Enabling First-Order Gradient-Based Learning for Equilibrium Computation in Markets

Nils Kohring [1] **Fabian R. Pieroth** [1] **Martin Bichler** [1]

Abstract

Understanding and analyzing markets is crucial, yet analytical equilibrium solutions remain largely infeasible. Recent breakthroughs in equilibrium computation rely on zeroth-order policy gradient estimation. These approaches commonly suffer from high variance and are computationally expensive. The use of fully differentiable simulators would enable more efficient gradient estimation. However, the discrete allocation of goods in economic simulations is a non-differentiable operation. This renders the first-order Monte Carlo gradient estimator inapplicable and the learning feedback systematically misleading. We propose a novel smoothing technique that creates a surrogate market game, in which first-order methods can be applied. We provide theoretical bounds on the resulting bias which justifies solving the smoothed game instead. These bounds also allow choosing the smoothing strength a priori such that the resulting estimate has low variance. Furthermore, we validate our approach via numerous empirical experiments. Our method theoretically and empirically outperforms zeroth-order methods in approximation quality and computational efficiency.

1. Introduction

Auctions are at the center of modern economic theory. Given some private valuation of goods available for purchase, participants must place bids on the market that maximize their expected payoff while remaining unaware of the other participants' valuations. In the seminal paper (Vickrey, 1961) the foundation for most auction theory results of today was laid. It is crucial to understand the strategic behavior in various auction applications, ranging from treasury and industrial procurement auctions to spectrum sales. Depending on the circumstances and behavioral assumptions, optimal strategies may differ drastically, starting from strategies, such as understating demand (bid-shading) (Krishna, 2009) and overstating demand (overbidding) (Ott & Beck, 2013), or much more convoluted strategies. However, computing such equilibria and approximations a priori remains challenging. Analytical equilibria can only be derived under strong assumptions such as in single-item auctions or the independent private values model.

A recent approach based on policy optimization uses randomized finite difference approximations of the gradient (Bichler et al., 2021). They proposed an algorithm called *neural pseudogradient ascent* (NPGA), which parametrizes the bidding strategies using neural networks and follows the approximate gradient dynamics of the game via simultaneous gradient ascent of all agents. The gradients are computed via *evolution strategies* (ES) (Salimans et al., 2017), which smoothen the objective by adding noise in the parameter space, thereby treating the environment as a black box. Compared with the well-known REINFORCE algorithm, where the actions are perturbed, this also results in zeroth-order gradient estimates with better precision and lower variance but much higher computational cost.

Under the differentiable programming paradigm, there is a growing interest in computing gradients for numerous reinforcement learning applications that allow for first-order gradient estimates. It is possible to create a full computational graph for applications with a certain amount of structure. First-order methods have the advantage of much lower variance, which leads to faster convergence rates to local minima of non-convex objective functions (Mohamed et al., 2020). However, there are two common problems in employing first-order methods. First, most reinforcement learning environments are provided only as black boxes. This implies that there is no explicit access to the underlying state transition function and the gradient can only be estimated by repeatedly evaluating the reward function. The wide applicability of zeroth-order policy optimization, like REINFORCE and more advanced actor-critic techniques (Schulman et al., 2017), contributes to their popularity. Sec-

[1]School of Computation, Information and Technology, Technical University of Munich. Correspondence to: Nils Kohring <nils.kohring@tum.de>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

ond, in some applications, such as the training of variational autoencoders, the computational path of the derivate is blocked (i.e., repeatably applying the chain rule to calculate the gradient of the reward with respect to the parameters of the policy) because it consists of sampling a random variable, which is a non-differentiable operation (Bangaru et al., 2021).

The situation is similar in auction games. The allocation of indivisible goods causes biased gradients of first-order methods. Example 1.1 showcases this observation. It was observed that the first-order Monte Carlo gradient estimate does not converge to equilibrium and quickly causes consistent zero-bidding (Bichler et al., 2021). From a mathematical standpoint, the single-sample (ex post) utility has a discontinuity. Thus, the sample mean of its exact gradients is an inadequate estimate for its true (ex ante) utility gradient (the expected utility over all possible valuations).

Example 1.1. Consider a first-price sealed-bid (FPSB) single-item auction. Two bidders compete for a single good, where the winner pays his or her bid. The derivative of the utility with respect to the bid is zero for losing bids and minus one for winning bids after a point of discontinuity. Either the bidder loses and receives no feedback or wins and could have won with an even smaller bid.

In this study, we propose transforming multi-agent auction games such that their utility functions are sufficiently regular for applying efficient first-order gradient methods while keeping the overall gradient dynamics close to the original game. In contrast to the original allocations of indivisible items, we use *soft* allocations instead. We effectively treat the items as divisible and allocate the proportional fraction of an item to the bidders based on their reported bids. An additional adaption to the pricing rule eliminates the discontinuity at the threshold of winning and losing an object. However, this comes at the expense of introducing a bias in the utility function. For example, a losing bidder has zero utility the original auction. However, in the smoothed auction, this bidder receives a small fraction of the good (and pays a correspondingly small price), such that the gradient indicates that a higher bid would have resulted in higher utility. The feedback to bid lower when winning remains of similar magnitude. Thus, there is always appropriate feedback on the current bidding strategy in the smoothed game. Figure 1 shows the utility function and its relaxed version.

This approach is applicable widely to economic models and general auction formats, such as sequential or simultaneous sales of multiple goods, as in combinatorial auctions with item bidding. It is further independent of the number of bidders, payment rule, risk preferences of the bidders, or correlations among the bidders' valuations. We demonstrate that the choice of a smoothing parameter follows

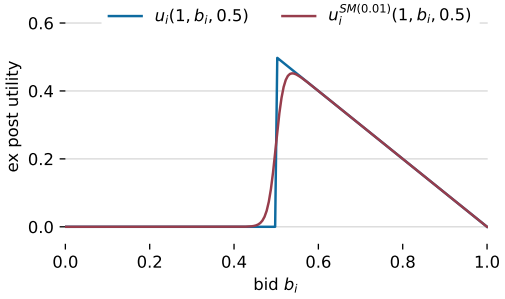

Figure 1. Ex post utility in the original FPSB single-item auction (blue) and its smoothed version (red) for a temperature of $\lambda = 0.01$ and a highest opponent bid of 0.5. The utility in the original auction is zero for losing bids and decreases linearly for winning bids.

a natural trade-off. Importantly, computing equilibria in multi-agent games is not straightforward and many negative results are known (Chasnov et al., 2020; Mazumdar et al., 2020; Letcher, 2020). Therefore, changes to the game dynamics must be implemented with great caution, and we can prove that an approximate equilibrium in the smoothed game still constitutes an approximate equilibrium in the original game.

Computational-wise, the smoothing only comes with the cost of tracking the gradients of the individual operations, upon which the game dynamics are built. Compared with NPGA, learning in the *smooth market* (SM) via the first-order estimator is more than ten times faster while yielding better results. For example, an iteration of NPGA in a small single-item auction with the default hyperparameters from (Bichler et al., 2021) takes approximately $0.16\,\mathrm{s}$, whereas first-order policy gradients applied to the SM take an average of $0.01\,\mathrm{s}$.

Our contribution can be summarized as follows: We introduce the SM and show that first-order methods provide an unbiased estimator of the utility gradient of the SM game. Furthermore, we provide theoretical guarantees showing that policy improvements in the SM result in improvements in the original game, and we provide theoretical and empirical insights showing that the empirical variance can be controlled. Finally, we demonstrate a substantial improvement to previous methods in performance and computational speed via multiple experiments.

2. Related Work

The theory of learning in games largely considers complete-information finite games, hence, traditional techniques rely on discretization. However, it is unclear how well a discretized strategy performs in the original continuous game

in general (Waugh et al., 2009) and it suffers from the curse of dimensionality. The first attempts to compute equilibria in imperfect-information auction games followed such an approach (Athey, 2001) or expressed the game as a limit of a sequence of complete-information games (Armantier et al., 2008). In larger combinatorial auctions equilibria were first computed with an algorithm that computes pointwise best responses in a discretization of the strategy space via Monte Carlo integration (Bosshard et al., 2020). Besides the aforementioned NPGA, an approach that similarly learns continuous-action strategies was proposed (Li & Wellman, 2021). Both algorithms learn bid functions via zeroth-order gradient estimates that are used during simultaneous gradient ascent in self-play. Our method considers a continuous surrogate game and enables the use of first-order gradient methods.

The idea of analytically smoothing markets is conceptually similar to that of differentiable physics simulations. Smooth approximations of the underlying dynamics were used in these simulations (Huang et al., 2021). Zeroth- and first-order methods were compared and the pros and cons of both when available were discussed (Suh et al., 2022). Furthermore, they demonstrated that the presence of discontinuities in the objective causes the first-order estimator to be biased, whereas the zeroth-order estimator remains unbiased. Smooth markets transfer these ideas to auctions.

3. Preliminaries

We restrict the formulations to the case of single-item auctions for brevity in the presentation. The extension to *auctions of multiple independent items* is straightforward and we present some experimental results for both cases.

3.1. Auctions as Bayesian Games

A *Bayesian auction game* is defined as a quintuple $G = (\mathcal{I}, \mathcal{A}, \mathcal{V}, F, u)$. $\mathcal{I} = \{1, \dots, n\}$ describes the set of bidders participating in the game. The set of possible bid profiles is given as $\mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$, where $\mathcal{A}_i$ is the set of bids available to agent $i \in \mathcal{I}$. Whereas $\mathcal{V} = \mathcal{V}_1 \times \cdots \times \mathcal{V}_n$ is the set of *valuation profiles*. $F\colon \mathcal{V} \to [0, 1]$ defines the joint prior probability distribution over valuation profiles, which is assumed to be common knowledge among all agents and atomless. F_i denotes agent i's marginal distribution of valuations. In this study, the index $\text{-}i$ denotes a profile of valuations, bids, or strategies for all bidders, except bidder i.

At the beginning of the game, nature draws a valuation profile $v \sim F$, and each agent i is informed of his or her valuation $v_i \in \mathcal{V}_i$. We denote by F_i the marginal distribution of bidder i and by $F_{\text{-}i|i}$ the conditional distribution of the opponents given v_i. Based on the drawn valuation v_i, each agent submits a bid b_i according to the *strategy*, *policy*, or *bid function* $\beta_i\colon \mathcal{V}_i \to \mathcal{A}_i$. We denote the resulting strategy space of bidder i as $\Sigma_i \subseteq \mathcal{A}_i^{\mathcal{V}_i}$ and the space of possible joint strategies as $\Sigma = \prod_i \Sigma_i$.

As part of the environment, the auctioneer collects these bids and applies an *auction mechanism* that determines allocations $x_i \in \{0, 1\}$ for each bidder i, such that the item is allocated to at most one bidder. Also, it determines payments $p(b) \in \mathbb{R}^n_{\geq 0}$ according to a payment rule p, which the agents must pay to the auctioneer. We will consider bidders with risk-neutral utility functions given by $u_i\colon \mathcal{V}_i \times \mathcal{A} \to \mathbb{R}$,

$$u_i(v_i, b) = v_i\, x_i(b) - p_i(b) \tag{1}$$

$$= \begin{cases} v_i - p_i(b) & b_i > \max b_{\text{-}i}, \\ 0 & \text{else}, \end{cases} \tag{2}$$

i.e., the players' utility is given by how much they value the good allocated to them minus the price to be paid. We will also write $u_i(v_i, b_i, b_{\text{-}i}) = u_i(v_i, b)$ with a slight abuse of notation. Thus, the bidders' utilities depend on all bidders' actions but only on their own valuations. They aim to maximize their utility u_i. We omit bidders with risk aversion or other forms of utility and valuation correlations for brevity. Notwithstanding, our treatment of equilibrium computation also extends to these settings. We will differentiate between the *ex ante* state of the game, where bidders know only the prior F, the *interim* state, where bidders additionally know their valuation $v_i \sim F_i$, and the *ex post* state, where all bids have been submitted; thus, $u_i(v_i, b)$ can be evaluated.

3.2. Equilibria

Nash equilibria (NE) are often regarded as the central solution concept in game theory. Informally, given the equilibrium strategy of the opponents in an NE, no agent has an incentive to unilaterally deviate. *Bayesian Nash equilibria* (BNE) extend this concept to games of incomplete information. Here, the expected utility over the distribution of opponent valuations is calculated instead. For a private valuation $v_i \in \mathcal{V}_i$, bid $b_i \in \mathcal{A}_i$, and opponent strategies $\beta_{\text{-}i} \in \Sigma_{\text{-}i}$, we denote the *interim utility* of bidder i as

$$\overline{u}_i(v_i, b_i, \beta_{\text{-}i}) \;=\; \mathbb{E}_{v_{\text{-}i}|v_i}[u_i(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))], \tag{3}$$

where $v_{\text{-}i}|v_i$ denotes the expectation over the opponent's conditional prior distribution given the valuation v_i. We also denote the *interim utility loss* of bid b_i incurred by not playing a best response, given v_i and $\beta_{\text{-}i}$ by:

$$\overline{\ell}_i(v_i, b_i, \beta_{\text{-}i}) \;=\; \sup_{b_i' \in \mathcal{A}_i} \overline{u}_i(v_i, b_i', \beta_{\text{-}i}) - \overline{u}_i(v_i, b_i, \beta_{\text{-}i}). \tag{4}$$

An *ε-Bayes Nash equilibrium* (ε-BNE) with $\varepsilon \geq 0$ is a strategy profile $\beta^* = (\beta_1^*, \dots, \beta_n^*) \in \Sigma$, such that no bidder can improve his or her interim expected utility more than ε

by deviating. Therefore, in an ε-BNE for all $i \in \mathcal{I}$, it holds that

$$\sup_{v_i \in \mathcal{V}_i} \overline{\ell}_i(v_i, \beta_i^*(v_i), \beta_{\text{-}i}^*) \;\leq\; \varepsilon. \tag{5}$$

A 0-BNE is simply called a BNE. In a BNE, every bidder's strategy maximizes his or her expected interim utility across his or her valuation space, given the opponents' strategies. While BNEs are often defined at the *interim* stage of the game, we also consider *ex ante* equilibria as strategy profiles that concurrently maximize each bidder's *ex ante* utility

$$\tilde{u}_i(\beta_i, \beta_{\text{-}i}) \;=\; \mathbb{E}_{v_i}[\overline{u}_i(v_i, \beta_i(v_i), \beta_{\text{-}i})]. \tag{6}$$

To estimate the worst-case interim utility loss $\overline{\ell}_{\text{max}}$, we choose an equidistant grid of n_{grid} alternative actions ranging from zero to the maximum valuation for all dimensions and calculate approximate best responses based on the average utility over a sample of n_{batch} prior distributions. Taking the maximum over all valuations and bidders then gives an estimate of $\overline{\ell}_{\text{max}}$, bounding ε for the ex ante case from above.

As a second metric, we additionally report the probability-weighted root mean squared error of the learned strategy β_i to the exact BNE strategy β_i^* for those settings where an analytical BNE is known. For a sample from the prior valuation of size n_{batch}, this approximates the L_2 distance $\|\beta_i - \beta_i^*\|_{\Sigma_i}$ of these two functions as

$$L_2(\beta_i) \;=\; \left(\frac{1}{n_{\text{batch}}} \sum_{v_i} \left(\beta_i(v_i) - \beta_i^*(v_i)\right)^2 \right)^{1/2}. \tag{7}$$

Unlike $\overline{\ell}_{\text{max}}$, this metric is much easier to compute and does not suffer the drawback that a strategy with a negatable small loss may still be arbitrarily distant from the actual BNE. However, it is only computable when an analytical BNE is available and may need multiple evaluations when there are multiple BNE.

3.3. Gradient Optimization Methods

Policy gradient methods are concerned with learning a parameterized policy β_{θ_i} that selects actions based on the current observations (Sutton & Barto, 2018). To maximize utility, bidder i updates the parameters θ_i according to gradient ascent. This process is intended to compute approximate ex ante BNEs, that is, to find mutual best responses of the bidders for all possible valuations. The exact gradient update for valuation v_i in iteration t is

$$\theta_i^t \;=\; \theta_i^{t-1} + \eta \cdot \nabla_{\theta_i^{t-1}} \overline{u}_i(v_i, \beta_{\theta_i^{t-1}}(v_i), \beta_{\theta_{\text{-}i}^{t-1}}). \tag{8}$$

This must be approximated in practice. Two common methods are zeroth- and first-order gradient approximations. The former solely relies on evaluating the objective function u_i, whereas the gradient $\nabla_{\theta_i} u_i$ can be evaluated in the latter.

As stated in the introduction, the discontinuous nature of the ex post utility function stems from the sampling of the opponents' priors and their corresponding actions. We encounter u_i from Equation 2 and its derivative (in general) is discontinuous in b_i. The observation of this inapplicability persists for all pricing regimes and behavioral assumptions that are commonly considered in auctions. Thus, an unbiased gradient estimate of the interim utility function *cannot* be derived by sampling the ex post gradient. Specifically, interchanging taking an expectation and differentiating is invalid:

$$\nabla_{\theta_i} \mathbb{E}_{v_{\text{-}i}|v_i}[u_i] \;\neq\; \mathbb{E}_{v_{\text{-}i}|v_i}[\nabla_{\theta_i} u_i]. \tag{9}$$

We supply the mathematical details in Appendix A. Therefore, the naive application of backpropagating the accumulated exact ex post gradients may not be expected to provide a meaningful estimate of the ex ante gradient. This study establishes a path towards valid first-order gradient estimates in auction games.

3.4. Zeroth-Order Approximation Methods

(Bichler et al., 2021) employed ES to circumvent the interchange of differentiation and integration. ES rely on a randomized finite difference approximation of the gradient based on perturbations in the parameter space of the neural networks which can be computed after averaging over the priors (Salimans et al., 2017). This is an alternative zeroth-order method to the REINFORCE algorithm. Unlike ES, REINFORCE relies on perturbations in the action space by using mixed strategies (typically Gaussian distributions) such that the gradient of the action probability density can be approximated. (Salimans et al., 2017) compared these estimates for RL applications and argued that the variance of the ES estimate can be significantly lower. We overload the notation for the ease of readability and write $u_i(\theta_i, v_{\text{-}i}) = u_i(v_i, \beta_{\theta_i}(v_i), \beta_{\theta_{\text{-}i}}(v_{\text{-}i}))$. For a hyperparameter $\sigma > 0$, the ES estimator can be derived from

$$\begin{aligned} &\nabla_{\theta_i} \mathbb{E}_{v_{\text{-}i}|v_i}\left[u_i(\theta_i, v_{\text{-}i})\right] \\ &\quad\approx \nabla_{\theta_i} \mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)} \mathbb{E}_{v_{\text{-}i}|v_i}\left[u_i(\theta_i + \sigma\epsilon, v_{\text{-}i})\right] &(10)\\ &\quad= \mathbb{E}_{\epsilon\sim\mathcal{N}(0,I)} \mathbb{E}_{v_{\text{-}i}|v_i}\left[\frac{\epsilon}{\sigma} u_i(\theta_i + \sigma\epsilon, v_{\text{-}i})\right]. &(11)\end{aligned}$$

The last term can now be approximated via sampling. However, the ES gradient estimate comes at massive computational costs. It requires a large number of additional environment evaluations for the sampled population values of ϵ. Parallelization is essentially unavailable, because it would reduce the number of samples from the prior when considering a fixed amount of memory. Latter of which is the main limiting factor in getting precise estimates of the expected utility in auction games. Thus, (Bichler et al., 2021) kept a large batch size and computed the ES sequentially using

a default population size of 64. Based on the variance of the estimate, (Salimans et al., 2017) argued that ES are an attractive choice if the number of episodes is large, which is not the case for single-round auctions.

4. Smoothing Single-Item Auctions

This section proposes the market-specific approach.

4.1. Allocation and Price Smoothing

The allocation of indivisible objects in auction games is typically modeled as a binary vector, with a one indicating that the item is allocated to the corresponding buyer. The set of legitimate allocations is defined as

$$\mathcal{X} = \Big\{ x \in \{0,1\}^n \;\Big|\; \sum_{i=1}^{n} x_i \leq 1 \Big\}. \tag{12}$$

For all commonly considered auctions, the allocations label the bids as winning or losing to maximize the auctioneer's revenue. They are calculated according to

$$x(b) = \operatorname*{arg\,max}_{x' \in \mathcal{X}} \sum_{i=1}^{n} b_i x_i'. \tag{13}$$

Typical auction mechanisms only differ in their payment rules. Two noteworthy examples are the first-price mechanism, where bidders pay what they bid and the celebrated VCG mechanism (second-price), where they pay for the harm they cause others by competing (Krishna, 2009).

These allocations result in the utilities not being continuous. Therefore, we propose relaxing the calculation of the allocations using the softmax function as a surrogate for the argmax operation:

$$x_i^{\mathrm{SM}(\lambda)}(b) = \frac{\exp\left(\frac{b_i}{\lambda}\right)}{\sum_{j=1}^{n} \exp\left(\frac{b_j}{\lambda}\right)}, \quad i = 1, \dots n. \tag{14}$$

The temperature $\lambda > 0$ denotes the smoothing strength. This can be interpreted as dividing the item among all bidders according to their proportional bid magnitudes, where $\sum_i x_i^{\mathrm{SM}(\lambda)}(b) = 1$ remains valid. The softmax asymptotically recovers the true argmax as λ approaches zero. As we are interested in a continuous utility surface, the discontinuity in the prices (only the winners pay) must also be considered. An obvious choice is to calculate the original prices of the good and then distribute the price according to the fractional allocations $x^{\mathrm{SM}(\lambda)}$:

$$p^{\mathrm{SM}}(b) = \sum_{j=1}^{n} p_j(b). \tag{15}$$

Hence, the ex post utility in the relaxed game takes the form

$$u_i^{\mathrm{SM}(\lambda)}(v_i, b) = \big(v_i - p^{\mathrm{SM}}(b)\big)\, x_i^{\mathrm{SM}(\lambda)}(b). \tag{16}$$

By definition, we have almost everywhere (a.e.) pointwise convergence of $x_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}(\,\cdot\,))$ to $x_i(v_i, b_i, \beta_{\text{-}i}(\,\cdot\,))$ as functions of $v_{\text{-}i}$, except at $b_i = \max b_{\text{-}i}$. Furthermore, the fractional prices $p^{\mathrm{SM}(\lambda)}(b_i, \beta_{\text{-}i}(\,\cdot\,))$ also converge a.e. pointwise to $p_i(b_i, \beta_{\text{-}i}(\,\cdot\,))$. Thus, the ex post utilities are recovered (a.e.) for ever smaller temperature. The resulting utilities are visualized for the special case of an FPSB auction (Figure 1). Throughout the rest of the article, we make the following regularity assumptions.

Assumption 4.1. Consider a Bayesian auction game G and assume:

1. The action $\mathcal{A}_i$ and valuation spaces $\mathcal{V}_i$ are compact intervals.
2. F is an atomless prior.
3. The bidding and pricing functions are measurable.

We regain continuity of the ex post utility and its gradient by this smoothing of allocations and payments. Specifically, we have the following theorem:

Theorem 4.2. *Let the conditions of Assumption 4.1 hold and assume the pricing function p^{SM}, the marginal density functions $\{f_{\text{-}i|i}\}_{v_i \in \mathcal{V}_i, i \in \mathcal{I}}$, and strategies $\{\beta_i\}_{i \in \mathcal{I}}$ to be Lipschitz continuous. Then, the estimator on the smooth interim utility's gradient by sampling from the smoothed ex post utilities' gradients is unbiased, i.e.,*

$$\nabla_{\theta_i} \overline{u}_i^{SM}(v_i, b_i) = \mathbb{E}_{v_{\text{-}i}|v_i}[\nabla_{\theta_i} u_i^{SM}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))], \tag{17}$$

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

We refer to Appendix A for the proof. Importantly, this relaxation technique is applicable to general markets with different payment rules, utility functions, or correlated priors. Compared with the ES gradient estimate, where the parameter space is perturbed, the SM gradient estimate perturbs the utility function. Thus, the origin of bias is different and can be controlled by σ for ES and by λ for SM.

4.2. Approximation Quality

We check the validity of the smoothing intervention by ensuring that the error to the original game dynamics can be controlled by choosing a sufficiently small value of λ. This ensures that conducting policy optimization in the smoothed game can be expected to result in policy improvements in the original game. Furthermore, this will clarify the question of an optimal choice of the temperature value.

Generally, analytically computing equilibria of the SM game is infeasible. Instead, we focus on comparing the expected interim and ex ante utilities in the original and SM game. A small error implies similar utility surfaces and gradient dynamics. Note that the ex post utilities can be quite different.

Suppose multiple bidders compete for a single commodity and bidder i has approximately the same bid magnitude as the strongest opponent. The smoothed allocation is close to one-half, whereas the true allocation is either zero or one. This would result in a significant difference in the ex post utility driven by the magnitude of the utility discontinuity in the original auction. The probability of such large errors decreases with smaller smoothing factors; however, this event cannot be completely ruled out. We verify in the following theorem, that the error in expected interim and ex ante utility approaches zero under mild assumptions on the auction format.

Theorem 4.3. *Let the conditions of Assumption 4.1 hold and suppose the payment rule p is bounded. Then, for bidder i, we have convergence in interim and ex ante utility:*

1. *Let $v_i \in \mathcal{V}_i$ and $b_i \in \mathcal{A}_i$, then*

$$\lim_{\lambda \to 0} \overline{u}_i^{SM(\lambda)}(v_i, b_i, \beta_{-i}) \;=\; \overline{u}_i(v_i, b_i, \beta_{-i}). \tag{18}$$

2. *Further assume β_i to be measurable. Then,*

$$\lim_{\lambda \to 0} \tilde{u}_i^{SM(\lambda)}(\beta_i, \beta_{-i}) \;=\; \tilde{u}_i(\beta_i, \beta_{-i}). \tag{19}$$

The proof is delegated to Appendix B. Theorem 4.3 ensures that for ever smaller λ, the bias in the expected utilities vanishes compared with the utilities in the original game. This implies that the smoothed gradients converge, thus justifying gradient-based learning in the perturbed game. Although Theorem 4.3 ensures convergence, it does not state how fast the error approaches zero. However, this information is crucial for practical applications. Therefore, we make the following additional assumptions on the auction format.

Assumption 4.4. For all $i \in \mathcal{I}$ assume:

1. β_i is strictly increasing and Lipschitz continuous.
2. β_i^{-1} is Lipschitz continuous.
3. There exists a uniform bound for all marginal conditional prior density functions $f_{i|\,.}$.
4. p_i is bounded.

Note that assuming Lipschitz continuous strategies is satisfied by common function approximations, e.g., neural networks. With these stronger assumptions, we can present a worst-case convergence rate of the interim and ex ante utility errors.

Proposition 4.5. *Consider an auction with n bidders that satisfies Assumptions 4.1 and 4.4. Then, the absolute interim and ex ante utility errors are of order $\mathcal{O}(\lambda)$.*

Proof Sketch. Use substitution on the opponents' bidding strategies, followed by iterated use of Hölder's inequality. The details of the proof can be found in Appendix C. □

Note that Restrictions 1 and 4 in Assumption 4.4 are standard in the literature (Krishna, 2009). Restriction 2 is slightly stronger by demanding that strategy β_i cannot become infinitely flat (e.g., a saddle-point would not be allowed). However, this restriction can be somewhat lifted resulting in a worse convergence rate. Details on this can be found in Appendix C. Finally, Restriction 3 holds for all commonly used prior distributions, however, it rules out perfect correlation. Based on the previous result, we can characterize how a learned ε-BNE of the SM game translates to an approximate BNE the original game:

Theorem 4.6. *In an auction with n bidders that satisfies Assumptions 4.1 and 4.4, let β^* be an ex ante ε-BNE in the smoothed game with smoothing parameter λ. Then β^* is an ex ante $\varepsilon + \mathcal{O}(\lambda)$-BNE of the original game.*

The proof can be found in Appendix D. The derived bounds in the previous results consider worst-case scenarios. However, we observed that the error may be significantly lower in practice. To rationalize this observation, we compare the worst-case bound to the exact error in a restricted setting. Consider an FPSB auction with two bidders, independent uniform priors, and a linear bidding function of the second bidder, $\beta_2(v_2) = sv_2 + t$. Then, the bound derived in Proposition 4.5 translates to

$$\left|\tilde{u}_1^{\mathrm{SM}(\lambda)}(\beta_1, \beta_2) - \tilde{u}_1(\beta_1, \beta_2)\right| \;\leq\; \frac{\ln(2)+1}{s}\lambda. \tag{20}$$

In Figure 2, we compare this bound (for bidder 2's BNE strategy with $s = 0.5$ and $t = 0$) to the exact interim utility error, which can be derived for this restricted setting (see Appendix E). The convergence rate of the interim utilities depends on the specific prior sample v_1 and bid b_1. The ex ante utilities converge more rapidly than predicted by the worst-case bound. We conjecture that this often holds in practice, resulting in better learning behavior than suggested by Proposition 4.5.

4.3. Choosing the Smoothing Temperature

Let us consider the question of an optimal smoothing strength. There is an incentive to keep temperature values as low as possible, such that the original game dynamics are distorted as little as possible. On the other hand, one does not want to decrease λ too low, as this causes numerical problems. The magnitude of the gradient goes towards infinity at the former discontinuity as λ decreases. Therefore, with finite sample size, the first-order gradient estimate might have a high empirical variance (Suh et al., 2022).

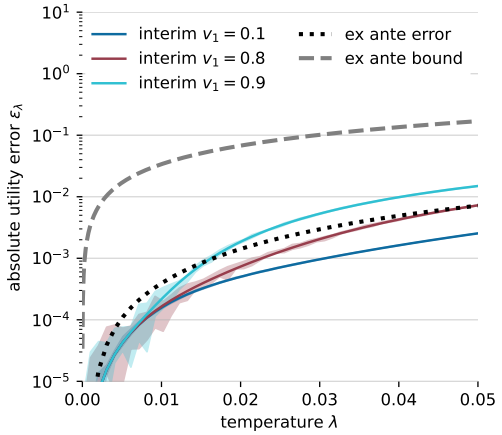

Figure 2. Comparison of the absolute utility errors. (i) The linear ex ante bound that holds for all valuations (gray dashed line) from Equation 20. (ii) Exact interim utility errors when both bidders act according to the BNE for some exemplary valuations (colorized lines) and their sampled mean values ± standard deviation (shaded areas). (iii) The approximate ex ante error (black dotted line).

We propose to use the utility sampling precision as a natural way to choose the temperature. For the special case presented in Figure 2 and the default batch size of 2^{18}, one can see that the sample precision is reached at about 10^{-4}. That is, for a drawn batch, the Monte Carlo estimation of ex ante utilities has a precision of about 10^{-4}, and we can no longer distinguish between the smoothed and original utilities. Therefore, one can use Proposition 4.5 to derive a lower bound for λ for a given sampling precision. As discussed at the end of Section 4.2, the true ex ante utility error is usually lower, so that one can choose a higher λ without losing any performance.

The empirical sampling precision is affected by several factors, such as the valuation and bidding ranges, the number of bidders, prior distributions, and complexity of bidding functions. Some of these influences can be standardized, e.g., by normalizing the bidding ranges. Ultimately, a sufficiently high batch size can overcome any bias introduced by aforementioned factors, such that it should be chosen as high as computationally possible to achieve an optimal sampling precision.

5. Empirical Results

We provide experimental evaluation of the new technique and compare the results with those of NPGA and REINFORCE by measuring how closely they approximate the analytical BNE. Results for settings with risk aversion or correlated valuations are similar and omitted for simplicity. Furthermore, we provide some insights and guidance on appropriate choices of λ and verify that our gradient estimate's variance is sufficiently small. We list all hyperparameters and details on the network architecture in Appendix G.

Table 1. Learning results in FPSB and SPSB auctions with different numbers m of items. We report the mean values of the L_2 and $\overline{\ell}_{\max}$ losses (smaller is better) and the time per iteration across five runs. We also report the standard deviation in parentheses for the losses.

	m	Algorithm	L_2	$\overline{\ell}_{\max}$	t/iter
FPSB	1	NPGA	0.011 (0.005)	0.005 (0.002)	0.155
		REINFORCE	0.021 (0.008)	**0.003 (0.000)**	**0.009**
		SM	**0.005 (0.003)**	0.004 (0.002)	**0.009**
	2	NPGA	0.013 (0.005)	0.010 (0.002)	0.150
		REINFORCE	0.041 (0.020)	0.016 (0.010)	**0.009**
		SM	**0.008 (0.002)**	**0.006 (0.003)**	**0.009**
	4	NPGA	0.028 (0.002)	0.021 (0.003)	0.148
		REINFORCE	0.064 (0.018)	0.039 (0.012)	**0.009**
		SM	**0.015 (0.004)**	**0.011 (0.004)**	**0.009**
	8	NPGA	0.104 (0.054)	0.127 (0.109)	0.206
		REINFORCE	0.187 (0.073)	0.331 (0.169)	**0.012**
		SM	**0.036 (0.003)**	**0.034 (0.009)**	**0.012**
SPSB	1	NPGA	0.012 (0.001)	0.002 (0.000)	0.170
		REINFORCE	0.028 (0.005)	0.002 (0.000)	**0.009**
		SM	**0.004 (0.001)**	**0.001 (0.000)**	0.011
	2	NPGA	0.018 (0.002)	0.003 (0.000)	0.264
		REINFORCE	0.082 (0.020)	0.009 (0.002)	0.011
		SM	**0.007 (0.001)**	**0.002 (0.000)**	**0.015**
	4	NPGA	0.043 (0.002)	0.011 (0.003)	0.457
		REINFORCE	0.140 (0.045)	0.028 (0.018)	**0.017**
		SM	**0.029 (0.003)**	**0.006 (0.002)**	0.024
	8	NPGA	0.214 (0.112)	0.299 (0.238)	0.869
		REINFORCE	0.320 (0.128)	0.262 (0.174)	**0.031**
		SM	**0.074 (0.002)**	**0.020 (0.002)**	0.043

5.1. Single-Item Auctions

For the two common payment rules of FPSB and second-price sealed-bid (SPSB) and a uniform prior on $[0, 1]$, we can measure the distance in action space to the unique BNE, as described in Equation 7 and compute an estimate of exploitability in the form of Equation 5. Table 1 shows the results. The losses are computed after training 2,000 iterations with each algorithm. The time per iteration, t/iter, decreases notably when comparing NPGA to SM across both payment rules, while also achieving a better approximation quality. Since the estimation of $\overline{\ell}_{\max}$ relies on a discretization of the action space and an exhaustive search thereon, L_2 detects smaller deviations, ceteris paribus. Although REINFORCE has a low iteration time, it is unable to learn high quality strategies due to its high variance (Section 5.3). We found that results for auctions with interdependent prior valuations or risk-aversion are quantitatively consistent with the results presented here.

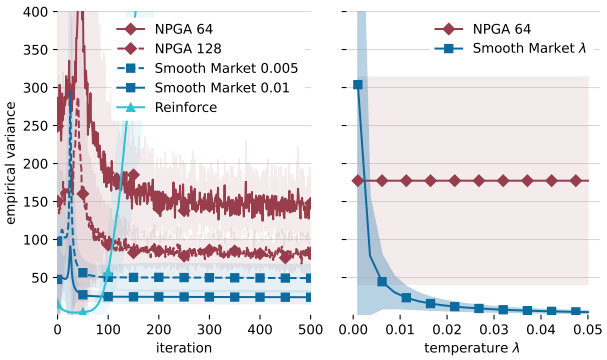

Figure 3. Empirical variance of the NPGA and REINFORCE zeroth-order and the SM first-order gradient estimates. Both zeroth-order methods are run in the original auction game. The mean values ± standard deviations over five runs each are depicted. *Left:* Comparing the variance throughout the learning procedure. *Right:* Comparing the variance for different smoothing temperatures (averaged over complete training runs).

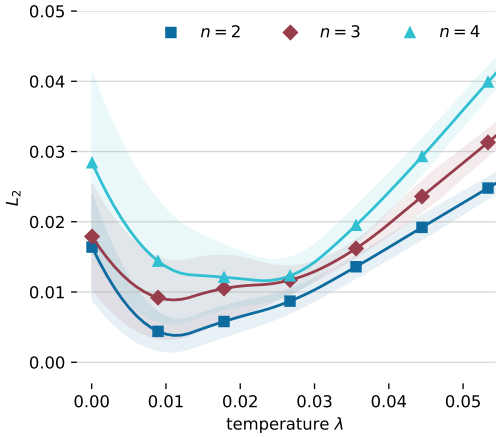

Figure 4. Action space distance for learned strategies to the BNE for different numbers of bidders and temperature values λ. The mean values ± standard deviations over five runs each are depicted.

5.2. Large Simultaneous Auctions

Furthermore, we study the separate sales of up to $m = 8$ distinctive goods and an increase in the number of bidders of up to $n = 4$. For simplicity, we do not consider any synergy effects on the items (this would include cases such as those where a bidder only values the bundle of two items but not either one of them individually), such that the BNE simplifies to the single-item strategy profile for each item separately. There are multiple motivations for these auctions. They can be considered as the base case of combinatorial auctions with item bidding and as a simple and practical alternative to full combinatorial auctions. Furthermore, combinatorial auctions with item bidding are being deployed, e.g., a bidder who is interested in a bundle of objects in parallel online display ad auctions or on a consumer shopping website is implicitly partaking in these auctions. Finally, asking a bidder to submit bids on all possible combinations of bundles ($2^m - 1$) is practically infeasible and there are positive results on the welfare properties of limiting the action space in this way (Bhawalkar & Roughgarden, 2011). Again, we draw i.i.d. uniform valuations on $[0, 1]$ and consider the FPSB and SPSB auctions. Learning in the SM game outperforms both previous approaches (Table 1). Since first-order methods are generally faster, we assume that the strong results in these settings will scale to even larger ones.

5.3. Empirical Variance

As stated in Section 4.3, there is a trade-off between low and high values of λ. Here, we consider the base setting of two bidders competing in a single-item FPSB auction. We decrease the batch size to 2^{16} as the single-sample gradients require more memory. Considering NPGA that is based on a sample of 64 evaluations of the objective by default, the empirical variance of the SM estimate is lower for all $\lambda > 0.002$ (compare intersection of Figure 3, right plot). Even after increasing NPGA's population size by a factor of two (which scales the run time in the same way), SM's variance remains lower for most choices, as can be seen in the left figure. The empirical variance of REINFORCE rapidly increases as the mixed-strategies get closer to the pure-strategy BNE. This degradation is to be expected when the learned variance of the Gaussian distributed actions decreases, see Exercise 13.4 of (Sutton & Barto, 2018).

Results for markets of different sizes are depicted in Figure 4. Keeping everything else fixed, the highest achievable performance decreases for larger markets, as is expected in multi-agent learning. The optimal smoothing strength is only affected indirectly via the bid magnitudes. At last, we note that the performance boost of larger batch sizes diminishes and best results are achieved for similar values of λ just below 0.01, indicating that the variance of the gradient estimate counteracts the lower bias. The results are presented in Appendix F.

6. Conclusion and Future Work

How can first-order gradient estimation methods be successfully applied to learning in auctions? We showed that our proposed smooth game formulation of strategic interactions in auctions provides a strong answer to this question. We established theoretical bounds on the bias caused by the smoothing, and an empirical evaluation verified that the variance of the gradient estimate can be controlled, leading

to low computational costs and high precision. Overall, we verified that equilibrium computation in smooth markets via fist-order gradient estimation is more efficient than previous learning methods.

Acknowledgements

We are grateful for funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation; grant no. BI 1059/I-9). We thank M. Oberlechner for valuable feedback on the research idea and S. Heidekrüger for the programming framework on top of which we have implemented our approach.

References

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

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

Bangaru, S. P., Michel, J., Mu, K., Bernstein, G., Li, T.-M., and Ragan-Kelley, J. Systematically differentiating parametric discontinuities. *ACM Transactions on Graphics (TOG)*, 40(4):1–18, 2021.

Bhawalkar, K. and Roughgarden, T. Welfare guarantees for combinatorial auctions with item bidding. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pp. 700–709. SIAM, 2011.

Bichler, M., Fichtl, M., Heidekrüger, S., Kohring, N., and Sutterer, P. Learning equilibria in symmetric auction games using artificial neural networks. *Nature Machine Intelligence*, 3(8):687–695, 2021.

Bogachev, V. I. *Measure Theory*, volume 1. Springer Science & Business Media, 2007.

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

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

Huang, Z., Hu, Y., Du, T., Zhou, S., Su, H., Tenenbaum, J. B., and Gan, C. Plasticinelab: A soft-body manipulation benchmark with differentiable physics. In *International Conference on Learning Representations*, 2021.

Katz, V. J. Change of variables in multiple integrals: Euler to cartan. *Mathematics Magazine*, 55:3–11, 1982. ISSN 0025-570X.

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

Letcher, A. On the impossibility of global convergence in multi-loss optimization. In *International Conference on Learning Representations*, 2020.

Li, Z. and Wellman, M. P. Evolution strategies for approximate solution of bayesian games. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 5531–5540, 2021.

Mazumdar, E., Ratliff, L. J., Jordan, M. I., and Sastry, S. S. Policy-gradient algorithms have no guarantees of convergence in linear quadratic games. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, pp. 860–868, 2020.

Mohamed, S., Rosca, M., Figurnov, M., and Mnih, A. Monte carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

Ott, M. and Beck, M. Incentives for overbidding in minimum-revenue core-selecting auctions. 2013.

Salimans, T., Ho, J., Chen, X., Sidor, S., and Sutskever, I. Evolution Strategies as a Scalable Alternative to Reinforcement Learning. *ArXiv*, March 2017.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., and Klimov, O. Proximal policy optimization algorithms. *ArXiv*, abs/1707.06347, 2017.

Suh, H. J., Simchowitz, M., Zhang, K., and Tedrake, R. Do differentiable simulators give better policy gradients? In *International Conference on Machine Learning*, pp. 20668–20696. PMLR, 2022.

Sutton, R. S. and Barto, A. G. *Reinforcement learning: An introduction*. MIT press, 2018.

Talvila, E. Necessary and sufficient conditions for differentiating under the integral sign. *The American Mathematical Monthly*, 108(6):544–548, 2001.

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

Waugh, K., Schnizlein, D., Bowling, M. H., and Szafron, D. Abstraction pathologies in extensive games. *AAMAS (2)*, 2009:781–8, 2009.

A. Proof of Theorem 4.2

The Leibniz integral rule states conditions for which the operator interchange of taking the limit and integrating is valid. Let us recall its measure theory variant using the notation of our application. For a rigorous treatment of the assumptions and different variants, we refer to the work of Talvila (2001), see Corollaries 5 and 8. We reformulate Condition 1 in our version using Fubini's Theorem.

Theorem A.1 (Leibniz integral rule)**.** *Let $[a, b] \subset \mathbb{R}$ and Ω with μ be a probability space. Suppose $f : [a, b] \times \Omega \to \mathbb{R}$ satisfies the conditions:*

1. *$f(x, \omega)$ is a measurable function in x and ω, and is integrable over $\omega \in \Omega$, for almost all $x \in [a, b]$.*
2. *For almost all $\omega \in \Omega$, $f(x, \omega)$ is absolutely continuous in x.*
3. *For all compact intervals $[c, d] \subset [a, b]$:*

$$\int_c^d \int_\Omega \left| \frac{\partial}{\partial x} f(x, \omega) \right| \mathrm{d}\mu(\omega) \mathrm{d}x < \infty. \tag{21}$$

Then

$$\frac{\partial}{\partial x} \int_\Omega f(x, \omega)\, \mathrm{d}\mu(\omega) \;=\; \int_\Omega \frac{\partial}{\partial x} f(x, \omega)\, \mathrm{d}\mu(\omega) \qquad \textit{a.e.} \tag{22}$$

Equation 22 is assumed to hold for backpropagation. It is needed for the approximation of ex ante gradients based the sample mean of ex post gradients (compare Equation 9 from main text). We are now ready to prove Theorem 4.2.

Proof. We show that under the assumptions made in Theorem 4.2, the Leibniz integral rule as formulated in Theorem A.1 holds. We proceed to show that the smooth ex post utilities $u_i^{\mathrm{SM}(\lambda)}$ are Lipschitz continuous, which essentially ensures all three conditions hold. So, for $i \in \mathcal{I}$, recall the smoothed ex post utility for some $v_i \in \mathcal{V}_i$ and $\lambda > 0$:

$$u_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i})) = \left(v_i - p^{\mathrm{SM}}(b_i, \beta_{\text{-}i}(v_{\text{-}i}))\right) x_i^{\mathrm{SM}(\lambda)}(b_i, \beta_{\text{-}i}(v_{\text{-}i})). \tag{23}$$

As the smooth pricing function p^{SM} is a sum over Lipschitz continuous functions, it is Lipschitz continuous. As $\beta_{\text{-}i}$ and $x_i^{\mathrm{SM}(\lambda)}$ are Lipschitz continuous, so is $x_i^{\mathrm{SM}(\lambda)}(b_i, \beta_{\text{-}i}(\,\cdot\,))$. Finally, as both p^{SM} and $x_i^{\mathrm{SM}(\lambda)}$ are bounded and Lipschitz continuous, their product is Lipschitz continuous as well. Therefore, $u_i^{\mathrm{SM}(\lambda)}$ is Lipschitz continuous in b_i and $v_{\text{-}i}$.

The Lipschitz continuity of u_i^{SM} ensures measurability, as well as integrability over $\mathcal{V}_{\text{-}i}$ for all $b_i \in \mathcal{A}_i$. Hence, Condition 1 holds. As Lipschitz continuity is stronger than absolute continuity, Condition 2 holds as well. Finally, note that due to Lipschitz continuity, there exists an L such that $\left| \frac{\partial}{\partial b_i} u_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i})) \right| \leq L$ for all $b_i \in \mathcal{A}_i$. This bound ensures that Condition 3 holds also.

In the case of conditional priors, consider the function $u_i^{\mathrm{SM}(\lambda)} f_{\text{-}i|\,\cdot\,}$. This function is again Lipschitz continuous as product of bounded Lipschitz continuous functions. Repeating the steps above for this function finishes the proof. □

Remark A.2. The original non-smooth ex post utility function u_i does not satisfy the conditions of Theorem A.1. For example, u_i is not even continuous in b_i, so that the second condition is violated.

B. Proof of Theorem 4.3

Proof. Let us start with the first statement. For the interim utility of bidder i in the original game, we have

$$\overline{u}_i(v_i, b_i, \beta_{\text{-}i}) = \mathbb{E}_{v_{\text{-}i}|v_i}\left[u_i(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))\right] \tag{24}$$

with the ex post utility from Equation 1 in the main text rewritten as

$$u_i(v_i, b_i, b_{\text{-}i}) = (v_i - p_i(b_i, b_{\text{-}i}))x_i(b_i, b_{\text{-}i}). \tag{25}$$

In the smoothed auction, we have

$$\overline{u}_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}) = \mathbb{E}_{v_{\text{-}i}|v_i}\left[u_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))\right] \tag{26}$$

with $u_i^{\mathrm{SM}(\lambda)}$ as defined in Equation 16 in the main text. Note that $u_i^{\mathrm{SM}(\lambda)}$ is integrable as composition of integrable functions. We first have a.e. pointwise convergence of $u_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))$ to $u_i(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))$ as λ approaches zero. That is, for all $v_{\text{-}i}$, except for $b_i = \beta_{\text{-}i}(v_{\text{-}i})$, the smaller λ gets the closer the allocations and the closer the utilities get.

Second, it is easy to see that $|u_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))|$ is bounded via

$$|u_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i}))| \leq |v_i - p^{\mathrm{SM}}(b_i, \beta_{\text{-}i}(v_{\text{-}i}))|$$

and noting that this is a composition of bounded functions. With these two conditions satisfied, we can apply the dominated convergence theorem in its a.e. version (see, e.g., Bogachev (2007), Theorem 2.8.1) on the terms from Equations 24 and 26 which proofs the first statement.

Let us now consider the ex ante utilities. From the interim convergence, we know that the expected interim utility $\overline{u}_i^{\mathrm{SM}(\lambda)}(v_i, b_i, \beta_{\text{-}i})$ converges pointwise to $\overline{u}_i(v_i, b_i, \beta_{\text{-}i})$ for all v_i and b_i. Again applying the dominated convergence theorem ensures equality of the expected utility in the ex ante state of the game. □

Remark B.1. Technically, a tie-breaking rule should be specified for x_i at the nullset $b_i = \beta_{\text{-}i}(v_{\text{-}i})$, but that is exactly the point which we neglect.

C. Proof of Proposition 4.5

The following section provides a linear bound on the error in interim and ex ante utility. For clarity, we restate all major assumptions.

Assumption C.1. Consider a Bayesian auction game G and assume:

1. The action $\mathcal{A}_i$ and valuation spaces $\mathcal{V}_i$ are compact intervals.
2. F is an atomless prior.
3. The bidding and pricing functions are measurable.

Assumption C.2. For all $i \in \mathcal{I}$ assume:

1. β_i is strictly increasing and Lipschitz continuous.
2. β_i^{-1} is Lipschitz continuous.
3. There exists a uniform bound for all marginal conditional prior density functions $f_{i|\cdot}$.
4. p_i is bounded.

Proof. We use the Hölder inequality throughout the proof, which we denote by (H). Whenever we use a specific assumption, we denote it by the corresponding number. We begin with the interim utility error, i.e., we aim to bound the following term for all $i \in \mathcal{I}, v_i \in \mathcal{V}_i$, and $\lambda > 0$:

$$\overline{\varepsilon}_i(v_i, b_i, \beta_{\text{-}i}, \lambda) := \left| \int_{\mathcal{V}_{\text{-}i}} u_i\left(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i})\right) - u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i})\right) \mathrm{d}F_{\text{-}i|i}(v_{\text{-}i}) \right|. \tag{27}$$

So, let $i \in \mathcal{I}$, $v_i \in \mathcal{V}_i$, $b_i \in \mathcal{A}_i$, and $\lambda > 0$ be arbitrary. By splitting up the integral into the individual opponents, we get

$$\overline{\varepsilon}_i(v_i, b_i, \beta_{\text{-}i}, \lambda) = \left| \int_{\mathcal{V}_1} \cdots \int_{\mathcal{V}_{i-1}} \int_{\mathcal{V}_{i+1}} \cdots \int_{\mathcal{V}_n} u_i\left(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i})\right) - u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i})\right) \right.$$

$$\left| \mathrm{d}F_{n|\text{-}n}(v_n) \dots \mathrm{d}F_{i+1|(1,\dots,i)}(v_{i+1}) \mathrm{d}F_{i-1|(1,\dots,i-2,i)}(v_{i-1}) \dots \mathrm{d}F_{1|i}(v_1) \right| \tag{28}$$

$$\overset{\text{(H),(A1.2),(A2.3)}}{\leq} \prod_{j \neq i} \|f_j\|_\infty \int_{\mathcal{V}_{\text{-}i}} \left| u_i\left(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i})\right) - u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, \beta_{\text{-}i}(v_{\text{-}i})\right) \right| \mathrm{d}v_{\text{-}i} =: (*_1). \tag{29}$$

We use $\|f_j\|_\infty$ to denote the uniform bound for any marginal conditional prior density function $f_j|\cdot$ of bidder j. Next, we perform a change of variables (Katz, 1982) through the inverse of the opponents' strategies

$$(*_1) \overset{\text{(A2.1),(A2.2)}}{=} \prod_{j \neq i} \|f_j\|_\infty \int_{\beta_{\text{-}i}(\mathcal{V}_{\text{-}i})} \left| \left(u_i\left(v_i, b_i, b_{\text{-}i}\right) - u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, b_{\text{-}i}\right) \right) \right| \cdot \left| \det \mathrm{D}\beta_{\text{-}i}^{-1}(b_{\text{-}i}) \right| \mathrm{d}b_{\text{-}i} \tag{30}$$

$$\overset{\text{(H),(A1.1)}}{\leq} \prod_{j \neq i} \|f_j\|_\infty \cdot \left\| \left(\beta_j^{-1}\right)' \right\|_\infty \int_{\beta_{\text{-}i}(\mathcal{V}_{\text{-}i})} \left| u_i\left(v_i, b_i, b_{\text{-}i}\right) - u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, b_{\text{-}i}\right) \right| \mathrm{d}b_{\text{-}i}. \tag{31}$$

Note that $\det \mathrm{D}\beta_{\text{-}i}^{-1}(b_{\text{-}i})$ is a diagonal matrix, as β_j only depends on v_j for every $j \in \mathcal{I}$, so that the determinate is given by the product of the individual inverse functions' derivatives. We continue with bounding the remaining integral. For this, define the set

$$A_{b_i} = \left\{ v_{\text{-}i} \in \mathcal{V}_{\text{-}i} \;\middle|\; \max_{\text{-}i} \beta_{\text{-}i}(v_{\text{-}i}) \geq b_i \right\}, \tag{32}$$

which includes all valuations of bidder i's opponents such that the item is *not* allocated to bidder i in the original game. The integral can then be split up in the following way

$$\int_{\beta_{\text{-}i}(\mathcal{V}_{\text{-}i})} \left| u_i\left(v_i, b_i, b_{\text{-}i}\right) - u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, b_{\text{-}i}\right) \right| \mathrm{d}b_{\text{-}i} \tag{33}$$

$$= \int_{\beta_{\text{-}i}\left(\mathcal{V}_{\text{-}i} \setminus A_{b_i}\right)} \left| u_i\left(v_i, b_i, b_{\text{-}i}\right) - u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, b_{\text{-}i}\right) \right| \mathrm{d}b_{\text{-}i} \qquad (=: (\#_1)) \tag{34}$$

$$+ \int_{\beta_{\text{-}i}\left(A_{b_i}\right)} \left| u_i\left(v_i, b_i, b_{\text{-}i}\right) - u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, b_{\text{-}i}\right) \right| \mathrm{d}b_{\text{-}i} \qquad (=: (\#_2)). \tag{35}$$

It remains to bound the integrals $(\#_1)$ and $(\#_2)$. We proceed with $(\#_2)$, i.e., the integral over the set, where the item is not allocated to bidder i. We get

$$(\#_2) = \int_{\beta_{\text{-}i}\left(A_{b_i}\right)} \left| u_i^{\mathrm{SM}(\lambda)}\left(v_i, b_i, b_{\text{-}i}\right) \right| \mathrm{d}b_{\text{-}i} = \int_{\beta_{\text{-}i}\left(A_{b_i}\right)} \left| \left(v_i - p^{\mathrm{SM}}(b_i, b_{\text{-}i})\right) x_i^{\mathrm{SM}(\lambda)}(b_i, b_{\text{-}i}) \right| \mathrm{d}b_{\text{-}i} \tag{36}$$

$$\overset{\text{(H),(A2.4),(A1.1)}}{\leq} \left\| v_i - p^{\mathrm{SM}}(b_i, \cdot) \right\|_{\infty_{|\beta_{\text{-}i}\left(A_{b_i}\right)}} \cdot \int_{\beta_{\text{-}i}\left(A_{b_i}\right)} \left| x_i^{\mathrm{SM}(\lambda)}(b_i, b_{\text{-}i}) \right| \mathrm{d}b_{\text{-}i}. \tag{37}$$

The additional subscript of the supremum norm indicates that the domain is limited to $\beta_{\text{-}i}(A_{b_i})$. This step reduced the problem for $(\#_2)$ to finding a bound for the integral over the soft-allocation function. Note that the softmax function is strictly positive and strictly decreasing in all components of $b_{\text{-}i}$. If $A_{b_i} = \emptyset$, the integral is zero and any positive number is an upper bound. Otherwise, there exists a $j \neq i$ such that $\beta_j(v_j) \geq b_i$ for all $(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_j, \dots v_n) \in A_{b_i}$. Therefore, we can bound the integral by

$$\int_{\beta_{\text{-}i}\left(A_{b_i}\right)} \left| x_i^{\mathrm{SM}(\lambda)}(b_i, b_{\text{-}i}) \right| \mathrm{d}b_{\text{-}i} = \int_{\beta_{\text{-}i}\left(A_{b_i}\right)} \frac{1}{1 + \sum_{j \neq i} \exp\left(\frac{b_j - b_i}{\lambda}\right)} \mathrm{d}b_{\text{-}i} \tag{38}$$

$$\leq \int_{\{b_j \geq b_i\}} \frac{1}{1 + \exp\left(\frac{b_j - b_i}{\lambda}\right)} \mathrm{d}b_j \tag{39}$$

$$\leq \lim_{M \to \infty} \int_{b_j = b_i}^{M} \frac{1}{1 + \exp\left(\frac{b_j - b_i}{\lambda}\right)} \mathrm{d}b_j \tag{40}$$

$$= \lim_{M\to\infty} \left[b_j - \lambda \ln\left(\exp\left(\frac{b_j}{\lambda}\right) + \exp\left(\frac{b_i}{\lambda}\right)\right)\right]_{b_j=b_i}^{M} \tag{41}$$

$$= \lim_{M\to\infty} M - \lambda \ln\left(\exp\left(\frac{M}{\lambda}\right) + \exp\left(\frac{b_i}{\lambda}\right)\right) - b_i + \lambda \ln\left(2\exp\left(\frac{b_i}{\lambda}\right)\right) \tag{42}$$

$$\leq \lim_{M\to\infty} M - \lambda \ln\left(\exp\left(\frac{M}{\lambda}\right)\right) - b_i + \lambda \ln(2) + \lambda \ln\left(\exp\left(\frac{b_i}{\lambda}\right)\right) \tag{43}$$

$$= \lambda \ln(2), \tag{44}$$

which finishes the bound for part $(\#_2)$.

We perform similar steps for the integral $(\#_1)$, that is taken over the opponents' valuations where bidder i gets the item. Using the definition of the smooth pricing function p^{SM} over this set, we get

$$(\#_1) = \int_{\beta_{\text{-}i}\left(\mathcal{V}_{\text{-}i}\setminus A_{b_i}\right)} \left| v_i - p_i(b_i, b_{\text{-}i}) - \left(v_i - p^{\mathrm{SM}}(b_i, b_{\text{-}i})\right) x_i^{\mathrm{SM}(\lambda)}(b_i, b_{\text{-}i}) \right| \mathrm{d}b_{\text{-}i} \tag{45}$$

$$= \int_{\beta_{\text{-}i}\left(\mathcal{V}_{\text{-}i}\setminus A_{b_i}\right)} \left| (v_i - p_i(b_i, b_{\text{-}i})) \left(1 - x_i^{\mathrm{SM}(\lambda)}(b_i, b_{\text{-}i})\right) \right| \mathrm{d}b_{\text{-}i} \tag{46}$$

$$\overset{\text{(H)},\text{(A2.4)}}{\leq} \|v_i - p_i(b_i, \cdot)\|_{\infty_{|\beta_{\text{-}i}\left(\mathcal{V}_{\text{-}i}\setminus A_{b_i}\right)}} \cdot \int_{\beta_{\text{-}i}\left(\mathcal{V}_{\text{-}i}\setminus A_{b_i}\right)} \left| 1 - x_i^{\mathrm{SM}(\lambda)}(b_i, b_{\text{-}i}) \right| \mathrm{d}b_{\text{-}i}. \tag{47}$$

It remains to bound the integral of $h(b_{\text{-}i}) := 1 - x_i^{\mathrm{SM}(\lambda)}(b_i, b_{\text{-}i})$ over the set where bidder i wins the item. Note that h is strictly positive and strictly increasing in all variables $b_{\text{-}i}$. If the set $\mathcal{V}_{\text{-}i} \setminus A_{b_i}$ is empty, we are done. Otherwise, we can bound the integral in the following way:

$$\int_{\beta_{\text{-}i}\left(\mathcal{V}_{\text{-}i}\setminus A_{b_i}\right)} h(b_{\text{-}i})\, \mathrm{d}b_{\text{-}i} = \int_{\beta_{\text{-}i}\left(\mathcal{V}_{\text{-}i}\setminus A_{b_i}\right)} 1 - \frac{1}{1 + \sum_{j\neq i} \exp\left(\frac{b_j - b_i}{\lambda}\right)} \mathrm{d}b_{\text{-}i} \tag{48}$$

$$= \int_{\beta_{\text{-}i}\left(\mathcal{V}_{\text{-}i}\setminus A_{b_i}\right)} \frac{\sum_{j\neq i} \exp\left(\frac{b_j - b_i}{\lambda}\right)}{1 + \sum_{j\neq i} \exp\left(\frac{b_j - b_i}{\lambda}\right)} \mathrm{d}b_{\text{-}i} \tag{49}$$

$$\leq \lim_{M\to-\infty} \int_M^{b_i} \cdots \int_M^{b_i} \frac{\sum_{j\neq i} \exp\left(\frac{b_j - b_i}{\lambda}\right)}{1 + \sum_{j\neq i} \exp\left(\frac{b_j - b_i}{\lambda}\right)} \mathrm{d}b_1 \ldots \mathrm{d}b_{i-1} \mathrm{d}b_{i+1} \ldots \mathrm{d}b_n \tag{50}$$

$$\leq \lim_{M\to-\infty} \int_M^{b_i} \cdots \int_M^{b_i} \sum_{j\neq i} \exp\left(\frac{b_j - b_i}{\lambda}\right) \mathrm{d}b_1 \ldots \mathrm{d}b_{i-1} \mathrm{d}b_{i+1} \ldots \mathrm{d}b_n \tag{51}$$

$$= \lim_{M\to-\infty} \sum_{j\neq i} \int_M^{b_i} \exp\left(\frac{b_j - b_i}{\lambda}\right) \mathrm{d}b_j \tag{52}$$

$$= \lim_{M\to-\infty} \sum_{j\neq i} \lambda \left[\exp\left(\frac{b_j - b_i}{\lambda}\right)\right]_{b_j=M}^{b_i} \tag{53}$$

$$= (n-1)\,\lambda. \tag{54}$$

Combining the derived statements, the interim utility error is bounded by

$$\bar{\varepsilon}_i(v_i, b_i, \beta_{\text{-}i}, \lambda) \;\leq\; K(v_i, b_i, \beta_{\text{-}i}) \cdot \lambda, \tag{55}$$

where

$$K(v_i, b_i, \beta_{\text{-}i}) = \left(\prod_{j\neq i} \left(\|f_j\|_\infty \cdot \left\| (\beta_j^{-1})' \right\|_\infty \right) \right)$$

$$\cdot \left(\ln(2) \left\| v_i - p^{\mathrm{SM}}(b_i, \cdot) \right\|_{\infty_{|\beta_{-i}(A_{b_i})}} + \left\| v_i - p(b_i, \cdot) \right\|_{\infty_{|\beta_{-i}(\mathcal{V}_{-i} \setminus A_{b_i})}} (n-1) \right). \tag{56}$$

Consequently, we can bound the ex ante utility error by performing similar steps as above by

$$\tilde{\varepsilon}_i(\beta_i, \beta_{-i}) := \left| \int_{\mathcal{V}_i} \overline{u}_i(v_i, \beta_i(v_i), \beta_{-i}) - \overline{u}_i^{\mathrm{SM}(\lambda)}(v_i, \beta_i(v_i), \beta_{-i}) \, \mathrm{d}F_i(v_i) \right| \tag{57}$$

$$\leq \lambda \int_{\mathcal{V}_i} K(v_i, \beta_i(v_i), \beta_{-i}) \, \mathrm{d}F_i(v_i) \tag{58}$$

$$\overset{\text{(H),(A2.3),(A2.1),(A2.2)}}{\leq} \tilde{K} \cdot \lambda, \tag{59}$$

where

$$\tilde{K} = \mu(\beta_i(\mathcal{V}_i)) \left(\prod_{j \in \mathcal{I}} \|f_j\|_\infty \cdot \left\| \left(\beta_j^{-1}\right)' \right\|_\infty \right) \left(\ln(2) \left\| g_i^{\mathrm{SM}} \right\|_\infty + (n-1) \|g_i\|_\infty \right), \tag{60}$$

where μ denotes the Borel-measure and for the functions

$$g_i^{\mathrm{SM}}(b_i, \beta_i^{-1}(b_i)) = \left\| \beta_i^{-1}(b_i) - p^{\mathrm{SM}}(b_i, \cdot) \right\|_{\infty_{|\beta_{-i}(A_{b_i})}}, \tag{61}$$

$$g_i(b_i, \beta_i^{-1}(b_i)) = \left\| \beta_i^{-1}(b_i) - p_i(b_i, \cdot) \right\|_{\infty_{|\beta_{-i}(\mathcal{V}_{-i} \setminus A_{b_i})}}. \tag{62}$$

□

Remark C.3. The proof uses the infinity norm of the Hölder inequality ($\|fg\|_1 \leq \|f\|_1 \|g\|_\infty$), so that one needs Condition 2 of Assumption 4.4. One can use the $\|\cdot\|_2$-norm in the inequality ($\|fg\|_1 \leq \|f\|_2 \|g\|_2$), so that one can weaken the assumption to $\left\| \left(\beta_i^{-1}\right)' \right\|_2 < \infty$. However, this leads to a worse convergence rate of the form $\mathcal{O}(\sqrt{\lambda})$.

Remark C.4. The result also holds for a special case of risk-averse bidders. Assume the ex post utility functions additionally depend on a risk parameter $\rho \in (0, 1]$:

$$u_i(v_i, b, \rho) = (v_i \, x_i(b) - p_i(b))^\rho \tag{63}$$

$$u_i^{\mathrm{SM}(\lambda)}(v_i, b, \rho) = \left(\left(v_i - p^{\mathrm{SM}}(b)\right) x_i^{\mathrm{SM}(\lambda)}(b) \right)^\rho. \tag{64}$$

By using Jensen's inequality for the concave mapping $x \mapsto x^\rho$, one can derive a convergence rate of $\mathcal{O}(\lambda^\rho)$ by performing analogous steps as in the proof above.

D. Proof of Theorem 4.6

Let us now prove that an ε-equilibrium in the smoothed game translates to an $\varepsilon + \mathcal{O}(\lambda)$-equilibrium in the original auction.

Proof. We know there exists a constant K such that $|\tilde{u}_i^{\mathrm{SM}(\lambda)}(\beta) - \tilde{u}_i(\beta)| \leq K\lambda$ due to Proposition 4.5. Therefore, the following holds for an ε-BNE in the smoothed game β^* and any strategy β_i:

$$\begin{aligned} \tilde{u}_i(\beta_i, \beta_{-i}^*) - \tilde{u}_i(\beta_i^*, \beta_{-i}^*) &= \tilde{u}_i(\beta_i, \beta_{-i}^*) - \tilde{u}_i^{\mathrm{SM}(\lambda)}(\beta_i, \beta_{-i}^*) + \tilde{u}_i^{\mathrm{SM}(\lambda)}(\beta_i, \beta_{-i}^*) \\ &\quad - \tilde{u}_i^{\mathrm{SM}(\lambda)}(\beta_i^*, \beta_{-i}^*) + \tilde{u}_i^{\mathrm{SM}(\lambda)}(\beta_i^*, \beta_{-i}^*) - \tilde{u}_i(\beta_i^*, \beta_{-i}^*) \qquad (65) \\ &\leq K\lambda + \varepsilon + K\lambda = \varepsilon + 2K\lambda. \qquad (66) \end{aligned}$$

This is equivalent to

$$\tilde{\ell}_i(\beta_i^*, \beta_{-i}^*) \leq \varepsilon + 2K\lambda. \tag{67}$$

□

E. A Special Case: Exact Errors

Consider the single-item FPSB auction with two bidders having uniform priors on the unit interval. Suppose that bidder 2 has a linear strategy $\beta_2(v_2) = s\,v_2$ with $s \in (0, 1]$. Then we can derive an error rate for bidder 1's absolute interim utility difference $\overline{\varepsilon}_\lambda = |\overline{u}_1 - \overline{u}_1^{\text{SM}}|$ and for valuation v_1 and bid $b_1 \leq s$.

For the interim utility of bidder 1 in the original game, we have

$$\overline{u}_1(v_1, b_1) = \int_{v_2=0}^{\frac{b_1}{s}} (v_1 - b_1)\,\mathrm{d}v_2 = \frac{v_1 - b_1}{s}\,b_1. \tag{68}$$

In the smoothed auction, we have

$$\overline{u}_1^{\text{SM}}(v_1, b_1) = \int_{v_2=0}^{1} (v_1 - \max\{b_1, s\,v_2\}) x_1^{\text{SM}}(b_1, s\,v_2)\,\mathrm{d}v_2. \tag{69}$$

When splitting the domain of the integral at $b_1 = s\,v_2$, the first integral evaluates to

$$-\frac{b_1 - v_1}{s}\left(\lambda \ln\left(\mathrm{e}^{\frac{b}{\lambda}} + 1\right) - \lambda \ln\left(2\mathrm{e}^{\frac{b_1}{\lambda}}\right) + b_1\right)$$

and the second to

$$\begin{aligned}
&\frac{1}{2s}\Big[2s\lambda v_2 \ln\left(\mathrm{e}^{\frac{sv_2}{\lambda} - \frac{b_1}{\lambda}} + 1\right) + 2\lambda^2 \operatorname{Li}_2\left(-\mathrm{e}^{\frac{sv_2}{\lambda} - \frac{b_1}{\lambda}}\right) - 2v_1\lambda \ln\left(\mathrm{e}^{\frac{sv_2}{\lambda}} + \mathrm{e}^{\frac{b_1}{\lambda}}\right) - s^2v_2^2 + 2v_1 s v_2\Big]_{v_2=\frac{b_1}{s}}^{1} \\
&= \frac{1}{s}\Big(\lambda(s \ln(\mathrm{e}^{s/\lambda - b_1/\lambda} + 1) - b_1 \ln(2)) + \lambda^2(\operatorname{Li}_2(-\mathrm{e}^{(s-b_1)/\lambda}) + \frac{1}{12}\pi^2) \\
&+ v_1\lambda(\ln(2) + b_1/\lambda - \ln(\mathrm{e}^{s/\lambda} + \mathrm{e}^{b_1/\lambda})) + \frac{1}{2}(b_1^2 - s^2) + v_1(s - b_1)\Big).
\end{aligned}$$

Combining these results, we arrive at an exact interim error of

$$\begin{aligned}
\overline{\varepsilon}_\lambda(v_1, b_1) = \frac{\lambda}{s}\Bigg(&- s \ln\left(\mathrm{e}^{\frac{s-b_1}{\lambda}} + 1\right) - \lambda\left(\operatorname{Li}_2\left(-\mathrm{e}^{\frac{s-b_1}{\lambda}}\right) + \frac{1}{12}\pi^2\right) \\
&+ v_1 \ln(\mathrm{e}^{s/\lambda} + \mathrm{e}^{b_1/\lambda}) + \frac{1}{\lambda}\left(\frac{s^2 - b_1^2}{2} - v_1 s\right) \\
&+ (b_1 - v_1)\left(\ln(\mathrm{e}^{-b_1/\lambda} + 1)\right)\Bigg).
\end{aligned} \tag{70}$$

Here, Li_2 is the dilogarithm. This result shows vastly different convergence rates across the valuation and action space. Let us assume both bidders are playing according to their BNE strategy, $\beta_i(v_i) = 0.5v_i$. Now, for the extreme case of $v_1 = 1$, $\overline{\varepsilon}_\lambda$ tends towards a linear function. At the other end of the spectrum, for $v_1 = 0.5$, $\overline{\varepsilon}_\lambda$ tends towards being constantly zero. In summary, the higher the valuation is the slower the convergence rate, with a linear rate in the worst case. Figure 2 in the main paper shows the convergence rates for smaller temperatures. We depict a selection of three valuations.

Again assuming that both bidders are playing according to their BNE strategy, the ex ante error

$$\tilde{\varepsilon}_\lambda(\beta_1) \;:=\; \int_{v_1=0}^{1} \overline{\varepsilon}_\lambda(v_1, \beta_1(v_1))\,\mathrm{d}v_1 \tag{71}$$

can be approximated by taking the sample mean of Equation 70. The expression goes towards zero quickly for ever smaller temperatures and is depicted in Figure 2 in the main paper.

F. Impact of Batch Size

We have run experiments for different batch sizes. The performance increase for ever larger batch sizes diminishes and optimal results are reached for temperature values just below 0.01 as can be seen in Table 2.

Batch size	$\min L_2$	λ
$2^{10} = 1{,}024$	0.0177	0.0239
$2^{14} = 16{,}384$	0.0044	0.0119
$2^{18} = 262{,}144$	0.0044	0.0089
$2^{22} = 4{,}194{,}304$	0.0042	0.0089

Table 2. Results for a selection of different batch sizes in the FPSB base setting. We report λ for which the L_2 loss is minimized. 2^{18} is the default value used during training and 2^{22} the maximal value possible on the GPU used.

G. Reproducibility and Hyperparameters

We have implemented all auctions and algorithms in the PyTorch framework. The code is available at GitHub via github.com/heidekrueger/bnelearn.

G.1. Learning

We use common hyperparameters across all settings except where noted otherwise. The feed-forward neural networks are fully connected with two hidden layers of ten nodes each with SeLU activations, as well as ReLU activations applied to the output layer. We model all bidders by a shared policy because the auctions considered are symmetric. Hence, learning is stabilized but limited to finding symmetric BNE. Furthermore, we perform supervised pretraining of 50 iterations towards truthful strategies to prevent degenerate initializations. All experiments are run on a single Nvidia GeForce 2080Ti GPU with 11 GB of memory and a batch size of 2^{18} for learning. Each experiment was repeated five times with 2,000 iterations. Furthermore, the following algorithm specific settings were used:

- For NPGA, we choose a population size of 64 and a variance of 1 for the normal distribution from which we draw population samples in parameter space. The variance is then scaled by the model size as is done in (Bichler et al., 2021).
- In the case of REINFORCE, the output dimension is increased by a factor of two because for each bid, a normal distribution (with its two parameters) is learned instead.
- For the smoothed game, we choose a temperature of 0.01.

G.2. Evaluating

A batch size of 2^{22} was used for the calculation of the L_2 loss. The choice of batch sizes was mainly driven by maxing out the GPU memory. Learning requires more memory than evaluating L_2, so the latter was possible to conduct with larger batch sizes. For the utility loss ε, we decreased the number of prior samples from the player currently under evaluation to $n_{\text{own-batch}} = 2^{10}$. For each of these valuations, his or her best response — with possible actions from an equidistant grid of size $n_{\text{gird}} = 2^{10}$ — is approximated over a sample of $n_{\text{opponent-batch}} = 2^{20}$ opponent valuations. A higher batch size for the opponents is necessary for reaching the required precision in estimating the utilities. In total, the calculation of ε requires $n_{\text{own-batch}} \cdot n_{\text{gird}} \cdot n_{\text{opponent-batch}} = 2^{40} > 1$ trillion game evaluations.