Equilibrium Learning in Combinatorial Auctions: Computing Approximate Bayesian Nash Equilibria via Pseudogradient Dynamics

Stefan Heidekrüger, Nils Kohring, Paul Sutterer, Martin Bichler
Department of Informatics, Technical University of Munich
stefan.heidekrueger@in.tum.de

Abstract

Applications of combinatorial auctions (CA) as market mechanisms are prevalent in practice, yet their Bayesian Nash equilibria (BNE) remain poorly understood. Analytical solutions are known only for a handful of specific cases; in the general case, finding BNE is known to be computationally hard. Previous work on numerical computation of BNE in auctions has relied on either solving model equations manually, calculating pointwise best-responses in strategy space, or iteratively solving restricted subgames. In this study, we present a generic yet scalable alternative multi-agent reinforcement learning method that uses policy networks for strategy representation and applies modified policy gradient dynamics in self-play. Most auctions are ex-post non-differentiable, so gradients may be unavailable or misleading, and we rely on suitable pseudogradient estimates instead. Although it is well-known that gradient dynamics cannot guarantee convergence to Nash equilibria in general, we observe fast and robust convergence to approximate BNE in a wide variety of auction games and present a sufficient condition for convergence.

Introduction

Auctions are widely used in advertising, procurement, or for spectrum sales (Cramton, Shoham, and Steinberg 2004; Milgrom 2017; Ashlagi, Monderer, and Tennenholtz 2011). Auction markets inherently involve incomplete information about competitors and strategic behavior of market participants. An important line of research in game theory has long studied decision making and equilibrium states in such markets which are typically modeled as Bayesian games.

It is well-known that equilibrium computation is hard: Finding Nash equilibria is known to be PPAD-complete even for normal-form games, which assume complete information and finite action spaces, and where a Nash equilibrium is guaranteed to exist (Daskalakis, Goldberg, and Papadimitriou 2009). In auction games modeled as Bayesian games, agents' values are drawn from some continuous prior value distribution and their action sets are described by continuous sets. For markets of a single item, the landmark results by Vickrey (1961) have enabled a deep understanding of common auction formats. For multi-item auctions and more specifically *combinatorial auctions*, in which players bid on *bundles* of multiple items simultaneously, there has been little progress. While the complexity of computing Bayes-Nash equilibria (BNE) is not well understood, Cai and Papadimitriou (2014) show that BNE computation for a specific combinatorial auctions is already (at least) PP-hard. Furthermore, finding an ϵ-approximation to a BNE is still NP-hard. Explicit solutions exist for very few specific environments, but in general, we neither know whether a BNE exists nor do we have a solution theory. Combinatorial auctions have become a pivotal research problem in algorithmic game theory (Roughgarden 2016) and they are now widely used in practice (Bichler and Goeree 2017). Understanding their equilibria is paramount, and access to scalable numerical methods for computing or approximating BNE can have a significant impact.

Learning in games suffers from the *nonstationarity problem*: Each player's objective depends on other agents' actions. Prior literature on explicit *equilibrium* learning has primarily focussed on complete-information games. In contrast, we focus on learning Bayes-Nash equilibria in games with continuous action space and continuous prior type distributions. The literature on equilibrium computation for these games is in its infancy and largely relies on best-response (BR) computations.

In this paper, we propose Neural Pseudogradient Ascent (NPGA) as an equilibrium learning method that follows modified gradient dynamics. While Nash-convergence of gradient-based learning has been widely studied in complete-information games, results and methods do not apply to Bayesian auction games: First, the underlying problem is equivalent to an infinite-dimensional variational inequality, for which we do not know an exact solution method. Second, the ex-post payoff functions in auction games are non-differentiable. Finally, multi-agent gradient dynamics are known to converge to Nash equilibria only in specific classes of games, even under complete information.

NPGA relies on self-play with neural policy networks, uses evolutionary strategies to compute gradients, and can exploit GPU hardware acceleration to massively parallelize the computations. In contrast to some previous work on numerical BNE computation, NPGA does not require any setting-specific information beyond evaluating auction outcomes themselves, and it can thus be applied to arbitrary Bayesian games. We discuss a sufficient condition for convergence of NPGA to a unique Bayes-Nash equilibrium and

provide extensive experimental results on single-item and combinatorial auctions, which pose a benchmark problem in algorithmic game theory. Interestingly, we observe convergence of NPGA to approximate BNE in a wide range of small- and medium-sized combinatorial auction environments and recover the analytical Bayes-Nash equilibrium whenever it is known.

The remainder of this paper is structured as follows: First, we formally introduce the model and problem before discussing related work. Next, we introduce and discuss NPGA, before applying it to a suite of benchmark combinatorial auctions. Finally, we summarize our findings and outline future research directions.

Problem statement

Bayesian Games and Combinatorial Auctions. A *Bayesian game* or *incomplete information game* is a quintuple $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$. $\mathcal{I} = \{1, \dots, n\}$ describes the set of agents participating in the game. $\mathcal{A} \equiv \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$ is the set of possible action profiles, with $\mathcal{A}_i$ being the set of actions available to agent $i \in \mathcal{I}$. $\mathcal{V} \equiv \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ is the set of *epistemic type profiles*. $F: \mathcal{V} \to [0,1]$ defines a joint prior probability distribution over type profiles that is assumed to be common knowledge among all agents. For any dependent random variable X, we denote its cumulative distribution function by F_X and its probability density function by f_X. For example, F_{v_i} denotes the marginal distribution of agent i's type. At the beginning of the game, nature draws a type profile $v \sim F$ and each agent i is informed of their own type $v_i \in \mathcal{V}_i$ only, thus the type constitutes private information based on which each agent chooses their action $b_i \in \mathcal{A}_i$. Each agent's *ex-post* utility function is then determined by $u_i : \mathcal{A} \times \mathcal{V}_i \to \mathbb{R}$, i.e. the agent's utility depends on all agents' actions but only on their own type. Agents aim to maximize their individual utility or *payoff* u_i.

Here, we consider *sealed-bid combinatorial auctions* (CA) on items $\mathcal{M} = \{1, \dots, m\}$. Each agent, or *bidder*, is allocated a (possibly empty) bundle $k_i \in \mathcal{K} \equiv 2^{\mathcal{M}}$ of items. Each agent's types $v_i \in \mathcal{V}_i$ are given by a vector of *private valuations* over bundles, i.e. $v_i \equiv (v_i(k))_{k \in \mathcal{K}}$. Bidders then submit actions, called *bids* b_i, according to some bid-language: In the general case, where bidders might be interested in any combination of items, bids are in $\mathcal{A}_i \subseteq \mathbb{R}_+^{|\mathcal{K}|}$, i.e. each player must submit 2^m bids. In practice this is prohibitive, and one commonly studies settings where valuations exhibit some structure that allows reducing the dimensionality of both the types and actions. The settings studied in this paper have type and action spaces $\mathbb{R}_+$ or $\mathbb{R}_+^2$.

After observing their own type v_i, bidders submit bids $b_i = \beta_i(v_i)$ chosen according to some *strategy* or *bid function* $\beta_i : \mathcal{V}_i \to \mathcal{A}_i$ that maps individual valuations to an action.[1] We denote by $\Sigma_i \subseteq \mathcal{A}_i^{\mathcal{V}_i}$ the resulting strategy space of bidder i and by $\Sigma \equiv \prod_i \Sigma_i$ the space of possible joint strategies. Note that even for deterministic strategies, the spaces Σ_i are infinite-*dimensional* unless $\mathcal{V}_i$ are finite. The auctioneer collects these bids, applies some *auction mechanism* that determines (a) an allocation $x \in \mathcal{K}^n$: each bidder i receives a (possibly empty) bundle $x_i \in \mathcal{K}$, s.t. each item $m \in \mathcal{M}$ is allocated to at most one bidder, and (b) payments $p \in \mathbb{R}^n$ that the agents have to pay to the auctioneer. For risk-neutral bidders, their utility functions are *quasi-linear*[2] and given by $u_i : \mathcal{V}_i \times \mathcal{A} \to \mathbb{R}$, $u_i(v_i, b_i, b_{-i}) = v_i(x_i) - p_i$, i.e. the utility of each player is given by how much she values her assigned allocation minus the price she has to pay for it. Throughout this paper, we will differentiate between the *ex-ante* state of the game, where players know only the priors F, the *ex-interim* state, where players additionally know their own valuation $v_i \sim F_{v_i}$, and the *ex-post* state, where all actions have been played and $u_i(v, b)$ can be observed.

Equilibria in Bayesian games. In non-cooperative game theory, Nash equilibria (NE) are the central equilibrium solution concept. An action profile b^* is a pure-strategy NE of a complete-information game $G = (\mathcal{I}, \mathcal{A}, u)$ if no agent has an incentive to deviate from b^* unilaterally. Bayesian-Nash equilibria (BNE) extend this notion to incomplete-information games, calculating the expected utility $\overline{u}$ over the conditional distribution of opponent valuations v_{-i}. For valuation $v_i \in \mathcal{V}_i$, action $b_i \in \mathcal{A}_i$ and fixed opponent strategies $\beta_{-i} \in \Sigma_{-i}$, we denote the *ex-interim utility* of bidder i by

$$\overline{u}_i(v_i, b_i, \beta_{-i}) \equiv \mathbb{E}_{v_{-i}|v_i}\left[u_i\left(v_i, b_i, \beta_{-i}(v_{-i})\right)\right]. \quad (1)$$

We also denote the *ex-interim utility loss* of action b_i incurred by not playing a BR action, given v_i and β_{-i}, by

$$\overline{\ell}_i(b_i; v_i, \beta_{-i}) = \sup_{b_i' \in \mathcal{A}_i} \overline{u}_i(v_i, b_i', \beta_{-i}) - \overline{u}_i(v_i, b_i, \beta_{-i}). \quad (2)$$

Note that $\overline{\ell}_i$ can generally not be observed in online-settings because it requires knowledge of a best-response.

An ex-interim *ϵ-Bayes Nash Equilibrium (ϵ-BNE)* is a strategy profile $\beta^* = (\beta_1^*, \dots, \beta_n^*) \in \Sigma$ such that for every type v, no agent can improve her own ex-interim expected utility by more than $\epsilon \geq 0$ by deviating from the common strategy profile. Thus, in an ϵ-BNE, we have:

$$\overline{\ell}_i\left(b_i; v_i, \beta_{-i}^*\right) \leq \epsilon \text{ for all } i \in \mathcal{I}, v_i \in \mathcal{V}_i \text{ and } b_i \in \mathcal{A}_i. \quad (3)$$

A 0-BNE is simply called BNE. While BNE are usually defined at the *ex-interim* stage of the game, we also consider *ex-ante* Bayesian equilibria as strategy profiles that concurrently maximize the players' *ex-ante* expected utility $\tilde{u}$. We analogously define $\tilde{u}$ and the *ex-ante utility losses* $\tilde{\ell}$ of a strategy profile $\beta \in \Sigma$ by

$$\tilde{u}_i(\beta_i, \beta_{-i}) \equiv \mathbb{E}_{v_i \sim F_{v_i}}\left[\overline{u}_i(v_i, \beta_i(v_i), \beta_{-i})\right] \text{ and } \quad (4)$$

$$\tilde{\ell}_i(\beta_i, \beta_{-i}) \equiv \sup_{\beta_i' \in \Sigma_i} \tilde{u}_i(\beta_i', \beta_{-i}) - \tilde{u}_i(\beta_i, \beta_{-i}). \quad (5)$$

Then, an ex-ante BNE $\beta^* \in \Sigma$ can be characterized by the equations $\tilde{\ell}_i(\beta_i^*, \beta_{-i}^*) = 0$ for all $i \in \mathcal{I}$. Clearly, every ex-interim BNE also constitutes an ex-ante equilibrium and the reverse holds almost surely, i.e. any ex-ante equilibrium fulfills Equation 3, except possibly on a nullset $V \subset \mathcal{V}$, i.e. with $\int_V df_v(v) = 0$. To see this, one may consider the equation $0 = \tilde{\ell}(\beta^*) = \mathbb{E}_{v_i}\left[\overline{\ell}(\beta^*, v_i)\right]$ and that by definition $\overline{\ell}(\beta, v_i) \geq 0$.

[1] *Mixed* strategies that randomize over actions would also be possible; we restrict ourselves to *pure* or *deterministic* strategies.

[2] Our method is also applicable to bidders with risk-averse utility functions, e.g. $u = \sqrt{v - p}$.

Related work

Nash-convergence of gradient dynamics in games has been studied in evolutionary game theory and multiagent learning. While earlier work considered mixed strategies over finite normal-form games (Zinkevich 2003; Bowling and Veloso 2002; Bowling 2005; Busoniu, Babuska, and De Schutter 2008), more recently, motivated by the emergence of Generative Adversarial Networks, there has been a focus on (complete-information) games with continuous action spaces and smooth utility functions (Mertikopoulos and Zhou 2019; Letcher et al. 2019; Balduzzi et al. 2018; Schaefer and Anandkumar 2019). A result found for many of the studied 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 Coarse Correlated equilibria (CCE), a solution concept weaker than Nash equilibria. An analogous result exists for finite-type Bayesian games, where no-regret learners are guaranteed to converge to a Bayesian CCE (Hartline, Syrgkanis, and Tardos 2015). In the present paper, we study equilibrium learning via gradient dynamics in *continuous-type* Bayesian games, specifically auctions, where they have not been investigated previously to our knowledge.

Earlier approaches to **BNE computation in auctions** were usually setting-specific and relied on reformulating Equation 3 as a system of differential equations (where possible), then solving this equation analytically or numerically (Krishna 2009; Ausubel and Baranov 2019). Armantier, Florens, and Richard (2008) introduced a method that is expresses the Bayesian game as the limit of a sequence 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 still 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. Most recently, Bosshard et al. (2017, 2020) proposed a method to find BNE in combinatorial auctions via smoothed best-response dynamics. The method explicitly computes point-wise best-responses in a fine-grained linearization of the strategy space via sophisticated Monte-Carlo integration. While avoiding reliance on setting specific knowledge, the best response computation suffers from the curse of dimensionality in larger games.

Pseudogradient dynamics in auction games

Next, we present our method for equilibrium computation in auctions, which we call Neural Pseudogradient Ascent (NPGA). On a high level, we propose following the ex-ante gradient dynamics of the game via simultaneous gradient ascent of all bidders. As we will see, however, computing the gradients themselves is not straightforward in the auction setting and we will need some modifications to established gradient dynamics methods such as (Zinkevich 2003; Silver et al. 2014). For now, assume that players can observe a gradient-oracle $\nabla_{\beta_i}\tilde{u}_i(\beta_i, \beta_{-i})$ with respect to the current strategy profile β^t in each iteration. Then the rule proposes that players perform a projected gradient update:

$$\beta_i^t \equiv \mathcal{P}_{\Sigma_i}\left(\beta_i^{t-1} + \Delta_i^t\right) \text{ with } \Delta_i^t \propto \nabla_{\beta_i}\tilde{u}_i(\beta_i, \beta_{-i}), \quad (6)$$

where $\mathcal{P}_{\Sigma_i}(\,\cdot\,)$ is the projection onto the set of feasible strategies for agent i. Several things must be noted about Equation 6: First, we consider the gradient dynamics of the *ex-ante* utility $\tilde{u}$, rather than ex-interim or ex-post utilities. The goal of an individual update step is thus to marginally improve the expected utility of player i across all possible joint valuations $v \sim F$. This perspective ultimately considers low-probability events less important than high-probability events, which is in contrast to some other methods, which explicitly aim to optimize *all* ex-interim states (Bosshard et al. 2017). Second, to compute the gradient oracle $\nabla_\beta \tilde{u}$ in self-play, we rely on access to other players' strategies, but evaluating each player's policy relies only on their own valuation. We thus follow the centralized-training, decentralized-execution framework common in multi-agent learning. Third, $\beta_i \in \Sigma_i$ are functions in an infinite-dimensional function space, so the gradient $\nabla_{\beta_i}\tilde{u}_i$ is itself a *functional* derivative. In our ex-ante perspective, we thus consider this to be the Gateaux derivative over the Hilbert space Σ_i, equipped with the inner product $\langle \psi, \beta_i \rangle = \mathbb{E}_{v_i \sim F_{v_i}}\left[\psi(v_i)^T \beta_i(v_i)\right]$ (which, in turn, defines the projection in Equation 6 as $\mathcal{P}_{\Sigma_i}(\beta) \equiv \arg\min_{\sigma \in \Sigma_i} \langle \sigma - \beta, \sigma - \beta \rangle$).

To implement this derivative in practice, we employ *policy networks* $\beta_i(v_i) \equiv \pi_i(v_i; \theta_i)$ specified by a neural network architecture and a corresponding parameter vector $\theta_i \in \mathbb{R}^{d_i}$. Importantly, given a suitable network architecture, one can ensure that all θ_i yield feasible bids, thus making the projection in the update step obsolete. In the empirical part of this study, we restrict ourselves to fully-connected feed-forward neural networks with ReLU activations in the output layer, which ensure nonnegative bids—the only feasibility constraint in the auctions we study. In any case, $d_i \in \mathbb{N}$ is finite and we thus transform the problem of choosing an infinite-dimensional strategy into choosing a finite-dimensional parameter vector θ_i.

Policy Pseudogradients. The *deterministic policy gradient theorem* (Silver et al. 2014) gives an established, canonical way to compute the payoff gradient with respect to the parameters θ: $\nabla_{\theta_i}\tilde{u}_i(\pi_i(\,\cdot\,;\theta_i), \beta_{-i}) = \mathbb{E}_{v \sim F}\left[\nabla_{\theta_i}\pi(v_i;\theta_i)\nabla_{b_i}u_i(v_i; b_i, \beta_{-i}(v_{-i}))|_{b_i=\pi_i(v_i;\theta_i)}\right]$. However, the regularity conditions required by the theorem are commonly violated in combinatorial auctions. In particular, 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 neither differentiable nor subdifferentiable in b_i. While this nondifferentiability does not extend to $\tilde{u}$, it nevertheless renders the policy gradient formula above inapplicable: Although the set of discontinuities is a v-nullset in practice, one can show that even on the differentiable intervals of $u_i(v_i, \cdot, b_{-i})$, its true gradient provides systematically misleading signals: Consider a first-price sealed-bid auction in which winning bidders pay their bid amount b_i. The utility graph is separated into two sections: Bidding lower than the highest opposing bid leads to zero payoff and thus no learning feedback, $\nabla_{b_i}u_i = 0$;

winning, however, *must* yield learning feedback to decrease the winning bid, $\nabla_{b_i} u_i = -1$. Back-propagation will thus lead to a steady decrease of bids in every iteration, until all players bid constant zero for any valuation.

To alleviate this, we instead estimate the policy gradient using a sampling approach based on evolutionary strategies (ES) (Salimans et al. 2017). To calculate $\nabla_\theta \tilde{u}$, we perturb the parameter vector P times, $\theta_{i;p} \equiv \theta_i + \varepsilon_p$, using zero-mean Gaussian noise $\varepsilon_p \sim \mathcal{N}(0, \sigma^2)$ for $p \in \{1, \ldots, P\}$, where P, σ are hyperparameters. We then calculate each perturbation's *fitness*, $\varphi_p \equiv \tilde{u}_i(\pi_i(v_i; \theta_{i;p}), \beta_{-i})$, via Monte-Carlo integration, and estimate the gradients as the fitness-weighted perturbation noise $\nabla_\theta^{ES} \equiv \frac{1}{\sigma^2 P} \sum_p \varphi_p \varepsilon_p$. Salimans et al. motivated their application of this gradient estimate to reinforcement learning as it's applicable to parallelization across large-scale CPU-clusters, but here we instead exploit its property that it gives an asymptotically unbiased estimator of $\nabla_\theta \tilde{u}$ even when $\nabla_b u$ itself is not well-defined. To summarize, NPGA "implements" Equation (6) via ES-pseudogradients and a NN parametrization of strategy functions which renders the projection step unnecessary:

$$\beta_i^t \equiv \pi_i(\,\cdot\,; \theta_i^t) \text{ with } \theta_i^t \equiv \theta_i^{t-1} + \Delta_i^t \text{ where } \Delta_i^t \propto \nabla_{\theta_i^t}^{ES} \quad (7)$$

Vectorizing Auction Evaluations. The only information about the game G needed in the computation of NPGA is access to the evaluation of $\tilde{u} = \mathbb{E}_{v \sim F}[u]$ for a given strategy profile. Given a vectorized implementation of the joint ex-post utility function u, estimating $\tilde{u}$ via Monte-Carlo integration over $\mathcal{V}$ is suitable to parallel execution on hardware accelerators such as GPUs. To this end, we built custom vectorized implementations of common auction mechanisms using the PyTorch framework (Paszke et al. 2017), allowing us to perform this evaluation multiple orders of magnitude faster compared to previous numerical work on auctions. For moderately sized auction games, allocations x can be computed in a vectorized fashion via full enumeration of feasible allocations. Common payment rules either have inherently vectorizable closed-form formulation or can be reformulated as the solution of a constrained quadratic program (e.g. the Vickrey-Clarke-Groves (VCG) mechanism or core-selecting pricing rules (Day and Cramton 2012)). To solve a large batch of the latter in parallel, we leverage a custom vectorized implementation of interior-point methods. (A similar approach was used by Amos and Kolter (2019).)

A convergence criterion. As discussed above, gradient dynamics do not generally converge to Nash. In differentiable, finite-dimensional, complete-information games (auctions are neither!), Mertikopoulos and Zhou (2019) show that strict monotonicity of the payoff gradients is likewise a sufficient condition for almost-sure convergence of gradient dynamics to a unique Nash equilibrium. Ui (2016) shows an analogous result for ex-post differentiable Bayesian games, in which payoff-monotonicity guarantees the existence of a unique BNE. However, the result likewise does not directly apply to auctions due to their ex-post non-differentiability. Instead, we give a slightly less restrictive criterion based on ex-interim payoff monotonicity that ensures convergence of gradient dynamics and whose formulation is compatible with auction games.

Definition 1 (Strict Ex-interim Payoff Monotonicity)**.** *Let $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$ be a Bayesian game, such that the individual ex-interim utilities are continuously differentiable in b_i with gradients bounded by a constant $Z > 0$ via $\|\nabla_{b_i} \overline{u}_i(v_i, b_i, \beta_{-i})\| \leq Z$. G is called* strictly (ex-interim) payoff-monotone*, if for all $i \in \mathcal{I}$, $\beta_{-i} \in \Sigma_{-i}$, $a_i, b_i \in \mathcal{A}_i$ and almost everywhere $v_i \in \mathcal{V}_i$ the following holds:*

$$\langle \nabla_{a_i} \overline{u}_i(v_i, a_i, \beta_{-i}) - \nabla_{b_i} \overline{u}_i(v_i, b_i, \beta_{-i}), a_i - b_i \rangle \;<\; 0. \quad (8)$$

While analytical verification of this criterion is elusive, except in special settings, it can (approximately) be checked numerically by sampling pairs of action profiles a, b for all players and using finite-difference gradient-estimators.

In the following, we provide a convergence result for NPGA under ex-interim monotonicity. For our convergence analysis, we will rely on certain properties of "appropriate" neural network architectures, defined below.

Definition 2 (Regular Convex Policy Network)**.** *A* Regular Convex Policy Network *is a neural network $\pi_i : \mathcal{V}_i \times \Theta_i \to \mathcal{A}_i$ with $\dim \Theta_i = d_i$ and the following properties:*

1. *π_i is a* convex neural network *in its parameters: For any convex objective function $g : \Sigma_i \to \mathbb{R}$, the map $\theta_i \mapsto g(\pi_i(\,\cdot\,, \theta_i))$ is convex.*
2. *π_i* universally approximates *Σ_i: There exists a $\delta > 0$, s.t. for all $\beta_i \in \Sigma_i$ there is a parameter vector $\theta_i \in \Theta_i$ with $\mathbb{E}_{v_i}[\|\beta_i(v_i) - \pi_i(v_i, \theta_i)\|] \leq \delta$.*
3. *π_i is* Lipschitz-continuous *in its parameters: $\exists L > 0$: $\forall \theta_i, \theta_i' \in \Theta_i$: $\mathbb{E}_{v_i}[\|\pi_i(v_i, \theta_i) - \pi_i(v_i, \theta_i')\|] \leq L \|\theta_i - \theta_i'\|$.*

Neural networks that are employed in practice (and in our empirical analysis) generally do not comply with this definition, but such networks have been shown to exist. For a concrete architecture, see Bach (2017) who studies wide single-hidden-layer networks with ReLU activations, in which only the output-layer weights are being trained. We'll state our main proposition before discussing this difference further:

Proposition 1. *Let $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$ be a Bayesian game such that the ex-post utilities exist, and such that the ex-interim payoff-gradients exist and fulfill strict ex-interim payoff monotonicity. Then, with an NN architecture as in Definition 2 and appropriate update step sizes, NPGA converges to an ex-ante ϵ-BNE of G, where $\epsilon \leq Z(2L\sigma\sqrt{d} + \delta)$.*

While existence and uniqueness of BNE in infinite-dimensional games are unknown in the general case, Proposition 1 guarantees efficient computability in a wide range of settings, some of which we explore in the next section. Still, it's important to note that there may be auctions for which payoff-monotonicity does not hold.

An abridged proof of Proposition 1 is given at the end of this article. We focus on the instructive parts of the proof and omit the more tedious derivations for brevity. As demonstrated in the proof, the use of Regular Convex Policy Networks transforms the training process into a problem of finding a Nash Equilibrium in a *concave*, finite-dimensional, complete-information game. Crucially, concavity of this game ensures existence of, and convergence to, a unique global equilibrium. Just as neural networks are known to find "good" solutions to nonconvex optimization

problems in practice despite a lack of theoretical guarantees, we will see below that we observe convergence to BNE when using standard neural network architectures that don't meet Definition 2: As such, we see Regular Convex Policy Networks as a helpful tool for theoretical analysis, but their implementation is generally neither practical nor desirable while common architectures achieve similar results.

Empirical Results

We evaluate NPGA on three suites of auction theoretic settings: First we validate our method on a suite of the most commonly studied auctions, i.e. single-item auctions with symmetric priors, before considering two suites of combinatorial auctions, the LLG and LLLLGG environments. In total, we study 21 different auction settings with different numbers of players, pricing rules, risk-profiles, and prior distributions of the valuations. In 18 of these settings, the (unique) BNE is known analytically, in three settings, no BNE is known.

Evaluaton Metrics. To evaluate the quality of strategy-profiles β learned by NPGA, we will provide four metrics: Whenever we have access to the analytical solution BNE β^*, we can check whether $\beta \to \beta^*$. We report each agent's

1. utility loss ℓ_i^* that results from unilaterally deviating from the BNE strategy profile β^* by playing the learned strategy β_i instead: $\ell_i^* \equiv \tilde{\ell}_i(\beta_i, \beta^*_{-i})$, compare Equation 5,
2. distance $\|\beta_i - \beta_i^*\|_{\Sigma_i}$ to the BNE in strategy space.

Both of these can be estimated via Monte-Carlo integration over the valuations $v \sim F$, i.e. given a batch of size H of valuations $(v_{h;i}, v_{h,-i})$, we approximate ℓ_i^* by the sample mean of $\tilde{\ell}_i(\beta_i(v_{h;i}), \beta^*_{-i}(v_{h;-i}))$ and $\|\beta_i - \beta_i^*\|$ by the RMSE of $\beta_i(v_{h;i})$ and $\beta_i^*(v_{h;i})$ in action-space.

However, we are also interested in judging the quality of β when no BNE is known. To do so, we also estimate the potential gains of deviating from the current strategy profile $\hat{\ell}_i \approx \tilde{\ell}_i(\beta_i; \beta_{-i})$ as well as an estimator $\hat{\epsilon}$ to the "true" epsilon of β (smallest ϵ s.t. β forms an ex-interim ϵ-BNE). In the absence of analytical solutions, one may periodically calculate these estimators and use them as a termination criterion once the desired precision is reached. As we will see, these additional metrics are expensive: To calculate the estimators $\hat{\ell}$ and $\hat{\epsilon}$ we introduce a grid of W equidistant points b_w per bidder, covering the action spaces $\mathcal{A}_i$. Now, given a valuation v_i and a bid b_i, we approximate the ex-interim utility loss $\bar{\ell}(v_i, b_i, \beta_{-i})$ of b_i at v_i via its sample-mean $\hat{\lambda}_i(v_i; b_i, \beta_{-i})$ over a batch of H opponent valuations $v_{h;-i}$. To evaluate $\hat{\lambda}_i$ at a single v_i, we thus need $(W+1)H$ auction evaluations. Running *another* batch H on i's valuation, we can then estimate the

3. worst-case ex-interim loss: $\hat{\epsilon} = \max_h \hat{\lambda}_i(v_{h,i}; \beta_i(v_{h,i}))$,
4. and the ex-ante loss: $\hat{\ell} = \frac{1}{H} \sum_h \hat{\lambda}_i(v_{h,i}; \beta_i(v_{h,i}), \beta_{-i})$.

Estimations of $\hat{\lambda}$ can be shared for both computations, nevertheless we need $\mathcal{O}(nWH^2)$ auction evaluations to calculate these metrics (in contrast, an iteration of NPGA requires only $\mathcal{O}(nPK)$ evaluations, with $P \ll W$). As the metrics $\hat{\epsilon}, \hat{\ell}$ are expensive to compute on dense grids, we use smaller batch sizes W and H than in evaluating ℓ^*, and calculate these metrics only in every 100th NPGA iteration.

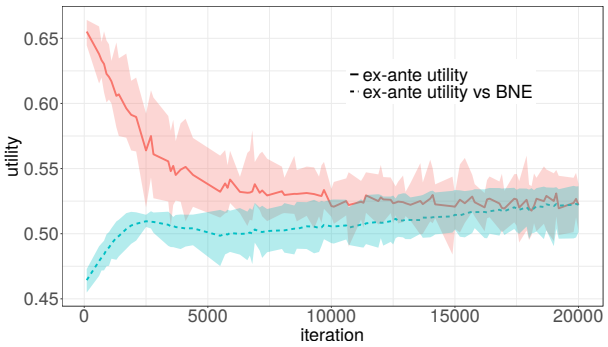

Figure 1: Learning curve of NPGA in 10-player FPSB auction with Gaussian priors, evaluated in self-play (red, solid) and against the BNE (blue, dotted). Mean/min/max over 10 repetitions.

Table 1: Performance of strategies learned by NPGA in single-item FPSB auctions. Results averaged over 10 runs of 5k (uniform risk-neutral) or 20k (others) iterations each.

valuations	n	ℓ^*	$\|\beta^* - \beta\|$	$\hat{\ell}$	$\hat{\epsilon}$	time sec/it
Uniform $\mathcal{U}(0,10)$ **risk-neutral**	**2**	0.0000	0.0072	0.0011	0.0059	0.31
	3	0.0001	0.0104	0.0007	0.0051	0.40
	5	0.0001	0.0194	0.0005	0.0053	0.46
	10	0.0001	0.0303	0.0003	0.0047	0.73
Uniform $\mathcal{U}(0,10)$ **risk-averse**	**2**	0.0003	0.0057	0.0012	0.0065	0.46
	3	0.0001	0.0069	0.0008	0.0048	0.52
	5	0.0001	0.0161	0.0006	0.0066	0.63
	10	0.0002	0.0383	0.0005	0.0085	0.93
Gaussian $\mathcal{N}(15, 10^2)$ **risk-neutral**	**2**	0.0079	0.3684	0.0443	0.4394	0.31
	3	0.0103	0.4478	0.0225	0.9723	0.39
	5	0.0172	0.8819	0.0176	1.7324	0.45
	10	0.0169	1.8801	0.0118	2.1660	0.68

We use common **hyperparameters** across almost all settings unless noted otherwise: Fully connected NNs with two hidden layers of 10 nodes each with SeLU (Klambauer et al. 2017) activations. ES-parameters P=64, σ=$\frac{1}{\sqrt{d_i}}$. Adam optimizer steps with default hyperparameters (Kingma and Ba 2017). To avoid degenerate initializations of θ (e.g. where one or more bidders bid constant zero due to dead ReLUs in the output layer), we perform supervised pre-training to the *truthful strategy* $\beta_i(v_i) = v_i$. All experiments were performed on a single Nvidia GeForce 2080Ti and Monte-Carlo batch-sizes were chosen to maximize GPU-RAM utilization: A learning batch size of K=2^{18}; primary evaluation batch size (for ℓ^*, $||\beta - \beta^*||$) of H=2^{22}; and secondary evaluation batch size H=2^{12} and grid size W=2^{10} (for $\hat{\ell}, \hat{\epsilon}$).

Single-Item Auctions

First-price sealed-bid (FPSB) auctions on a single item, in which the highest-bidding player wins the item and pays her own bid as price, are the best-known auctions and for many configurations their BNE are known analytically (Krishna

Table 2: Results of NPGA in LLG-settings with independent and correlated valuations. Values are means of 10 runs of 5,000 iterations. For FPSB, no BNE is known; for correlated priors, estimating $\hat{\ell}, \hat{\epsilon}$ is not straightforward. Negative ℓ^* are artefacts of the sample variance of F_v at available precision.

priors	payment	bidder	ℓ^*	$\|\beta^*-\beta\|$	$\hat{\ell}$	$\hat{\epsilon}$	time sec/it
independent	n.-VCG	locals	0.0001	0.0050	0.0002	0.0009	0.84
		global	0.0000	0.0269	0.0000	0.0001	
	n.-bid	locals	-0.0002	0.0073	0.0003	0.0013	0.79
		global	0.0000	0.0424	0.0000	0.0001	
	n.-zero	locals	-0.0001	0.0078	0.0002	0.0019	0.79
		global	0.0000	0.0088	0.0000	0.0001	
	FPSB	locals	–	–	0.0009	0.0031	0.65
		global	–	–	0.0016	0.0064	
correlated	n.-VCG	locals	-0.0001	0.0042	–	–	0.80
		global	0.0000	0.0305	–	–	
	n.-bid	locals	0.0003	0.0064	–	–	0.83
		global	0.0000	0.0498	–	–	
	n.-zero	locals	0.0001	0.0059	–	–	0.81
		global	0.0000	0.0072	–	–	

2009). We apply NPGA to 12 such FPSB settings with varying number of players, valuation priors and risk attitudes. The results are given in Table 1.

In all settings, we observe convergence of NPGA to the analytical BNE in terms of ex-ante payoff, both when evaluated in self-play and against opponents playing the BNE. However, we also see that sometimes there's no full norm-convergence in the strategy space: This indicates that NPGA learns *ex-ante* BNE as the algorithm is intended, but may bid suboptimally in "unimportant" regions of the valuation space, e.g. when there are many players and i's valuation is low (see Gaussian-10p setting). Nevertheless, even when the strategy-space norm is nonzero, the learned strategy becomes indistinguishable from the BNE in terms of ex-ante utility: Figure 1 shows the learning curve of a run in the Gaussian-10p setting for which the norm has not converged. Additionally, we observe that the exploitability-estimate $\hat{\ell}$, while not exactly equal to ℓ^*, is consistent in order of magnitude and may thus serve as a suitable proxy for convergence in the absence of known BNE. For a complete treatment of the single-item setting, we also implemented Second Price auctions, where NPGA robustly found the (truthful) BNE.

Combinatorial Auctions

Local-global combinatorial auctions will serve as our main benchmark for BNE compuation. In such auctions, there are two groups of bidders, locals and globals: Globals g are interested in larger bundles of items while their priors allow them to draw higher valuations, so local bidders l need to coordinate to outbid the globals. We consider settings where locals have (possibly correlated) uniform priors $v_{ik} \sim \mathcal{U}(0, \overline{v}_i)$ with $\overline{v}_l=1, \overline{v}_g=2$ (for each bundle $k \in \mathcal{K}_i$). The 3-player LLG setting is a standard setting in auction theory and one of the smallest CAs that requires strategic cooperation between bidders; the larger 6-player LLLLGG setting is the most complex environment in which approximate BNE have been computed to date (Bosshard et al. 2017).

The **LLG setting** includes two local bidders and one global bidder that bid on $m=2$ items. Local bidders $i=1,2$ are each interested in the bundle $\{i\}$, while the global bidder wants the package $\{1,2\}$ of both items. The setting exhibits a mix of competition and cooperation and has been extensively studied in the context of different *core-selecting* pricing rules (Day and Cramton 2012). Closed-form solutions of the unique, local-symmetric BNE under three such rules are known in the LLG setting: the nearest-VCG rule, the nearest-zero rule, and the nearest-bid rule. The interested reader is referred to Ausubel and Baranov (2019). In these equilibria, the global bidder bids truthfully, while the local bidders' BNE strategies differ in each payment rule and depending on the correlation between locals' priors. We evaluate NPGA on these core payment rules with independent and correlated priors (γ=0.5) as well as with first-price payments with independent priors, for which no exact BNE is known. Numerical results are presented in Table 2. We again observe that NPGA converges to the BNE in all six settings where it is known. In fact, after low hundreds of iterations, we can no longer detect a difference in utility to the true BNE with available measurement precision, while still observing slight differences in strategy-space distance: Figure 2 depicts the strategy learned by NPGA after 5,000 iterations in comparison to the analytical BNE strategy for the nearest-zero payment rule. In the FPSB auction, no BNE is known, but $\hat{\ell} \approx 10^{-3}$ (vs utilities of 0.426 (global), 0.149 (locals)) indicate that exploitability of β is minuscule.

In the **LLLLGG setting**, four local and two global bidders compete for six items. Each bidder is interested in two overlapping bundles of size 2 (locals) or 4 (global), with actions sets $\mathcal{A}_i = \mathbb{R}^2_+$. We apply NPGA to LLLLGG with first-price and nearest-vcg rules, where no BNE are known. In this larger setting, computing the clearing prices p is computationally expensive and forms the bottleneck: Nearest-VCG prices require solving a linear- and a subsequent quadratic

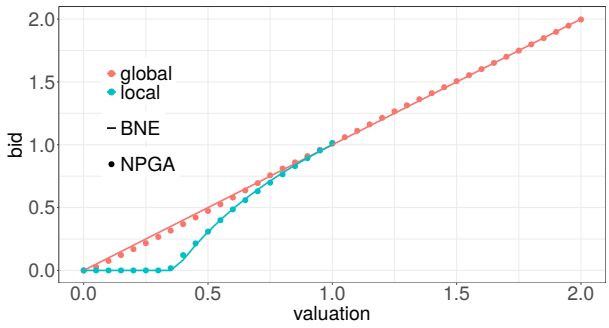

Figure 2: Learned (dots) and BNE (lines) strategies in LLG-setting with nearest-zero core payment rule.

Table 3: Results and runtime of NPGA after 5k (1k) iterations in the LLLLGG first-price (nearest-VCG) auction over 10 (2) repititions. Values are mean and (standard deviation).

payment	bidder	$\hat{\ell}$	$\hat{\epsilon}$	time sec/iter
first-price	locals	0.0015 (0.0003)	0.0109 (0.0025)	0.97
	globals	0.0010 (0.0002)	0.0077 (0.0016)	(0.005)
near.-VCG	locals	0.0013 (0.0003)	0.0052 (0.0012)	275.22
	globals	0.0011 (0.0006)	0.0098 (0.0059)	(0.670)

optimization problem for each instance (Day and Cramton 2012). As iterations of NPGA need many thousands of samples, we rely on a custom interior-point solver that can solve batches of quadratic optimization problems on the GPU. Still, we make the following hyperparameter adjustments in the nearest-VCG setting: $P{=}32$; $K{=}2^{14}$; $H{=}2^7$ and $W{=}2^8$ on two experiments of 1,000 iterations each. Results are shown in table 3. Under both pricing rules, NPGA learns strategy profiles with an estimated ex-ante utility loss $\hat{\ell}{<}0.002$ for both local and global bidders. Global (local) bidders achieve stable average utilities of 0.238 (0.18) in first-price and 0.181 (0.201) in nearest-vcg, thus the estimated loss indicates that players can be exploited for less than 1% of their achieved utility.

Conclusion and future work

This paper explores equilibrium learning in Bayesian games, one of the large unsolved problems in algorithmic game theory. Gradient dynamics are challenging in Bayesian auction games for several reasons: these games are not differentiable, and the continuous type- and action spaces make efficient representation difficult or expensive. We propose Neural Pseudogradient Ascent as a numerical method that relies on policy pseudogradients. We hope that our approach will lead to further study of gradient dynamics in game-theoretical and microeconomic settings where they have previously been considered inapplicable.

In experiments, we validate our method on standard single-item auctions and combinatorial auctions, which constitute a pivotal problem in algorithmic game theory with many practical applications. We find that NPGA converges to approximate BNE for central benchmark problems in this field, and we prove a sufficient criterion under which almost sure convergence to equilibria is guaranteed. In summary, the method can provide an effective numerical tool to compute approximate BNE not only for combinatorial auctions but also for other types of Bayesian games without setting-specific customization, while running on consumer hardware and leveraging GPU-parallelization for performance.

Proof Outline of Proposition 1

Proof. We will proceed as follows: We will approximate the infinite-dimensional Bayesian-game by a finite-dimensional (but continuous action), complete-information game. Under strict monotonicity, the regularity conditions above and with Regular Convex Policy Networks, we will argue that NPGA almost surely finds an approximation of the unique NE in this finite-dimensional game. We then give a bound on the ex-ante loss in the original game for this strategy profile found by NPGA, thus certifying an ex-ante ϵ-BNE.

Existence of the ex-interim gradients (Def. 1) implies that the ex-ante utilities $\tilde{u}_i(\beta_i, \beta_{-i}){=}\mathbb{E}_{v_i}\left[\overline{u}_i(v_i, \beta(v_i), \beta_{-i})\right]$ are Gâteaux-diff'ble in the Hilbert spaces Σ_i with Gâteaux-gradients $\nabla_{\beta_i}\tilde{u}_i[\beta](v_i){=}\nabla_{b_i}\overline{u}_i(v_i, b_i, \beta_{-i})|_{b_i=\beta_i(v_i)}$. Ex-interim monotonicity then implies (details ommitted) $\langle\nabla_{\beta_i}\tilde{u}_i[\beta_i, \beta_{-i}]{-}\nabla_{\alpha_i}\tilde{u}_i[\alpha_i, \beta_{-i}], \beta_i{-}\alpha_i\rangle_{\Sigma_i}{<}0$, i.e. the ex-ante gradients are strictly monotone operators on Σ_i, thus the $\tilde{u}_i$ are strictly concave (Bauschke and Combettes 2017).

With NNs as in Definition 2, the functions $\breve{u}_i(\theta_i) \equiv \tilde{u}_i(\pi_i(\,\cdot\,, \theta_i), \beta_{-i})$ are then also strictly concave in θ_i for any opponent strategies β_{-i}. We can then construct a finite-dimensional, complete-information *Parameter Game* $\breve{G}$, in which all players approximate their strategies β in G using policy networks and we interpret the parameters $\theta \in \mathbb{R}^d$ of the networks as *the action* of the new game: $\breve{G} \equiv (\mathcal{I}, \Theta, \breve{u})$. As this game is finite-dimensional and concave, Mertikopoulos and Zhou (2019) establish that it has a unique Nash equilibrium $\breve{\theta}^*$ and the *dual averaging* (DA) algorithm converges almost surely to $\breve{\theta}^*$ given an unbiased and finite-variance oracle of the gradients $\nabla_{\theta_i}\breve{u}_i(\theta_i; \theta_{-i})$. Next, we argue that $\breve{\theta}^*$ induces an approximate BNE in the original Bayesian Game G before analyzing how NPGA implements DA in $\breve{G}$ with noisy feedback, thus finding a good approximation of $\breve{\theta}^*$. Let $\breve{\theta}^*$ thus be the Nash equilibrium of $\breve{G}$. Then for any player i, $\breve{\theta}^*_i$ is a best response (BR) to $\breve{\theta}^*_{-i}$ and $\pi_i(\,\cdot\,, \breve{\theta}^*_i)$ is an ex-ante BR to $\pi_{-i}(\,\cdot\,, \breve{\theta}^*_{-i})$ in the Bayesian Game with the *restricted strategy space* expressible by the network. Due to universal approximation properties of π_i, however, any BR β^*_i in the *unrestricted* game G must be close in function space to $\pi_i(\,\cdot\,, \breve{\theta}^*_i)$, and the ex-ante utility loss incurred by not playing β^*_i instead of $\pi_i(\,\cdot\,; \breve{\theta}^*_i)$ is bounded: In fact, with the assumed regularity conditions, one can prove the following for arbitrary θ_{-i} (details ommitted): If $\breve{\theta}^*{\in}\Theta_i$ and $\beta^*_i{\in}\Sigma_i$ are BRs to θ_{-i} in $\breve{G}$ and G, respectively, then $\tilde{\ell}_i(\breve{\theta}^*; \theta_{-i}) \leq Z\delta$. In the NE, all $\breve{\theta}^*_i$ are BRs, so we have $\tilde{\ell}(\breve{\theta}^*) \leq Z\delta$.

Finally, we show that NPGA finds a good approximation of $\breve{\theta}^*$. As deliberated above, we choose the NN architecture in such a way that Θ becomes unconstrained, i.e. any parameter $\theta_i{\in}\mathbb{R}^{d_i}$ is feasible, where d_i is the dimension of the network for player i. On an *unconstrained* action set Θ, however, dual averaging (DA) with Euclidean regularization is equivalent to Online Gradient Ascent on $\breve{u}$ (Zinkevich 2003), thus NPGA implements DA on $\tilde{u}$ with gradient oracle ∇^{ES}_{θ}.

To use the convergence result cited above of NPGA to $\breve{\theta}$, it would remain to show that the Neural Pseudogradients $\nabla^{ES}\tilde{u}$ are finite-variance and unbiased estimators of the true gradients $\nabla_\theta\breve{u}$. This is unfortunately violated for strictly positive ES-noise variance σ^2 as used by NPGA. However, for $\sigma{>}0$ we can set $\breve{u}^\sigma_i \equiv \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,\sigma^2 I)}\left[\breve{u}_i(\theta_i + \varepsilon, \theta_{-i})\right]$ and introduce yet another finite-dimensional game $\breve{G}^\sigma = (\mathcal{I}, \Theta, \breve{u}^\sigma)$. Now, one can show (details ommitted) that (1) $\breve{G}^\sigma$ is likewise concave, that (2) the ES-gradients are finite-variance and unbiased estimators of $\breve{u}^\sigma$, and (3) that the loss of any parameter choice θ_i in $\breve{G}$ is bounded by that in $\breve{G}^\sigma$ via $\breve{\ell}_i(\theta_i, \theta_{-i}) \leq \breve{\ell}^\sigma_i(\theta_i, \theta_{-i}) + 2ZL\sqrt{d_i}\sigma$. Due to (1), $\breve{G}^\sigma$ again admits a unique NE θ^*, and with (2) and Definition 1, NPGA converges to θ^* almost surely for appropriate step sizes via the results in Mertikopoulos and Zhou (2019).

To summarize, we showed that NPGA finds a parameter profile θ^* that forms a NE of $\breve{G}^\sigma$ and which retains an-ex ante loss in G of $\tilde{\ell}_i(\theta^*) \leq Z\delta + 2ZL\sqrt{d_i}\sigma$. Thus, NPGA converges almost surely to an ex-ante ϵ-BNE of G. □

References

Amos, B.; and Kolter, J. Z. 2019. OptNet: Differentiable Optimization as a Layer in Neural Networks. *arXiv:1703.00443 [cs, math, stat]* URL http://arxiv.org/abs/1703.00443. Comment: ICML 2017.

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. ISSN 08837252, 10991255. doi:10.1002/jae.1040. URL http://doi.wiley.com/10.1002/jae.1040.

Ashlagi, I.; Monderer, D.; and Tennenholtz, M. 2011. Simultaneous Ad Auctions. *Mathematics of Operations Research* 36(1): 1–13.

Ausubel, L. M.; and Baranov, O. 2019. Core-Selecting Auctions with Incomplete Information. *International Journal of Game Theory* ISSN 1432-1270. doi:10.1007/s00182-019-00691-3. URL https://doi.org/10.1007/s00182-019-00691-3.

Bach, F. 2017. Breaking the Curse of Dimensionality with Convex Neural Networks. *Journal of Machine Learning Research* 18(19): 1–53. ISSN 1533-7928. URL http://jmlr.org/papers/v18/14-546.html.

Balduzzi, D.; Racaniere, S.; Martens, J.; Foerster, J.; Tuyls, K.; and Graepel, T. 2018. The Mechanics of N-Player Differentiable Games. *arXiv:1802.05642 [cs]; ICML* URL http://arxiv.org/abs/1802.05642. Comment: ICML 2018, final version.

Bauschke, H. H.; and Combettes, P. L. 2017. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Cham: Springer International Publishing. ISBN 978-3-319-48310-8 978-3-319-48311-5. doi:10.1007/978-3-319-48311-5. URL http://link.springer.com/10.1007/978-3-319-48311-5.

Bichler, M.; and Goeree, J. K. 2017. *Handbook of Spectrum Auction Design*. Cambridge University Press.

Bosshard, V.; Bünz, B.; Lubin, B.; and Seuken, S. 2017. Computing Bayes-Nash Equilibria in Combinatorial Auctions with Continuous Value and Action Spaces. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 119–127. Melbourne, Australia: International Joint Conferences on Artificial Intelligence Organization. ISBN 978-0-9992411-0-3. doi:10.24963/ijcai.2017/18. URL https://www.ijcai.org/proceedings/2017/18.

Bosshard, V.; Bünz, B.; Lubin, B.; and Seuken, S. 2020. Computing Bayes-Nash Equilibria in Combinatorial Auctions with Verification. *Journal of Artificial Inelligence Research* URL http://arxiv.org/abs/1812.01955. Comment: 35 pages.

Bowling, M. 2005. Convergence and No-Regret in Multiagent Learning. In Saul, L. K.; Weiss, Y.; and Bottou, L., eds., *Advances in Neural Information Processing Systems 17*, 209–216. MIT Press. URL http://papers.nips.cc/paper/2673-convergence-and-no-regret-in-multiagent-learning.pdf.

Bowling, M.; and Veloso, M. 2002. Multiagent Learning Using a Variable Learning Rate. *Artificial Intelligence* 136(2): 215–250. ISSN 0004-3702. doi:10.1016/S0004-3702(02)00121-2. URL http://www.sciencedirect.com/science/article/pii/S0004370202001212.

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. ISSN 1094-6977. doi:10.1109/TSMCC.2007.913919.

Cai, Y.; and Papadimitriou, C. 2014. Simultaneous Bayesian Auctions and Computational Complexity. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation - EC '14*, 895–910. Palo Alto, California, USA: ACM Press. ISBN 978-1-4503-2565-3. doi:10.1145/2600057.2602877. URL http://dl.acm.org/citation.cfm?doid=2600057.2602877.

Cramton, P.; Shoham, Y.; and Steinberg, R. 2004. Combinatorial Auctions. Technical Report 04mit, University of Maryland, Department of Economics - Peter Cramton. URL https://ideas.repec.org/p/pcc/pccumd/04mit.html.

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

Day, R. W.; and Cramton, P. 2012. Quadratic Core-Selecting Payment Rules for Combinatorial Auctions. *Operations Research* 60(3): 588–603. ISSN 0030-364X. doi:10.1287/opre.1110.1024. URL https://pubsonline.informs.org/doi/abs/10.1287/opre.1110.1024.

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., eds., *Advances in Neural Information Processing Systems 28*, 3061–3069. Curran Associates, Inc. URL http://papers.nips.cc/paper/6016-no-regret-learning-in-bayesian-games.pdf.

Kingma, D. P.; and Ba, J. 2017. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]* URL http://arxiv.org/abs/1412.6980. Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015.

Klambauer, G.; Unterthiner, T.; Mayr, A.; and Hochreiter, S. 2017. Self-Normalizing Neural Networks. *arXiv:1706.02515 [cs, stat]* URL http://arxiv.org/abs/1706.02515. Comment: 9 pages (+ 93 pages appendix).

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

Letcher, A.; Balduzzi, D.; Racaniere, S.; Martens, J.; Foerster, J.; Tuyls, K.; and Graepel, T. 2019. Differentiable Game Mechanics. *arXiv:1905.04926 [cs, stat]* URL http://arxiv.org/abs/1905.04926. Comment: JMLR 2019, journal version of arXiv:1802.05642.

Mertikopoulos, P.; and Zhou, Z. 2019. Learning in Games with Continuous Action Sets and Unknown Payoff Functions. *Mathematical Programming* 173(1): 465–507. ISSN

1436-4646. doi:10.1007/s10107-018-1254-8. URL https://doi.org/10.1007/s10107-018-1254-8.

Milgrom, P. 2017. *Discovering Prices: Auction Design in Markets with Complex Constraints*. Columbia University Press.

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. In *NIPS-W*.

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. ISSN 00043702. doi:10.1016/j.artint.2012.09.007. URL http://linkinghub.elsevier.com/retrieve/pii/S0004370212001191.

Roughgarden, T. 2016. Twenty Lectures on Algorithmic Game Theory — Algorithmics, Complexity, Computer Algebra and Computational Geometry. URL https://www.cambridge.org/de/academic/subjects/computer-science/algorithmics-complexity-computer-algebra-and-computational-g/twenty-lectures-algorithmic-game-theory,https://www.cambridge.org/de/academic/subjects/computer-science/algorithmics-complexity-computer-algebra-and-computational-g.

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]* URL http://arxiv.org/abs/1703.03864.

Schaefer, F.; and Anandkumar, A. 2019. Competitive Gradient Descent. In Wallach, H.; Larochelle, H.; Beygelzimer, A.; d\textquotesingle Alché-Buc, F.; Fox, E.; and Garnett, R., eds., *Advances in Neural Information Processing Systems 32*, 7625–7635. Curran Associates, Inc. URL http://papers.nips.cc/paper/8979-competitive-gradient-descent.pdf.

Silver, D.; Lever, G.; Heess, N.; Degris, T.; Wierstra, D.; and Riedmiller, M. 2014. Deterministic Policy Gradient Algorithms. In *International Conference on Machine Learning*, 387–395. URL http://proceedings.mlr.press/v32/silver14.html.

Ui, T. 2016. Bayesian Nash Equilibrium and Variational Inequalities. *Journal of Mathematical Economics* 63: 139–146. ISSN 0304-4068. doi:10.1016/j.jmateco.2016.02.004. URL http://www.sciencedirect.com/science/article/pii/S0304406816000124.

Vickrey, W. 1961. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of finance* 16(1): 8–37.

Zinkevich, M. 2003. Online Convex Programming and Generalized Infinitesimal Gradient Ascent. In *Proceedings of the Twentieth International Conference on International Conference on Machine Learning*, ICML'03, 928–935. Washington, DC, USA: AAAI Press. ISBN 978-1-57735-189-4.