one can see that the sample precision is reached at about 10^{-4}. That is, for a drawn batch, the Monte Carlo estimation of ex ante utilities has a precision of about 10^{-4}, and we can no longer distinguish between the smoothed and original utilities. Therefore, one can use Proposition 4.5 to derive a lower bound for λ for a given sampling precision. As discussed at the end of Section 4.2, the true ex ante utility error is usually lower, so that one can choose a higher λ without losing any performance.

The empirical sampling precision is affected by several factors, such as the valuation and bidding ranges, the number of bidders, prior distributions, and complexity of bidding functions. Some of these influences can be standardized, e.g., by normalizing the bidding ranges. Ultimately, a sufficiently high batch size can overcome any bias introduced by aforementioned factors, such that it should be chosen as high as computationally possible to achieve an optimal sampling precision.

5. Empirical Results

We provide experimental evaluation of the new technique and compare the results with those of NPGA and REINFORCE by measuring how closely they approximate the analytical BNE. Results for settings with risk aversion or correlated valuations are similar and omitted for simplicity. Furthermore, we provide some insights and guidance on appropriate choices of λ and verify that our gradient estimate's variance is sufficiently small. We list all hyperparameters and details on the network architecture in Appendix G.

Table 1. Learning results in FPSB and SPSB auctions with different numbers m of items. We report the mean values of the L_2 and $\overline{\ell}_{\max}$ losses (smaller is better) and the time per iteration across five runs. We also report the standard deviation in parentheses for the losses.

	m	Algorithm	L_2	$\overline{\ell}_{\max}$	t/iter
FPSB	1	NPGA	0.011 (0.005)	0.005 (0.002)	0.155
		REINFORCE	0.021 (0.008)	**0.003 (0.000)**	**0.009**
		SM	**0.005 (0.003)**	0.004 (0.002)	**0.009**
	2	NPGA	0.013 (0.005)	0.010 (0.002)	0.150
		REINFORCE	0.041 (0.020)	0.016 (0.010)	**0.009**
		SM	**0.008 (0.002)**	**0.006 (0.003)**	**0.009**
	4	NPGA	0.028 (0.002)	0.021 (0.003)	0.148
		REINFORCE	0.064 (0.018)	0.039 (0.012)	**0.009**
		SM	**0.015 (0.004)**	**0.011 (0.004)**	**0.009**
	8	NPGA	0.104 (0.054)	0.127 (0.109)	0.206
		REINFORCE	0.187 (0.073)	0.331 (0.169)	**0.012**
		SM	**0.036 (0.003)**	**0.034 (0.009)**	**0.012**
SPSB	1	NPGA	0.012 (0.001)	0.002 (0.000)	0.170
		REINFORCE	0.028 (0.005)	0.002 (0.000)	**0.009**
		SM	**0.004 (0.001)**	**0.001 (0.000)**	0.011
	2	NPGA	0.018 (0.002)	0.003 (0.000)	0.264
		REINFORCE	0.082 (0.020)	0.009 (0.002)	0.011
		SM	**0.007 (0.001)**	**0.002 (0.000)**	**0.015**
	4	NPGA	0.043 (0.002)	0.011 (0.003)	0.457
		REINFORCE	0.140 (0.045)	0.028 (0.018)	**0.017**
		SM	**0.029 (0.003)**	**0.006 (0.002)**	0.024
	8	NPGA	0.214 (0.112)	0.299 (0.238)	0.869
		REINFORCE	0.320 (0.128)	0.262 (0.174)	**0.031**
		SM	**0.074 (0.002)**	**0.020 (0.002)**	0.043

5.1. Single-Item Auctions

For the two common payment rules of FPSB and second-price sealed-bid (SPSB) and a uniform prior on $[0, 1]$, we can measure the distance in action space to the unique BNE, as described in Equation 7 and compute an estimate of exploitability in the form of Equation 5. Table 1 shows the results. The losses are computed after training 2,000 iterations with each algorithm. The time per iteration, t/iter, decreases notably when comparing NPGA to SM across both payment rules, while also achieving a better approximation quality. Since the estimation of $\overline{\ell}_{\max}$ relies on a discretization of the action space and an exhaustive search thereon, L_2 detects smaller deviations, ceteris paribus. Although REINFORCE has a low iteration time, it is unable to learn high quality strategies due to its high variance (Section 5.3). We found that results for auctions with interdependent prior valuations or risk-aversion are quantitatively consistent with the results presented here.

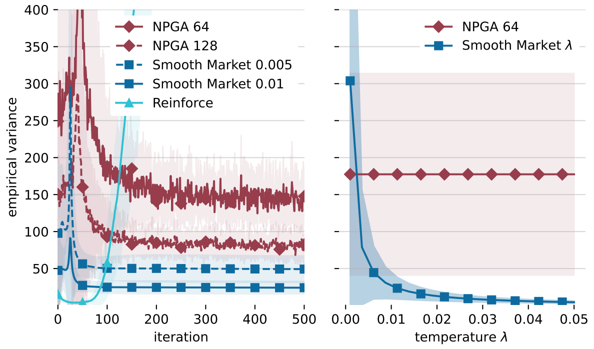


Figure 3. Empirical variance of the NPGA and REINFORCE zeroth-order and the SM first-order gradient estimates. Both zeroth-order methods are run in the original auction game. The mean values ± standard deviations over five runs each are depicted. *Left:* Comparing the variance throughout the learning procedure. *Right:* Comparing the variance for different smoothing temperatures (averaged over complete training runs).

5.2. Large Simultaneous Auctions

Furthermore, we study the separate sales of up to $m = 8$ distinctive goods and an increase in the number of bidders of up to $n = 4$. For simplicity, we do not consider any synergy effects on the items (this would include cases such as those where a bidder only values the bundle of two items but not either one of them individually), such that the BNE simplifies to the single-item strategy profile for each item separately. There are multiple motivations for these auctions. They can be considered as the base case of combinatorial auctions with item bidding and as a simple and practical alternative to full combinatorial auctions. Furthermore, combinatorial auctions with item bidding are being deployed, e.g., a bidder who is interested in a bundle of objects in parallel online display ad auctions or on a consumer shopping website is implicitly partaking in these auctions. Finally, asking a bidder to submit bids on all possible combinations of bundles $(2^m - 1)$ is practically infeasible and there are positive results on the welfare properties of limiting the action space in this way (Bhawalkar & Roughgarden, 2011). Again, we draw i.i.d. uniform valuations on $[0, 1]$ and consider the FPSB and SPSB auctions. Learning in the SM game outperforms both previous approaches (Table 1). Since first-order methods are generally faster, we assume that the strong results in these settings will scale to even larger ones.

5.3. Empirical Variance

As stated in Section 4.3, there is a trade-off between low and high values of λ. Here, we consider the base setting of two bidders competing in a single-item FPSB auction. We

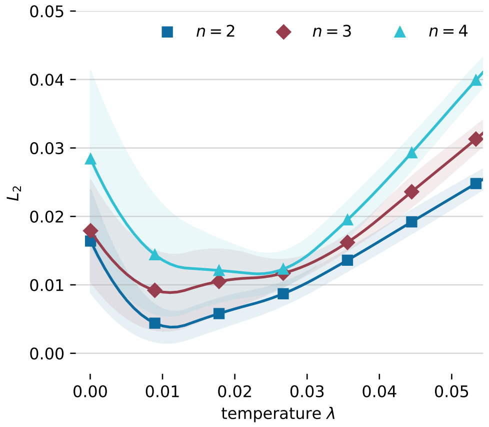


Figure 4. Action space distance for learned strategies to the BNE for different numbers of bidders and temperature values λ. The mean values ± standard deviations over five runs each are depicted.

decrease the batch size to 2^{16} as the single-sample gradients require more memory. Considering NPGA that is based on a sample of 64 evaluations of the objective by default, the empirical variance of the SM estimate is lower for all $\lambda > 0.002$ (compare intersection of Figure 3, right plot). Even after increasing NPGA's population size by a factor of two (which scales the run time in the same way), SM's variance remains lower for most choices, as can be seen in the left figure. The empirical variance of REINFORCE rapidly increases as the mixed-strategies get closer to the pure-strategy BNE. This degradation is to be expected when the learned variance of the Gaussian distributed actions decreases, see Exercise 13.4 of (Sutton & Barto, 2018).

Results for markets of different sizes are depicted in Figure 4. Keeping everything else fixed, the highest achievable performance decreases for larger markets, as is expected in multi-agent learning. The optimal smoothing strength is only affected indirectly via the bid magnitudes. At last, we note that the performance boost of larger batch sizes diminishes and best results are achieved for similar values of λ just below 0.01, indicating that the variance of the gradient estimate counteracts the lower bias. The results are presented in Appendix F.

6. Conclusion and Future Work

How can first-order gradient estimation methods be successfully applied to learning in auctions? We showed that our proposed smooth game formulation of strategic interactions in auctions provides a strong answer to this question. We established theoretical bounds on the bias caused by the smoothing, and an empirical evaluation verified that the variance of the gradient estimate can be controlled, leading to low computational costs and high precision. Overall, we verified that equilibrium computation in smooth markets via fist-order gradient estimation is more efficient than previous learning methods.

Acknowledgements

We are grateful for funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; grant no. BI 1059/I-9). We thank M. Oberlechner for valuable feedback on the research idea and S. Heidekrüger for the programming framework on top of which we have implemented our approach.

References

Armantier, O., Florens, J.-P., and Richard, J.-F. Approximation of Nash equilibria in Bayesian games. *Journal of Applied Econometrics*, 23(7):965–981, 2008.

Athey, S. Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica*, 69(4):861–889, 2001.

Bangaru, S. P., Michel, J., Mu, K., Bernstein, G., Li, T.-M., and Ragan-Kelley, J. Systematically differentiating parametric discontinuities. *ACM Transactions on Graphics (TOG)*, 40(4):1–18, 2021.

Bhawalkar, K. and Roughgarden, T. Welfare guarantees for combinatorial auctions with item bidding. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pp. 700–709. SIAM, 2011.

Bichler, M., Fichtl, M., Heidekrüger, S., Kohring, N., and Sutterer, P. Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence*, 3(8):687–695, 2021.

Bogachev, V. I. *Measure Theory*, volume 1. Springer Science & Business Media, 2007.

Bosshard, V., Bünz, B., Lubin, B., and Seuken, S. Computing bayes-nash equilibria in combinatorial auctions with verification. *Journal of Artificial Intelligence Research*, 69:531–570, 2020.

Chasnov, B., Ratliff, L., Mazumdar, E., and Burden, S. Convergence analysis of gradient-based learning in continuous games. In *Uncertainty in Artificial Intelligence*, pp. 935–944. PMLR, 2020.

Huang, Z., Hu, Y., Du, T., Zhou, S., Su, H., Tenenbaum, J. B., and Gan, C. Plasticinelab: A soft-body manipulation benchmark with differentiable physics. In *International Conference on Learning Representations*, 2021.

Katz, V. J. Change of variables in multiple integrals: Euler to cartan. *Mathematics Magazine*, 55:3–11, 1982. ISSN 0025-570X.

Krishna, V. *Auction theory*. Academic press, 2009.

Letcher, A. On the impossibility of global convergence in multi-loss optimization. In *International Conference on Learning Representations*, 2020.

Li, Z. and Wellman, M. P. Evolution strategies for approximate solution of bayesian games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 5531–5540, 2021.

Mazumdar, E., Ratliff, L. J., Jordan, M. I., and Sastry, S. S. Policy-gradient algorithms have no guarantees of convergence in linear quadratic games. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 860–868, 2020.

Mohamed, S., Rosca, M., Figurnov, M., and Mnih, A. Monte carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

Ott, M. and Beck, M. Incentives for overbidding in minimum-revenue core-selecting auctions. 2013.

Salimans, T., Ho, J., Chen, X., Sidor, S., and Sutskever, I. Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *ArXiv*, March 2017.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms. *ArXiv*, abs/1707.06347, 2017.

Suh, H. J., Simchowitz, M., Zhang, K., and Tedrake, R. Do differentiable simulators give better policy gradients? In *International Conference on Machine Learning*, pp. 20668–20696. PMLR, 2022.

Sutton, R. S. and Barto, A. G. *Reinforcement learning: An introduction*. MIT press, 2018.

Talvila, E. Necessary and sufficient conditions for differentiating under the integral sign. *The American Mathematical Monthly*, 108(6):544–548, 2001.

Vickrey, W. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961.

Waugh, K., Schnizlein, D., Bowling, M. H., and Szafron, D. Abstraction pathologies in extensive games. *AAMAS (2)*, 2009:781–8, 2009.

Chapter 8

Conclusion

Concluding this dissertation, the following sections summarize and discuss its key results in light of the literature on learning to bid and point out future research directions.

8.1 Summary and Discussion

This dissertation investigated if and how RL-based equilibrium computation can be leveraged in microeconomic models. This may help economists and policymakers to develop a better understanding of markets and make more informed decisions in the long run. Despite equilibria being computationally complex in general and convergence guarantees being limited to a few scenarios, this dissertation provides strong theoretical and empirical results on the success of utilizing RL to compute approximate equilibria.

Prior research has been mainly focused on two extremes. On the one hand, the market simulations were drastically simplified by discretization, assuming independent priors, risk neutrality, et cetera. These either allowed solving the ODE equilibrium condition (analytically or numerically) or the application of simple learning dynamics. This becomes especially contentious as many equilibria in auction theory are sensitive to the underlying assumptions. Even results from closely related markets do not usually have explanatory power. On the other hand, large-scale experiments have been conducted, e.g., for online advertisement auctions, where no general understanding of the dynamics of the applied algorithms could be established. Instead, the qualitative assessment usually compared the results to expert-designed heuristics or simply checked if the utilities increased during learning. We contributed to bridging the gap between these two extremes. Our framework allows for a multitude of different settings with variable market sizes in terms of participants and goods, ranging from single-item to larger combinatorial auctions, assumptions on the information structure of the agents, and their behavioral preferences. We provide the most extensive simulation suite of single-sided and two-sided auctions. Using the policy gradient method NPGA, we were able to approximate equilibria in auctions for which no analytical equilibrium was known previously. These include multi-unit auctions with valuation interdependencies or asymmetric bargaining markets. Our approach maintains tractability of the underlying dynamics that allows the verification of the equilibrium proximity in terms of the utility left on the table. This gives insights into the generally robust convergence of policy gradient methods. Despite these positive results, we were able to construct a non-degenerate example where current convergence guarantees from MARL do not hold. Refining the precise conditions for equilibrium convergence of learning dynamics in markets is the most pressing question for future research.

8.2 Future Work

There are a couple of promising directions for future research that we want to outline. As we have briefly discussed, it remains open if a set of reasonable assumptions exists explaining the global convergence of learning to bid that we consistently observe empirically. What is the broadest possible class of markets for which a convergent learning algorithm exists, and which algorithm is that? Do auctions satisfy some regularity condition not yet connected to equilibrium convergence? We have seen the violation of currently available convergence conditions in bilateral bargaining, i.e., concavity and monotonicity of the utility functions (in the sense of Rosen (1965) and Mertikopoulos and Zhou (2019), respectively), but were able to show local convergence in a restricted setting with linear strategies (Bichler et al., 2022, see chapter 6). Mertikopoulos (2019) suggests, in a spirit similar to Letcher et al. (2019a), to further consult the theory of dynamical systems that provides a unified way of explaining and decomposing learning dynamics and their convergence to equilibria or cycling and chaotic behavior.

From the practitioner's point of view, scalability and applicability to more realistic domains is arguably of the highest relevance. Applications include procurement auctions conducted along industrial supply chains, high-stakes spectrum auctions, or display ad auctions for which some early empirical research already exists. One stepping stone for learning strategic interaction in these scenarios is the extension of existing simulation frameworks to markets of sequential sales. Agents are then able to adapt their bids and offers depending on past and current sales and prices. Early experiments in a simple sequential auction with unit-demand bidders (Krishna, 2009, Chapter 15) show promising results for gradient-based learning.

On the technical side, it remains open if the specialized first-order gradient estimation technique can be generalized to more markets or if general zeroth-order methods are favorable. NPGA is widely applicable but comes with a high computational burden, whereas the first-order smoothing approach is currently limited to the independent sale of goods. So the question of generalizability comes up. PPO, as the state-of-the-art actor-critic approach for continuous games and control problems, may also be a viable alternative. There are results on the convergence of PPO and, more generally, on actor-critic methods (Perkins et al., 2017; Liu et al., 2019), but it remains open, which results are transferable to equilibrium computation in markets.

Already with some promising initial work, automated mechanism design may help design markets that maximize certain criteria. Dütting et al. (2014) use a neural network to model and learn the optimal auction itself. Depending on the underlying assumptions, bidders may either be considered truthful or strategic. In the latter case, they may also follow learning dynamics in an inner loop while the mechanism is learned in an outer loop. Can this hierarchical learning setup be implemented in a robust manner, and would non-interpretable pricing regimes be accepted in practice?

Another path forward is to depart from the black-box function approximation approach towards an attempt to gain insight into the underlying functional relationship between the valuations and the bids. Let us again consider the simple example of sequentially selling one unit at a time via an FPSB (under independent priors and risk neutrality). Then, the optimal strategy is linear in the valuation across all stages, and only the slope parameter changes. Some authors have suggested simply learning this factor of the linear function, but this approach is based on the expert knowledge that bidding functions must be of a linear type and, thus, it does not generalize. It fails as soon as the assumption of independent priors is relaxed, and optimal bidding strategies become non-linear. Neural network training becomes especially difficult in large markets with more decision-making options and periods. In the case of sequential sales, the action space's dimension increases with the number of stages, making it even harder to learn via function approximation. Most notably, the strategies for low-probability events are poorly learned. So this question arises if one can design a procedure to find the exact functional form of bid functions. One may hope to apply concepts from symbolic optimization or regression (D'Ascoli et al., 2022) to learn the functional form. It would allow for more interpretability and better generalization across stages. However, a naive application of these concepts would obviously struggle when no closed-form solutions exist.

Bibliography

S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

S. Athey. Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. *Econometrica*, 69(4):861–889, 2001.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th annual foundations of computer science*, pages 322–331. IEEE, 1995.

P. Bajari. Comparing competition and collusion: A numerical approach. *Economic Theory*, 18:187–205, 2001.

S. R. Balseiro and Y. Gur. Learning in repeated auctions with budgets: Regret minimization and equilibrium. *Management Science*, 65(9):3952–3968, 2019.

J. Bhandari and D. Russo. Global optimality guarantees for policy gradient methods. *Preprint*, 2019.

M. Bichler. *Market design: a linear programming approach to auctions and matching*. Cambridge University Press, 2017.

M. Bichler, M. Fichtl, S. Heidekrüger, N. Kohring, and P. Sutterer. Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence*, 3(8):687–695, 2021.

M. Bichler, N. Kohring, M. Oberlechner, and F. R. Pieroth. Learning equilibrium in bilateral bargaining games. *European Journal of Operational Research (EJOR)*, 2022.

M. Bichler, N. Kohring, and S. Heidekrüger. Learning equilibria in asymmetric auction games. *INFORMS Journal on Computing*, 35(3):523–542, 2023.

V. Bosshard, B. Bünz, B. Lubin, and S. Seuken. Computing Bayes-Nash equilibria in combinatorial auctions with verification. *Journal of Artificial Intelligence Research*, 69:531–570, 2020.

L. Bottou. Online learning and stochastic approximations. *Online Learning in Neural Networks*, 17(9):142, 1998.

G. W. Brown. Iterative solution of games by fictitious play. In *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, 1951.

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

Y. Cai and C. Papadimitriou. Simultaneous Bayesian auctions and computational complexity. In *ACM Conference on Economics and Computation*, pages 895–910, 2014.

E. Calvano, G. Calzolari, V. Denicolò, J. E. Harrington, and S. Pastorello. Protecting consumers from collusive prices due to ai. *Science*, 370(6520):1040–1042, 2020.

S. Campo, I. Perrigne, and Q. Vuong. Asymmetry in first-price auctions with affiliated private values. *Journal of Applied Econometrics*, 18(2):179–207, 2003.

N. Cesa-Bianchi and G. Lugosi. *Prediction, learning, and games*. Cambridge University Press, 2006.

B. Chasnov, L. Ratliff, E. Mazumdar, and S. Burden. Convergence analysis of gradient-based learning in continuous games. In *Proceedings of the 35th Uncertainty in Artificial Intelligence Conference (UAI)*, volume 115, pages 935–944. PMLR, 2020.

S.-F. Cheng, D. M. Reeves, Y. Vorobeychik, and M. P. Wellman. Notes on equilibria in symmetric games. In *International Workshop on Game Theoretic and Decision Theoretic Agents*, pages 71–78. Research Collection School of Computing and Information Systems, 2004.

V. Conitzer and T. Sandholm. Failures of the VCG mechanism in combinatorial auctions and exchanges. In *Proceedings of the fifth International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 521–528, 2006.

S. D'Ascoli, P.-A. Kamienny, G. Lample, and F. Charton. Deep symbolic regression for recurrence prediction. In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, volume 162, pages 4520–4536. PMLR, 2022.

C. Daskalakis and V. Syrgkanis. Learning in auctions: Regret is hard, envy is easy. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 219–228. IEEE, 2016.

C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou. The complexity of computing a Nash equilibrium. *Communications of the ACM*, 52(2):89–97, 2009.

G. Debreu. A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences*, 38(10):886–893, 1952.

P. Dütting and T. Kesselheim. Best-response dynamics in combinatorial auctions with item bidding. *Games and Economic Behavior*, 134:428–448, 2022.

P. Dütting, T. Kesselheim, and É. Tardos. Mechanism with unique learnable equilibria. In *ACM Conference on Economics and Computation*, pages 877–894, 2014.

M. Ewert, S. Heidekrüger, and M. Bichler. Approaching the overbidding puzzle in all-pay auctions: Explaining human behavior through Bayesian optimization and equilibrium learning. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1586–1588, 2022.

Z. Feng, G. Guruganesh, C. Liaw, A. Mehta, and A. Sethi. Convergence analysis of no-regret bidding algorithms in repeated auctions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 5399–5406, 2021.

G. Fibich and N. Gavish. Numerical simulations of asymmetric first-price auctions. *Games and Economic Behavior*, 73(2):479–495, 2011.

H. Flanders. Differentiation under the integral sign. *The American Mathematical Monthly*, 80(6):615–627, 1973.

L. Flokas, E. Vlatakis-Gkaragkounis, T. Lianeas, P. Mertikopoulos, and G. Piliouras. No-regret learning and mixed Nash equilibria: They do not mix. In *Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 1–24, 2020.

J. N. Foerster, R. Y. Chen, M. Al-Shedivat, S. Whiteson, P. Abbeel, and I. Mordatch. Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. International Foundation for Autonomous Agents and Multiagent Systems, 2018.

Y. Freund and R. E. Schapire. Adaptive game playing using multiplicative weights. *Games and Economic Behavior*, 29:79–103, 1999.

D. Fudenberg, F. Drew, D. K. Levine, and D. K. Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

I. Gemp, T. Anthony, J. Kramar, T. Eccles, A. Tacchetti, and Y. Bachrach. Designing all-pay auctions using deep learning and multi-agent simulation. *Scientific Reports*, 12(1):16937, 2022.

S. Ghadimi and G. Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

A. Giannou, K. Lotidis, P. Mertikopoulos, and E.-V. Vlatakis-Gkaragkounis. On the convergence of policy gradient methods to Nash equilibria in general stochastic games. In *Proceedings of the 36th International Conference on Neural Information Processing Systems (NeurIPS)*, pages 1–43, 2022.

T. Groves. Incentives in teams. *Econometrica*, 41(4):617–631, 1973.

M. Guo and V. Conitzer. Worst-case optimal redistribution of VCG payments in multi-unit auctions. *Games and Economic Behavior*, 67(1):69–98, 2009.

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

S. Hart and A. Mas-Colell. Uncoupled dynamics do not lead to Nash equilibrium. *American Economic Review*, 93(5):1830–1836, 2003.

J. Hartline, V. Syrgkanis, and E. Tardos. No-regret learning in Bayesian games. *Advances in Neural Information Processing Systems (NeurIPS)*, 28, 2015.

J. Heinrich and D. Silver. Deep reinforcement learning from self-play in imperfect-information games. *Preprint*, 2016.

J. Hofbauer and W. H. Sandholm. On the global convergence of stochastic fictitious play. *Econometrica*, 70(6):2265–2294, 2002.

C. A. Holt Jr. Competitive bidding for contracts under alternative auction procedures. *Journal of Political Economy*, 88(3):433–445, 1980.

Y.-P. Hsieh, P. Mertikopoulos, and V. Cevher. The limits of min-max optimization algorithms: Convergence to spurious non-critical sets. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139, pages 4337–4348. PMLR, 2021.

T. P. Hubbard and H. J. Paarsch. On the numerical solution of equilibria in auction models with asymmetries within the private-values paradigm. In *Handbook of Computational Economics*, volume 3, pages 37–115. Elsevier, 2014.

J. Jin, C. Song, H. Li, K. Gai, J. Wang, and W. Zhang. Real-time bidding with multi-agent reinforcement learning in display advertising. In *Proceedings of the ACM International Conference on Information and Knowledge Management*, pages 2193–2201. Association for Computing Machinery, 2018.

T. R. Kaplan and S. Zamir. Multiple equilibria in asymmetric first-price auctions. *Economic Theory Bulletin*, 3:65–77, 2015.

H. Karimi, J. Nutini, and M. Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak-łojasiewicz condition. In *Proceedings of the Machine Learning and Knowledge Discovery in Databases: European Conference (ECML PKDD)*, pages 795–811. Springer, 2016.

J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of the International Conference on Neural Networks (ICNN)*, volume 4, pages 1942–1948. IEEE, 1995.

P. Klemperer. Auction theory: A guide to the literature. *Journal of Economic Surveys*, 13 (3):227–286, 1999.

N. Kohring, C. Fröhlich, S. Heidekrüger, and M. Bichler. Equilibrium computation for auction games via multi-swarm optimization. *AAAI-22 Workshop on Reinforcement Learning in Games*, 2022.

N. Kohring, F. R. Pieroth, and M. Bichler. Enabling first-order gradient-based learning for equilibrium computation in markets. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 17327–17342. PMLR, 2023.

Y. Kolumbus and N. Nisan. Auctions between regret-minimizing agents. In *Proceedings of the ACM Web Conference 2022*, pages 100–111, 2022.

V. Krishna. *Auction theory*. Academic Press, 2009.

J. Kwon and P. Mertikopoulos. A continuous-time approach to online optimization. *Journal of Dynamics and Games*, 4(2):125–148, 2017.

M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Pérolat, D. Silver, and T. Graepel. A unified game-theoretic approach to multiagent reinforcement learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 2017.

J. D. Lee, M. Simchowitz, M. I. Jordan, and B. Recht. Gradient descent only converges to minimizers. In *Proceedings of the Conference on Learning Theory*, pages 1246–1257. PMLR, 2016.

W. Leininger, P. B. Linhart, and R. Radner. Equilibria of the sealed-bid mechanism for bargaining with incomplete information. *Journal of Economic Theory*, 48(1):63–106, 1989.

C. E. Lemke and J. T. Howson, Jr. Equilibrium points of bimatrix games. *Journal of the Society for Industrial and Applied Mathematics*, 12(2):413–423, 1964.

A. Letcher, D. Balduzzi, S. Racanière, J. Martens, J. Foerster, K. Tuyls, and T. Graepel. Differentiable game mechanics. *Journal of Machine Learning Research*, 20(84):1–40, 2019a.

A. Letcher, J. Foerster, D. Balduzzi, T. Rocktäschel, and S. Whiteson. Stable opponent shaping in differentiable games. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2019b.

Z. Li and M. P. Wellman. Evolution strategies for approximate solution of Bayesian games. In *Proceedings of AAAI Conference on Artificial Intelligence*, volume 35, pages 5531–5540, 2021.

B. Liu, Q. Cai, Z. Yang, and Z. Wang. *Neural proximal/trust region policy optimization attains globally optimal policy*. Curran Associates Inc., 2019.

W. Liu. Learning approximate Bayes-Nash equilibria with opponent-learning awareness. Master's thesis, Technical University of Munich, 2021.

R. C. Marshall, M. J. Meurer, J.-F. Richard, and W. Stromquist. Numerical analysis of asymmetric first price auctions. *Games and Economic Behavior*, 7(2):193–220, 1994.

E. Mazumdar, L. J. Ratliff, and S. S. Sastry. On gradient-based learning in continuous games. *SIAM Journal on Mathematics of Data Science*, 2(1):103–131, 2020.

P. Mertikopoulos. Online optimization and learning in games: Theory and applications. *Grenoble 1 UGA-Université Grenoble Alpes, Habilitation*, 2019.

P. Mertikopoulos and Z. Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1–2):465–507, 2019.

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

P. R. Milgrom and R. J. Weber. A theory of auctions and competitive bidding. *Econometrica: Journal of the Econometric Society*, pages 1089–1122, 1982.

K. Miyasawa. On the convergence of learning processes in a 2x2 non-zero-person game. Technical report, Princeton University NJ, 1961.

V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

S. Mohamed, M. Rosca, M. Figurnov, and A. Mnih. Monte Carlo gradient estimation in machine learning. *J. Mach. Learn. Res.*, 21(132):1–62, 2020.

D. Monderer and L. S. Shapley. Potential games. *Games and Economic Behavior*, 14(1): 124–143, 1996.

J. Nash. Non-cooperative games. *Annals of Mathematics*, 54(2), 1951.

Y. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120(1):221–259, 2009.

G. Noti and V. Syrgkanis. Bid prediction in repeated auctions with learning. In *Proceedings of the Web Conference 2021*, pages 3953–3964, 2021.

M. J. Osborne and A. Rubinstein. *A course in game theory.* MIT press, 1994.

C. Papadimitriou and G. Piliouras. Game dynamics as the meaning of a game. *ACM SIGecom Exchanges*, 16(2):53–63, 2019.

D. Parker and C. Kirkpatrick. The economic impact of regulatory policy: A literature review of quantitative evidence. *Organisation for Economic Cooperation and Development*, 2012.

S. Perkins, P. Mertikopoulos, and D. S. Leslie. Mixed-strategy learning with continuous action sets. *IEEE Transactions on Automatic Control*, 62(1):379–384, 2017.

Z. Rabinovich, E. Gerding, M. Polukarov, and N. R. Jennings. Generalised fictitious play for a continuum of anonymous players. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 245–250, 2009.

Z. Rabinovich, V. Naroditskiy, E. H. Gerding, and N. R. Jennings. Computing pure Bayesian-Nash equilibria in games with finite actions and continuous types. *Artificial Intelligence*, 195:106–139, 2013.

L. J. Ratliff, S. A. Burden, and S. S. Sastry. On the characterization of local Nash equilibria in continuous games. *IEEE Transactions on Automatic Control*, 61(8):2301–2307, 2016.

D. M. Reeves and M. P. Wellman. Computing best-response strategies in infinite games of incomplete information. In *Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI)*, 2004.

J. G. Riley and W. F. Samuelson. Optimal auctions. *The American Economic Review*, 71 (3):381–392, 1981.

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

J. B. Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pages 520–534, 1965.

T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever. Evolution strategies as a scalable alternative to reinforcement learning. *Preprint*, 2017.

J. B. Sanders, J. D. Farmer, and T. Galla. The prevalence of chaotic dynamics in games with many players. *Scientific Reports*, 8(1):4902, 2018.

J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. Proximal policy optimization algorithms. *Preprint*, 2017.

S. Shalev-Shwartz. *Online learning: Theory, algorithms, and applications*. Hebrew University, 2007.

D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller. Deterministic policy gradient algorithms. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, volume 32, pages 387–395. PMLR, 2014.

D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

J. C. Spall. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 37(3):332–341, 1992.

S. Srinivasan, M. Lanctot, V. Zambaldi, J. Pérolat, K. Tuyls, R. Munos, and M. Bowling. Actor-critic policy optimization in partially observable multiagent environments. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018.

H. J. Suh, M. Simchowitz, K. Zhang, and R. Tedrake. Do differentiable simulators give better policy gradients? In *Proceedings of the 39th International Conference on Machine Learning (ICML)*, pages 20668–20696. PMLR, 2022.

J. Tan, R. Khalili, H. Karl, and A. Hecker. Multi-agent reinforcement learning for long-term network resource allocation through auction: A V2X application. *Computer Communications*, 194:333–347, 2022.

E. Tardos and V. V. Vazirani. Basic solution concepts and computational issues. *Algorithmic Game Theory*, pages 3–28, 2007.

T. Ui. Correlated equilibrium and concave games. *International Journal of Game Theory*, 37(1):1–13, 2008.

M. Vannoni and M. Morelli. Regulation and economic growth: A ‘contingent’ relationship. *Center for Economic and Policy Research*, 2021.

W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.

J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 1944.

K. Waugh, D. Schnizlein, M. Bowling, and D. Szafron. Abstraction pathologies in extensive games. In *Proceedings of the 8th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, volume 2, pages 781–788, 2009.

J. Weed, V. Perchet, and P. Rigollet. Online learning in repeated auctions. In *Proceedings of the Conference on Learning Theory*, pages 1562–1583. PMLR, 2016.

R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Reinforcement Learning*, pages 5–32, 1992.

K. Zhang, Z. Yang, and T. Başar. *Multi-agent reinforcement learning: A selective overview of theories and algorithms*, pages 321–384. Springer International Publishing, 2021.

M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML)*, pages 928–936, 2003.

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