Learning equilibrium in bilateral bargaining games

Martin Bichler, Nils Kohring, Matthias Oberlechner, Fabian Pieroth

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

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.

Keywords: Auctions bidding, Game theory, Machine learning

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 and 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 (Leininger et al., 1989; Satterthwaite and Williams, 1989; Kadan, 2007; Gresik, 2011; 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 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 and 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 (Jofré et al., 2007; Bigi et al., 2013). 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) (**?**) 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 montonicity. 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 and 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 and 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 and 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 and Schotter (1989) experimentally analyze the properties of the average mechanism and find linear equilibrium strategies also in the lab. Furthermore, Satterthwaite and 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 and Samuelson, 1983). Additionally, Satterthwaite and 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 and Dobzinski, 2021) or randomized mechanisms (Garratt and 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 and 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 (Zinkevich, 2003; Bowling and Veloso, 2002; Bowling, 2005; Busoniu et al., 2008), more recently, motivated by the emergence of GANs, there has been a focus on (complete-information) continuous games (Mertikopoulos and Zhou, 2019; Letcher et al., 2019; Balduzzi et al., 2018; Schaefer and Anandkumar, 2019; Bailey and Piliouras, 2018). 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) (Vickrey, 1961; Krishna, 2009; Ausubel and Baranov, 2020). 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 (?). 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 $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_{n_B} \times \mathcal{A}_{n_B+1} \times \cdots \times \mathcal{A}_{n_B+n_S}$ 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 $\mathcal{V} = \mathcal{V}_1 \times \cdots \times \mathcal{V}_{n_B+n_S}$ 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., $\beta_i : \mathcal{V}_i \rightarrow \mathcal{A}_i$. The strategy profile is denoted by $\beta = (\beta_1, \dots, \beta_{n_B+n_S}) = (\beta_i, \beta_{-i})$ 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

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

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 $i \in \{1,\dots,n_B\}$ gets an item if and only if $x_i(b) = 1$. A seller $j \in \{n_B+1,\dots,n_B+n_S\}$ 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

$$u_i^{QL}(v_i, b) = \begin{cases} x_i(b) \cdot v_i - p_i(b) & \text{for } i \in \{1,\dots,n_B\}, \\ 0 & \text{else.} \end{cases} \tag{4}$$

The sellers' ex-post utilities are respectively

$$u_j^{QL}(v_j, b) = \begin{cases} p_j(b) - x_j(b) \cdot v_j & \text{for } j \in \{n_B+1,\dots,n_B+n_S\}, \\ 0 & \text{else.} \end{cases} \tag{5}$$

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 $u_i^{\mathrm{RA}} = (u_i^{\mathrm{QL}})^\rho$ 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 $b = (b_1, \dots, b_{n_B+n_S})$, the mechanism sorts them according to a natural ordering, i.e.,

$$b_1 \geq b_2 \geq \dots \geq b_{n_B} \quad \text{and} \quad b_{n_B+1} \leq b_{n_B+2} \leq \dots \leq b_{n_B+n_S}, \tag{6}$$

[1]Other common mechanisms for two-sided markets are not as strategically complex **?**Blumrosen and Dobzinski (2021) or applicable McAfee (1992).

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.*

[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.

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 and 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

$$\begin{aligned} &u_B^{\text{interim}}(v_B, \beta_B(v_B), \beta_S) \\ &= \mathbb{1}_{\{\beta_B(v_B) \geq \underline{b_S}\}} \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} (v_B - P(\beta_B(v_B), y)) f_S(\beta_S^{-1}(y))(\beta_S^{-1})'(y) dy \end{aligned} \tag{7}$$

and

$$\begin{aligned} &u_S^{\text{interim}}(v_S, \beta_B, \beta_S(v_S)) \\ &= \mathbb{1}_{\{\overline{b_B} \geq \beta_S(v_S)\}} \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} (P(x, \beta_S(v_S)) - v_S) f_B(\beta_B^{-1}(x))(\beta_B^{-1})'(x) dx, \end{aligned} \tag{8}$$

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.

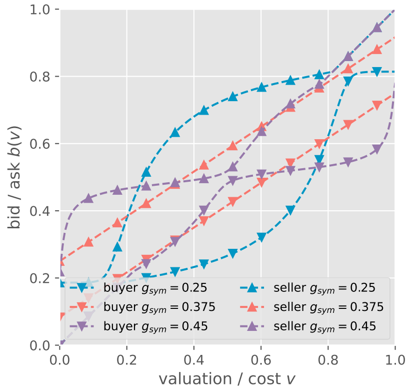
A few articles discuss the related equilibrium problem in the asymmetric independent private values model of one-sided auctions, which also results in a system of non-linear ODEs (Hubbard and Paarsch, 2014). Because the Lipschitz condition is not satisfied for the system, much of the theory concerning 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 and 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 and 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 and Williams (1989)):

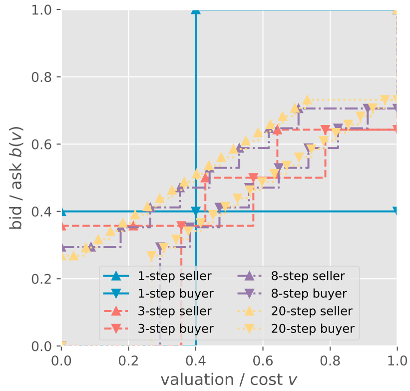
$$\beta_B(v_B; k) = \begin{cases} \frac{1}{1+k} v_B + \frac{k(1-k)}{2(1+k)}, & \text{if } v_B \in \left[\frac{1-k}{2}, 1\right], \\ h_B(v_B), & \text{else,} \end{cases} \tag{10}$$

$$\beta_S(v_S; k) = \begin{cases} \frac{1}{2-k} v_S + \frac{1-k}{2}, & \text{if } v_S \in \left[0, \frac{2-k}{2}\right], \\ h_S(v_S), & \text{else.} \end{cases} \tag{11}$$

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 and Satterthwaite, 1983).



(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.

Figure 1: Exemplary equilibrium strategies for symmetric and step function equilibria classes.

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 and Samuelson (1983) and 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 Figure 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

$$\begin{aligned} &0 < a_1 < a_2 < \cdots < a_n < 1, \\ &z_1 = a_1, \qquad z_i = a_i + \frac{x_{i-1}}{(x_i - x_{i-1})}\frac{(a_i - a_{i-1})}{2} \text{ for } i = 2, \ldots, n, \\ &x_i = a_i - \frac{(1 - z_{i+1})}{(z_{i+1} - z_i)}\frac{(a_{i+1} - a_i)}{2} \text{ for } i = 1, \ldots, n-1, \qquad x_n = a_n. \end{aligned}$$

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 $0 < a_1 < \cdots < a_n < 1$. 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

[3]Note that these equilibrium strategies do not satisfy Assumption 1.

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. Figure 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). 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 and 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 and 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. However, without knowing the parametric form of the bid function, it is impossible to study the properties of the expected utility functions.

To keep the analysis of the utilities tractable, 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 $|t_B|, |t_S| \leq K < \infty$,*
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):

$$\begin{aligned} & u_B^{\text{ante}}(m_B, t_B, m_S, t_S, k) \\ & = -\frac{1}{6{m_B}^2 m_S}(m_B + t_B - t_S)^2 \\ & \quad \cdot (t_B - t_S + m_B(m_B + t_B + 2t_S - 2) + m_B k(m_B + t_B - t_S)). \end{aligned} \tag{17}$$

Similarly, the seller's ex-ante utility is

$$\begin{aligned} & u_S^{\text{ante}}(m_B, t_B, m_S, t_S, k) \\ & = -\frac{1}{6 m_B {m_S}^2}(m_B + t_B - t_S)^3 \\ & \quad + \frac{1}{6 m_B {m_S}^2} m_S (m_B + t_B - t_S)^2 (m_B + t_B + 2t_S + k m_B + k t_B - k t_S). \end{aligned} \tag{18}$$

Figure 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 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.

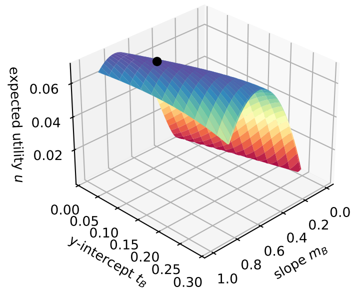


Figure 2: Ex-ante utility of the buyer under an opposing seller that plays according to the linear BNE strategy. The maximal utility on the feasible action set (black dot) is achieved by also playing the BNE strategy $\beta_B(v) = \frac{2}{3}v + \frac{1}{12}$. Note that the points are restricted to the feasible set according to Assumption 1.

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 an approximate game based on a discretized version of the 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 $p \in \{1, \dots, P\}$, 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 pseudocode 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 asymmetric, and we train two neural networks, one for the buyer and one for the seller. In larger

[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

---

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 and 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 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) (**?**) aims for distributional strategies in a discretized version of the auction game. Distributional strategies (Milgrom and 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 (**?**, 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 (**?**, 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:

$$L_2(\beta_i) \ = \ \left(\frac{1}{n_{\text{batch}}} \sum_{v_i} \left(\beta_i(v_i) - \beta^*_i(v_i)\right)^2 \right)^{\frac{1}{2}}. \tag{21}$$

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':

$$\begin{aligned} \lambda_i(v_i, b', \beta) = &\frac{1}{n_{\text{batch}}} \\ &\cdot \sum_{v_{-i}|v_i} \left(u_i\left(v_i, b', \beta_{-i}(v_{-i})\right) - u_i\left(v_i, \beta_i(v_i), \beta_{-i}(v_{-i})\right)\right). \end{aligned} \tag{24}$$

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.

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 Figure 1(a), whereas SODA converges to different step function equilibria depending on the initialization.

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 rand.**	**buyer**	0.0090 (0.0040)	0.0009 (0.0002)
		seller	0.0089 (0.0039)	0.0003 (0.0003)

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 2,000 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 sub-optimal equilibrium from the class of differentiable equilibria (see Figure 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

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)

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 (Figure 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 mechanism, the bids derived from the learned distributional strategy closely match the analytical equilibrium regardless of different initializations.

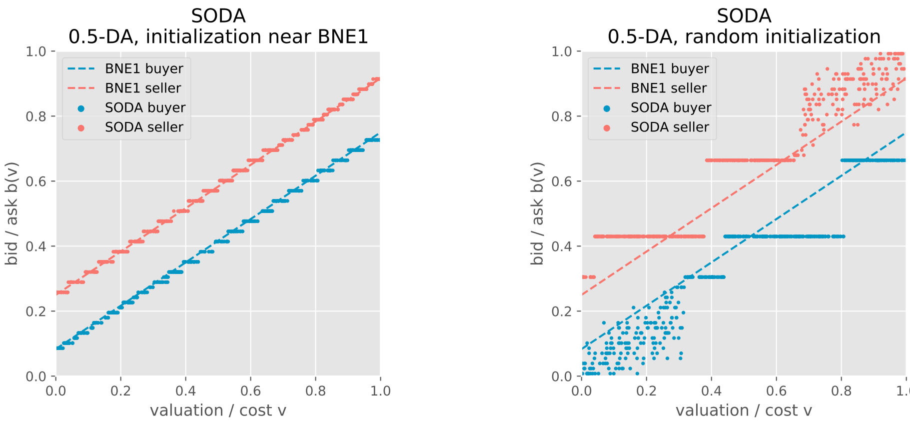


Figure 3: 500 bids sampled from the strategies computed with SODA after initialization near the linear equilibrium BNE1 (left) and after random initilization (right) for the average mechanism with uniform prior.

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.

Table 3: Mean and standard deviation over ten runs of 2,000 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)

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 and 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). Figure 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

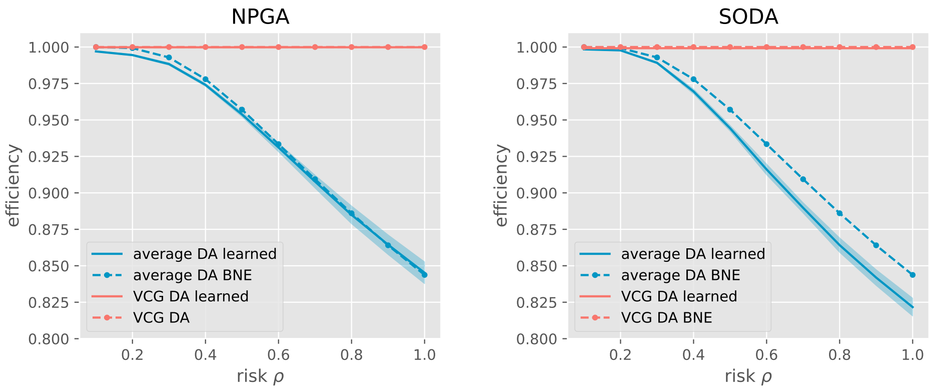


Figure 4: Mean and standard deviation of efficiency for NPGA (left) and SODA (right) applied to the average and VCG double auction with different risk parameters. Dashed lines depict efficiency in the linear BNE.

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)

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).

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 such that it is unclear, whether equilibrium computation would converge in this setting. The assumption of linear bid functions allows us to study the expected utility landscape in much more detail than would be possible in richer environments. We show that in equilibrium the utility functions are concave in large domains. 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 was known so far. This includes bilateral bargaining with Gaussian priors or risk averse traders. In the appendix, we report experiments with multiple buyers, multiple sellers, or both. This way, the paper pushes the boundaries of equilibrium computation and contributes to the understanding of equilibrium learning in the simplest and arguably most well-known model of trade.

We gratefully acknowledge financial support from the German National Science Foundation through Grant DFG BI-1057/I-9.

References

Andrade, G. P., Frongillo, R., and Piliouras, G. (2021). Learning in matrix games can be arbitrarily complex. 134:159–185.

Armantier, O., Florens, J.-P., and Richard, J.-F. (2008). Approximation of Nash equilibria in Bayesian games. *Journal of Applied Econometrics*, 23(7):965–981.

Ausubel, L. M. and Baranov, O. (2020). Core-selecting auctions with incomplete information. *International Journal of Game Theory*, 49:251–273.

Bailey, J. P. and Piliouras, G. (2018). Multiplicative weights update in zero-sum games. In *Proceedings of the 2018 ACM Conference on Economics and Computation*, pages 321–338. ACM.

Balduzzi, D., Racaniere, S., Martens, J., Foerster, J., Tuyls, K., and Graepel, T. (2018). The mechanics of n-player differentiable games. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 354–363. PMLR.

Benaim, M. and Hirsch, M. W. (1999). Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games. *Games and Economic Behavior*, 29:36–72.

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

Bigi, G., Castellani, M., Pappalardo, M., and Passacantando, M. (2013). Existence and solution methods for equilibria. *European Journal of Operational Research*, 227(1):1–11.

Blumrosen, L. and Dobzinski, S. (2021). (almost) efficient mechanisms for bilateral trading. *Games and Economic Behavior*, 130:369–383.

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

Bosshard, V., Bünz, B., Lubin, B., and 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 Saul, L. K., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*, pages 209–216. MIT Press.

Bowling, M. and Veloso, M. (2002). Multiagent learning using a variable learning rate. *Artificial Intelligence*, 136(2):215–250.

Busoniu, L., Babuska, R., and 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.

Butcher, J. C. (2008). Runge–Kutta Methods. In *Numerical Methods for Ordinary Differential Equations*, chapter 3, pages 137–316. John Wiley & Sons, Ltd.

Cavazzuti, E., Pappalardo, M., and Passacantando, M. (2002). Nash equilibria, variational inequalities, and dynamical systems. 114(3):491–506.

Chasnov, B., Ratliff, L., Mazumdar, E., and Burden, S. (2020). Convergence analysis of gradient-based learning in continuous games. In *Uncertainty in Artificial Intelligence*, pages 935–944. PMLR.

Chatterjee, K. and Samuelson, W. (1983). Bargaining under incomplete information. *Operations research*, 31(5):835–851.

Cripps, M. W. and 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. and Gavish, N. (2011). Numerical simulations of asymmetric first-price auctions. *Games and Economic Behavior*, 73(2):479–495.

Fichtl, M., Oberlechner, M., and 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. and Levine, D. K. (2009). Learning and Equilibrium. *Annual Review of Economics*.

Garratt, R. and Pycia, M. (2020). Efficient bilateral trade. *Unpublished paper, SSRN*, 2444279:365.

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. and 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.

Guo, W., Jordan, M. I., and Lin, T. (2021). A variational inequality approach to bayesian regression games. In *60th IEEE Conference on Decision and Control (CDC)*, pages 795–802. IEEE Press.

Hagerty, K. M. and Rogerson, W. P. (1987). Robust trading mechanisms. 42(1):94–107. Publisher: Elsevier.

Hartline, J., Syrgkanis, V., and Tardos, E. (2015). No-Regret Learning in Bayesian Games. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28*, pages 3061–3069. Curran Associates, Inc.

Hormann, W. (2013). *Automatic Nonuniform Random Variate Generation*. Springer.

Hubbard, T. P. and Paarsch, H. J. (2014). Chapter 2 - On the Numerical Solution of Equilibria in Auction Models with Asymmetries within the Private-Values Paradigm. In Schmedders, K. and Judd, K. L., editors, *Handbook of Computational Economics*, volume 3 of *Handbook of Computational Economics Vol. 3*, pages 37–115. Elsevier.

Jofré, A., Rockafellar, R. T., and Wets, R. J. (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.

Klambauer, G., Unterthiner, T., Mayr, A., and Hochreiter, S. (2017). Self-normalizing neural networks. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

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

Leininger, W., Linhart, P., and Radner, R. (1989). Equilibria of the sealed-bid mechanism for bargaining with incomplete information. *Journal of Economic Theory*, 48(1):63–106.

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

Matlab (2020). *version 9.11.0.1809720 (R2021b).* The MathWorks Inc., Natick, Massachusetts.

McAfee, R. (1992). A dominant strategy double auction. *Journal of Economic Theory*, 56(2):434–450.

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

Milgrom, P. R. and Weber, R. J. (1985). Distributional strategies for games with incomplete information. *Mathematics of Operations Research*, 10(4):619–632.

Myerson, R. B. and Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265–281.

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., and Lerer, A. (2017). Automatic differentiation in pytorch.

Rabinovich, Z., Naroditskiy, V., Gerding, E. H., and Jennings, N. R. (2013). Computing pure Bayesian-Nash equilibria in games with finite actions and continuous types. *Artificial Intelligence*, 195:106–139.

Radner, R. and Schotter, A. (1989). The sealed-bid mechanism: An experimental study. *Journal of Economic Theory*, 48(1):179–220.

Ratliff, L. J., Burden, S. A., and Sastry, S. S. (2016). On the Characterization of Local Nash Equilibria in Continuous Games. *IEEE Transactions on Automatic Control*, 61(8):2301–2307.

Rosen, J. B. (1965). Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica: Journal of the Econometric Society*, pages 520–534.

Rustichini, A., Satterthwaite, M. A., and Williams, S. R. (1994). Convergence to efficiency in a simple market with incomplete information. *Econometrica: Journal of the Econometric Society*, pages 1041–1063.

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

Sanders, J. B., Farmer, J. D., and Galla, T. (2018). The prevalence of chaotic dynamics in games with many players. *Scientific Reports*, 8(1):1–13.

Satterthwaite, M. A. and Williams, S. R. (1989). Bilateral trade with the sealed bid k-double auction: Existence and efficiency. *Journal of Economic Theory*, 48(1):107–133.

Satterthwaite, M. A. and Williams, S. R. (2002). The optimality of a simple market mechanism. *Econometrica*, 70(5):1841–1863.

Satterthwaite, M. A., Williams, S. R., and Zachariadis, K. E. (2022). Price discovery using a double auction. *Games and Economic Behavior*, 131:57–83.

Schaefer, F. and Anandkumar, A. (2019). Competitive gradient descent. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

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.

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, pages 928–935, Washington, DC, USA. AAAI Press.

Appendix A. Reproducibility and Hyperparameters

All our experiments are run with the following learning parameters, if not specified otherwise.

Appendix A.1. 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 11GB 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 A.5 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.

Table A.5: 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	–

Appendix A.2. 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 2,000 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 Ghz and 16GB 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 A.5 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 Wilson (1985); Rustichini et al. (1994). Overall, the inefficiency in a k-double auction decreases for symmetrically growing markets (Cripps and 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.

Appendix B.1. 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 B.6 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. Figure B.5 illustrates this observation for the case of four buyers and one seller. The buyers' strategy is very close to being truthful (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.

Table B.6: Mean and standard deviation over ten runs of 2,000 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)

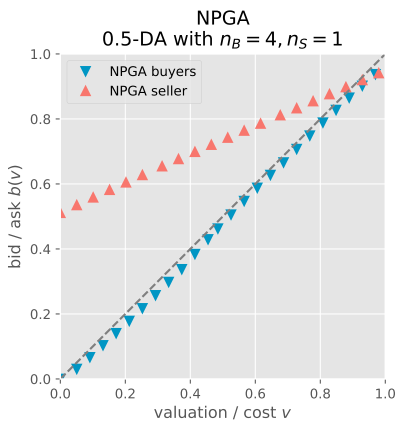


Figure B.5: The strategies of four buyers and one seller after 2,000 iterations with NPGA for a uniform prior in the average mechanism.

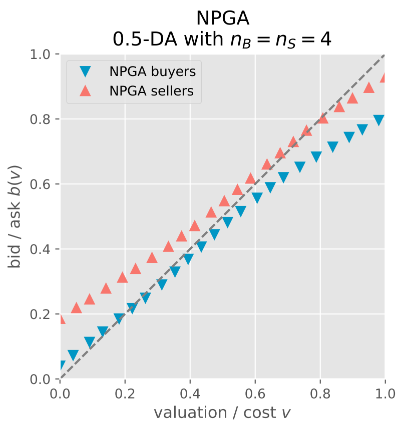


Figure B.6: The strategies of buyers and seller after 4,000 iterations with NPGA for a uniform prior with four buyers and sellers in the average-auction.

Appendix B.2. 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 and 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.

Figure B.6 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 B.7. The distance to truthful strategies is decreasing with an increasing number of buyers and sellers.

Table B.7: 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)

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 $\beta_B(v_B) \geq \underline{b_S}$:

$$
\begin{aligned}
&\mathbb{E}_{v_S \sim f_S}\left[u_B(v_B, b_B, \beta_S(v_S))\right] \\
&= \int_{\Omega_S} u_B(v_B, \beta_B(v_B), \beta_S(v_S)) \cdot f_S(v_S) dv_S \\
&= \int_{\beta_S^{-1}(\hat{\Omega}_S)} u_B(v_B, \beta_B(v_B), \beta_S(v_S)) \cdot f_S(v_S) dv_S \\
&\overset{(*_1)}{=} \int_{\hat{\Omega}_S} u_B(v_b, \beta_B(v_B), y) \cdot f_S(\beta_S^{-1}(y)) \cdot |(\beta_S^{-1})'(y)| dy \\
&\overset{\text{prop. 1}}{=} \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} (v_B - P(\beta_B(v_B), y)) \cdot f_S(\beta_S^{-1}(y)) \cdot (\beta_S^{-1})'(y) dy \\
&\overset{\text{PI}}{=} \left[(v_B - P(\beta_B(v_B), y)) \cdot F_S(\beta_S^{-1}(y))\right]_{y=\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} \\
&\quad + \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} \frac{d}{dy} P(\beta_B(v_B), y) \cdot F_S(\beta_S^{-1}(y)) dy \\
&\overset{(*_2)}{=} (v_B - P(\beta_B(v_B), \beta_B(v_B))) \cdot F_S\left(\beta_S^{-1}\left(\min\{\beta_B(v_B), \overline{b_S}\}\right)\right) \\
&\quad + \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} \frac{d}{dy} P(\beta_B(v_B), y) \cdot F_S(\beta_S^{-1}(y)) dy.
\end{aligned}
$$

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 $\beta_S(v_S) \leq \overline{b_B}$.

$$
\begin{aligned}
&\mathbb{E}_{v_B \sim f_B}\left[u_S(v_S, \beta_B(v_B), b_S)\right] \\
&= \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} (P(x, \beta_S(v_S)) - v_S) \cdot f_B(\beta_B^{-1}(x)) \cdot (\beta_B^{-1})'(x) dx \\
&\overset{\text{PI}}{=} \left[(P(x, \beta_S(v_S)) - v_S) \cdot F_B(\beta_B^{-1}(x))\right]_{x=\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} \\
&\quad - \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} \frac{d}{dx} P(x, \beta_S(v_S)) \cdot F_B(\beta_B^{-1}(x)) dx \\
&= \left(P(\overline{b_B}, \beta_S(v_S)) - v_S\right) \\
&\quad - \left(P\left(\max\{\beta_S(v_S), \underline{b_B}\}, \beta_S(v_S)\right) - v_S\right) \cdot F_B\left(\beta_B^{-1}\left(\max\{\beta_S(v_S), \underline{b_B}\}\right)\right)
\end{aligned}
$$

$$- \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} \frac{d}{dx} P(x, \beta_S(v_S)) \cdot F_B(\beta_B^{-1}(x)) dx.$$

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

$$\begin{aligned} & u_B^{\text{interim}}(v_B, \beta_B(v_B), \beta_S) \\ & = \mathbb{1}_{\{\beta_B(v_B) \geq \underline{b_S}\}} \cdot \Bigg((v_B - \beta_B(v_B)) \cdot F_S(\beta_S^{-1}(\min\{\beta_B(v_B), \overline{b_S}\})) \\ & \quad + (1-k) \cdot \int_{\underline{b_S}}^{\min\{\beta_B(v_B), \overline{b_S}\}} F_S(\beta_S^{-1}(y)) dy \Bigg), \end{aligned} \tag{C.1}$$

and the seller's interim utility by

$$\begin{aligned} & u_S^{\text{interim}}(v_S, \beta_B, \beta_S(v_S)) \\ & = \mathbb{1}_{\{\overline{b_B} \geq \beta_S(v_S)\}} \cdot \Bigg(\left(k \cdot \overline{b_B} + (1-k) \cdot \beta_S(v_S) - v_S\right) \cdot F_B(\beta_B^{-1}(\overline{b_B})) \\ & \quad - \left(k \max\left(\beta_S(v_S), \underline{b_B}\right) + (1-k)\beta_S(v_S) - v_S\right) \cdot F_B(\beta_B^{-1}(\max\{\beta_S(v_S), \underline{b_B}\})) \\ & \quad - k \cdot \int_{\max\{\beta_S(v_S), \underline{b_B}\}}^{\overline{b_B}} F_B(\beta_B^{-1}(x)) dx \Bigg). \end{aligned} \tag{C.2}$$

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. Figures D.7 through D.10 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

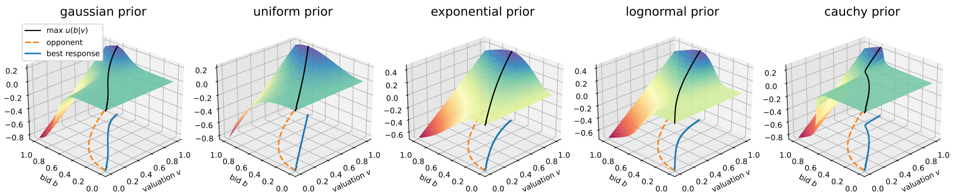


Figure D.10: Utility of the buyer under an opposing seller that plays according to some concave strategy: $\beta(v) = \sqrt{v}$.

Using Eq. (2) and Eq. (7), the buyer's ex-ante utility is given by

$$\begin{aligned} u_B^{\text{ante}}(\beta_B, \beta_S, k) &= u_B^{\text{ante}}(m_B, t_B, m_S, t_S, k) && \text{(E.1)} \\ &= \mathbb{E}_{v_B \sim f_B}\left[u_B^{\text{interim}}(v_B, m_B v_B + t_B, (m_S, t_S), k)\right] && \text{(E.2)} \\ &= \int_{\frac{1}{m_B}(t_S - t_B)}^{1} u_B^{\text{interim}}(v_B, m_B v_B + t_B, (m_S, t_S), k) dv_B. && \text{(E.3)} \end{aligned}$$

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 $\beta_B(v_B) = m_B v_B + t_B < \underline{b_S} = t_S$. 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

$$\begin{aligned} &u_B^{\text{interim}}(v_B, m_B v_B + t_B, (m_S, t_S), k) \\ &= (v_B - \beta_B(v_B)) \cdot F_S(\beta_S^{-1}(\beta_B(v_B))) + (1-k) \cdot \int_{\underline{b_S}}^{\beta_B(v_B)} F_S(\beta_S^{-1}(y)) dy && \text{(E.4)} \\ &= (v_B - m_B v_B - t_B) \cdot F_S(\beta_S^{-1}(m_B v_B + t_B)) && \text{(E.5)} \\ &+ (1-k) \int_{t_S}^{m_B v_B + t_B} F_S\left(\frac{1}{m_S}(y - t_S)\right) dy \\ &= (v_B - m_B v_B - t_B) \cdot F_S\left(\frac{1}{m_S}(m_B v_B + t_B - t_S))\right) && \text{(E.6)} \\ &+ (1-k) \int_{t_S}^{m_B v_B + t_B} F_S\left(\frac{1}{m_S}(y - t_S)\right) dy \\ &\overset{(*1)}{=} (v_B - m_B v_B - t_B) \cdot \frac{1}{m_S}(m_B v_B + t_B - t_S) && \text{(E.7)} \\ &+ (1-k) \int_{t_S}^{m_B v_B + t_B} \frac{1}{m_S}(y - t_S) dy \\ &= \frac{1}{m_S} \cdot \left[(v_B - m_B v_B - t_B)(m_B v_B + t_B - t_S) + (1-k)\left[\frac{1}{2}y^2 - t_S y\right]_{y=t_S}^{m_B v_B + t_B}\right] && \text{(E.8)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{m_S}\left[-m_B^2 v_B^2 - 2m_B t_B v_B + m_B v_B^2 + t_S m_B v_B - t_B^2 + t_B v_B + t_S t_B - t_S v_B\right] \\
&\quad + \frac{1-k}{2m_S}\left[m_B^2 v_B^2 + 2m_B t_B v_B - 2m_B t_S v_B + t_B^2 - 2t_B t_S + t_S^2\right] && \text{(E.9)} \\
&= \frac{1}{m_S}\left[\left(-m_B^2 + m_B\right) v_B^2 + \left(-2m_B t_B + t_S m_B + t_B - t_S\right) v_B - t_B^2 + t_S t_B\right] \\
&\quad + \frac{1-k}{2m_S}\left[m_B^2 v_B^2 + \left(2m_B t_B - 2m_B t_S\right) v_B + \left(t_B - t_S\right)^2\right]. && \text{(E.10)}
\end{aligned}
$$

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

$$
\begin{aligned}
&u_B^{\text{ante}}(m_B, t_B, m_S, t_S, k) \\
&= \left(\frac{m_B - m_B^2}{m_S} + \frac{1-k}{2m_S} m_B^2\right) \cdot \int_{\frac{1}{m_B}(t_S - t_B)}^{1} v_B^2 dv_B \\
&\quad + \left(\frac{t_S m_B - 2m_B t_B + t_B - t_S}{m_S} + \frac{1-k}{2m_S}\left(2m_B t_B - 2m_B t_S\right)\right) \cdot \int_{\frac{1}{m_B}(t_S - t_B)}^{1} v_B dv_B \\
&\quad + \left(\frac{t_S t_B - t_B^2}{m_S} + \frac{(1-k)\left(t_B - t_S\right)^2}{2m_S}\right)\left(1 - \frac{t_S - t_B}{m_B}\right) && \text{(E.11)} \\
&= \left(-\frac{m_B(m_B + k m_B - 2)}{2m_S}\right)\left(\frac{\left(t_B - t_S\right)^3}{3m_B^3} + \frac{1}{3}\right) \\
&\quad + \left(-\frac{t_S - t_B + m_B t_B + k m_B t_B - k m_B t_S}{m_S}\right)\left(\frac{1}{2} - \frac{\left(t_S - t_B\right)^2}{2m_B^2}\right) \\
&\quad + \left(\frac{t_S t_B - t_B^2}{m_S} + \frac{(1-k)\left(t_B - t_S\right)^2}{2m_S}\right)\left(1 - \frac{t_S - t_B}{m_B}\right). && \text{(E.12)}
\end{aligned}
$$

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.

$$
\begin{aligned}
&= \left(-\frac{1}{2m_S}\right) \cdot \left(\frac{1}{3m_B^2}\right)\left(m_B + k m_B - 2\right)\left(\left(t_B - t_S\right)^3 + m_B^3\right) \\
&\quad + \left(-\frac{1}{2m_S}\right) \cdot \left(\frac{1}{3m_B^2}\right) \cdot 3\left(t_S - t_B + m_B t_B + k m_B t_B - k m_B t_S\right)\left(m_B^2 - \left(t_S - t_B\right)^2\right) \\
&\quad + \left(-\frac{1}{2m_S}\right) \cdot \left(\frac{1}{3m_B^2}\right) \cdot 3m_B\left(2t_S t_B - 2t_B^2 + (1-k)(t_B - t_S)^2\right)\left(t_S - t_B - m_B\right) \\
& && \text{(E.13)} \\
&= -\frac{1}{6m_S m_B^2}\left(k m_B^4 + m_B t_B^3 - m_B t_S^3 - 6t_B t_S^2 + 6t_B^2 t_S - 2m_B^3 + m_B^4 - 2t_B^3 + 2t_S^3\right) \\
&\quad - \frac{1}{6m_S m_B^2}\left(k m_B t_B^3 - k m_B t_S^3 + 3m_B t_B t_S^2 - 3m_B t_B^2 t_S + 3k m_B t_B t_S^2 - 3k m_B t_B^2 t_S\right) \\
&\quad - \frac{1}{6m_S m_B^2}\left(3m_B^3 t_B - 3m_B t_B^3 - 3m_B^2 t_B + 3m_B^2 t_S + 9t_B t_S^2 - 9t_B^2 t_S + 3t_B^3 - 3t_S^3 - 3k m_B t_B^3\right) \\
&\quad - \frac{1}{6m_S m_B^2}\left(3k m_B^3 t_B + 3k m_B t_S^3 - 3k m_B^3 t_S - 3m_B t_B t_S^2 + 6m_B t_B^2 t_S - 9k m_B t_B t_S^2 + 9k m_B t_B^2 t_S\right)
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{1}{6m_S m_B^2}\left(3m_B^2t_B^2 - 3m_B^2t_S^2 + 3m_Bt_B^3 + 3m_Bt_S^3 + 3km_Bt_B^3 - 3km_Bt_S^3 - 3m_Bt_Bt_S^2\right)\\
&-\frac{1}{6m_S m_B^2}\left(-3m_Bt_B^2t_S + 3km_B^2t_B^2 + 3km_B^2t_S^2 + 9km_Bt_Bt_S^2 - 9km_Bt_B^2t_S - 6km_B^2t_Bt_S\right)
\end{aligned} \tag{E.14}
$$

$$
= -\frac{(m_B + t_B - t_S)^2\big(t_B - t_S + m_B\,(m_B + t_B + 2t_S - 2) + m_Bk\,(m_B + t_B - t_S)\big)}{6m_B^2 m_S}. \tag{E.15}
$$

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 Figure 2, which shows the local maximum at that point.

We repeat this process for the seller's ex-ante utility.

$$
\begin{aligned}
&u_S^{\text{ante}}(m_B, t_B, m_S, t_S, k)\\
&= \mathbb{E}_{v_S\sim f_S}\left[u_S^{\text{interim}}(v_S, (m_B, t_B), m_Sv_S + t_S, k)\right] && \text{(E.16)}\\
&= \int_0^{\frac{1}{m_S}(m_B+t_B-t_S)} u_S^{\text{interim}}(v_S, (m_B, t_B), m_Sv_S + t_S, k)dv_S && \text{(E.17)}\\
&= -\frac{(m_B + t_B - t_S)^3 - m_S\,(m_B + t_B - t_S)^2\,(m_B + t_B + 2t_S + km_B + kt_B - kt_S)}{6m_Bm_S{}^2}. && \text{(E.18)}
\end{aligned}
$$

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 (Mertikopoulos and 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 (**?**). 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.

[5] Rosen originally referred to strict monotonicity as diagonal strict concavity.

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 Eq. (E.15) and Eq. (18) in Appendix E, we can derive the payoff gradients

$$\begin{aligned} v_{\mathrm{ls}}(m_B, t_B, m_S, t_S) &= \begin{pmatrix} \frac{d}{dm_B} u_B^{\mathrm{ante}} \\ \frac{d}{dt_B} u_B^{\mathrm{ante}} \\ \frac{d}{dm_S} u_S^{\mathrm{ante}} \\ \frac{d}{dt_S} u_S^{\mathrm{ante}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{(t_B - t_S)^3}{3\, m_B{}^3\, m_S} - \frac{6\, m_B + 9\, t_B - 3\, t_S - 4}{12\, m_S} + \frac{(t_B + t_S)\,(t_B - t_S)^2}{4\, m_B{}^2\, m_S} \\ -\frac{(m_B + t_B - t_S)\left(t_B - m_B - t_S + \frac{3\, m_B\, t_B}{2} + \frac{m_B\, t_S}{2} + \frac{3\, m_B{}^2}{2}\right)}{2\, m_B{}^2\, m_S} \\ \frac{(m_B + t_B - t_S)^3}{3\, m_B\, m_S{}^3} - \frac{(m_B + t_B - t_S)^2\,(m_B + t_B + t_S)}{4\, m_B\, m_S{}^2} \\ \frac{(m_B + t_B - t_S)^2 - m_S\,(m_B + t_B - t_S)\left(\frac{m_B}{2} + \frac{t_B}{2} + \frac{3\, t_S}{2}\right)}{2 m_B\, m_S{}^2} \end{pmatrix}. \end{aligned}$$

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

$$\langle v_{\mathrm{ls}}(y') - v_{\mathrm{ls}}(y), y' - y \rangle = \frac{3{,}631}{3{,}000{,}000{,}000} > 0.$$

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.

Appendix G.1. Convergence of Gradient-based Learning

Consider a set of $\mathcal{I} = \{1, \dots, n\}$ agents, an action space $\mathbb{R}^d = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_n}$ (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 $(f_1, \dots, f_n)$ 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

$$\omega(\theta) = (D_1 f_1(\theta), \dots, D_n f_n(\theta)), \tag{G.1}$$

and the *game Jacobian*, i.e., the Jacobian of ω, by

$$J(\theta) = \begin{bmatrix} D_1^2 f_1(\theta) & \cdots & D_{1n} f_1(\theta) \\ \vdots & \ddots & \vdots \\ D_{n1} f_n(\theta) & \cdots & D_n^2 f_n(\theta) \end{bmatrix}. \tag{G.2}$$

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 $\theta^* \in \mathbb{R}^d$.

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., spec$(J(\theta^*)) \subset \mathbb{C}_+^\circ$), 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 $\mathcal{G} = (f_1, \dots, f_n)$ 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.*

Appendix G.2. 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 Eq. (E.15) and Eq. (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

$$\begin{aligned}
&\omega(m_B, t_B, m_S, t_S) \\
&= \begin{pmatrix} \frac{\partial}{\partial m_B} u_B^{\text{ante}} \\ \frac{\partial}{\partial t_B} u_B^{\text{ante}} \\ \frac{\partial}{\partial m_S} u_S^{\text{ante}} \\ \frac{\partial}{\partial t_S} u_S^{\text{ante}} \end{pmatrix} && \text{(G.3)} \\
&= \begin{pmatrix} \frac{3m_B(t_B-t_S)^2(t_B+t_S)-m_B^3(6m_B+9t_B-3t_S-4)+4(t_B-t_S)^3}{12m_B^3 m_S} \\ -\frac{(m_B+t_B-t_S)\left(t_B-m_B-t_S+\frac{3}{2}m_B(t_B+t_S+m_B)\right)}{2m_B^2 m_S} \\ -\frac{(m_B+t_B-t_S)^2\left(2(t_S-t_B-m_B)\frac{3}{2}m_S(m_B+t_B+t_S)\right)}{6m_B m_S^3} \\ -\frac{(m_B+t_B-t_S)\left(t_S-t_B-m_B+\frac{1}{2}m_S(m_B+t_B+3t_S)\right)}{2m_B m_S^2} \end{pmatrix}. && \text{(G.4)}
\end{aligned}$$

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

$$\begin{aligned}
&J(m_B, t_B, m_S, t_S) && \text{(G.5)} \\
&= \begin{bmatrix} D_B^2 u_B^{\text{ante}}(m_B, t_B, m_S, t_S) & D_{B,S} u_B^{\text{ante}}(m_B, t_B, m_S, t_S) \\ D_{S,B} u_S^{\text{ante}}(m_B, t_B, m_S, t_S) & D_S^2 u_S^{\text{ante}}(m_B, t_B, m_S, t_S) \end{bmatrix}.
\end{aligned}$$

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

$$d_{B,B}^{1,1} = -\frac{m_B\left(m_B^3 + (t_B - t_S)^2 (t_B + t_S)\right) + 2(t_B - t_S)^3}{2m_B^4 m_S},$$

$$
\begin{aligned}
d_{B,B}^{1,2} &= -\frac{m_B\left(3m_B^2+(t_S-t_B)(3t_B+t_S)\right)-4(t_B-t_S)^2}{4m_B^3 m_S},\\
d_{B,B}^{2,1} &= -\frac{m_B\left(3m_B^2+(t_S-t_B)(3t_B+t_S)\right)-4(t_B-t_S)^2}{4m_B^3 m_S},\\
d_{B,B}^{2,2} &= -\frac{t_B-t_S+\frac{3}{2}m_B(t_B-t_S+m_B)}{m_B^2 m_S}.
\end{aligned}
$$

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

$$
\begin{aligned}
d_{B,S}^{1,1} &= \frac{m_B\left(m_B^2(6m_B+9t_B-3t_S-4)-3(t_B+t_S)(t_B-t_S)^2\right)-4(t_B-t_S)^3}{12m_B^3 m_S^2},\\
d_{B,S}^{1,2} &= -\frac{-m_B{}^3+m_Bt_B{}^2+2m_Bt_Bt_S-3m_Bt_S{}^2+4t_B{}^2-8t_Bt_S+4t_S{}^2}{4m_B{}^3m_S},\\
d_{B,S}^{2,1} &= \frac{(m_B+t_B-t_S)\left(t_B-m_B-t_S+\frac{3}{2}m_B(t_B+t_S+m_B)\right)}{2m_B{}^2m_S{}^2},\\
d_{B,S}^{2,2} &= \frac{2t_B-2t_S+m_Bt_B+m_Bt_S+m_B{}^2}{2m_B{}^2m_S}.
\end{aligned}
$$

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

$$
\begin{aligned}
d_{S,B}^{1,1} &= \frac{4(m_B+t_B-t_S)^2(2m_B-t_B+t_S)-3m_S(m_B+t_B-t_S)\left(2m_B{}^2+m_Bt_B+m_Bt_S-t_B{}^2+t_S{}^2\right)}{12m_B^2m_S^3},\\
d_{S,B}^{1,2} &= \frac{4(m_B+t_B-t_S)^2-m_S(m_B+t_B-t_S)(3m_B+3t_B+t_S)}{4m_Bm_S{}^3},\\
d_{S,B}^{2,1} &= \frac{m_B^2(2-m_S)-2(t_B-t_S)^2+m_S(t_B-t_S)(t_B+3t_S)}{4m_B^2m_S^2},\\
d_{S,B}^{2,2} &= \frac{m_B(2-m_S)+2t_B-2t_S-m_S(t_B+t_S)}{2m_Bm_S^2}.
\end{aligned}
$$

Lastly,

$$
D_S^2u_S^{\text{ante}}(m_B,t_B,m_S,t_S)=\begin{pmatrix} d_{S,S}^{1,1} & d_{S,S}^{1,2} \\ d_{S,S}^{2,1} & d_{S,S}^{2,2} \end{pmatrix},
$$

where

$$
\begin{aligned}
d_{S,S}^{1,1} &= -\frac{(m_B+t_B-t_S)^3-\frac{1}{2}m_S(m_B+t_B-t_S)^2(m_B+t_B+t_S)}{m_B {m_S}^4}, \\
d_{S,S}^{1,2} &= \frac{m_S\,(m_B+t_B-t_S)(m_B+t_B+3t_S){-4}\,(m_B+t_B-t_S)^2}{4m_B {m_S}^3}, \\
d_{S,S}^{2,1} &= \frac{m_S\,(m_B+t_B-t_S)(m_B+t_B+3t_S){-4}\,(m_B+t_B-t_S)^2}{4m_B {m_S}^3}, \\
d_{S,S}^{2,2} &= \frac{-m_B\,(m_S+2)-2t_B+2t_S-m_S\,(t_B-3t_S)}{2m_B m_S^2}.
\end{aligned}
$$

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

$$
\begin{aligned}
D_B^2 u_B^{\text{ante}}\left(\frac{2}{3}, \frac{1}{12}, \frac{2}{3}, \frac{1}{4}\right) &= \begin{pmatrix} -\frac{189}{256} & -\frac{135}{128} \\ -\frac{135}{128} & -\frac{27}{16} \end{pmatrix}, \\
D_S^2 u_S^{\text{ante}}\left(\frac{2}{3}, \frac{1}{12}, \frac{2}{3}, \frac{1}{4}\right) &= \begin{pmatrix} -\frac{81}{256} & -\frac{81}{128} \\ -\frac{81}{128} & -\frac{27}{16} \end{pmatrix}
\end{aligned}
$$

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

$$\lambda\left(J(\theta^*)\right) = \begin{pmatrix} \lambda_1\left(J(\theta^*)\right) \\ \lambda_2\left(J(\theta^*)\right) \\ \lambda_3\left(J(\theta^*)\right) \\ \lambda_4\left(J(\theta^*)\right) \end{pmatrix} = \begin{pmatrix} \frac{\sigma_3\sqrt{21}4664119193^{3/4}\sigma_2\sigma_1\left({\sigma_{11}}^{3/4}+\sigma_5+\sigma_6\right)\sigma_7}{69036339115606897664} \\ \frac{\sigma_3\sqrt{21}4664119193^{3/4}\sigma_2\sigma_1\left({\sigma_{11}}^{3/4}-\sigma_5+\sigma_6\right)\sigma_7}{69036339115606897664} \\ \frac{\sigma_3\sqrt{21}4664119193^{3/4}\sigma_2\sigma_1\left(-{\sigma_{11}}^{3/4}+\sigma_4+\sigma_6\right)\sigma_7}{69036339115606897664} \\ -\frac{\sigma_3\sqrt{21}4664119193^{3/4}\sigma_2\sigma_1\left({\sigma_{11}}^{3/4}+\sigma_4-\sigma_6\right)\sigma_7}{69036339115606897664} \end{pmatrix},$$

where

$$\begin{aligned}
\sigma_1 &= \left(29221932781 - 6048\sqrt{1402682838}\mathrm{i}\right)^{1/6}, \\
\sigma_2 &= \left(-29221932781 + 6048\sqrt{1402682838}\mathrm{i}\right)^{1/4}, \\
\sigma_3 &= (-1)^{3/4}, \\
\sigma_4 &= \sqrt{-\sigma_8 + \sigma_9 - \sigma_{10} - 9487417\sqrt{\sigma_{11}}}, \\
\sigma_5 &= \sqrt{\sigma_8 + \sigma_9 - \sigma_{10} - 9487417\sqrt{\sigma_{11}}}, \\
\sigma_6 &= 63\sqrt{3}{\sigma_{12}}^{1/6}{\sigma_{11}}^{1/4}, \\
\sigma_7 &= \Big({\sigma_{12}}^{1/3}\left(-153710549 + 72\sqrt{1402682838}\mathrm{i}\right) + 121846{\sigma_{12}}^{2/3} \\
&\quad -473375263673 - 319752\sqrt{1402682838}\mathrm{i}\Big)^{1/4} \\
\sigma_8 &= 161784\sqrt{87665798343 + 18144\sqrt{1402682838}\mathrm{i}}, \\
\sigma_9 &= 8882{\sigma_{12}}^{1/3}\sqrt{\sigma_{11}}, \\
\sigma_{10} &= {\sigma_{12}}^{2/3}\sqrt{\sigma_{11}}, \\
\sigma_{11} &= 4441{\sigma_{12}}^{1/3} + {\sigma_{12}}^{2/3} + 9487417, \\
\sigma_{12} &= 29221932781 + 6048\sqrt{1402682838}\mathrm{i}.
\end{aligned}$$

One can numerically verify that all eigenvalues have a strictly negative real part. Therefore, it holds that $\mathrm{spec}(J(\theta^*)) \subset \mathbb{C}_-^o$.

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. $\square$

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 $k \in \{\frac{0}{10}, \frac{1}{10}, \ldots, \frac{10}{10}\}$.*