Mathematics of Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Mode Connectivity in Auction Design

Christoph Hertrich, Yixin Tao, László A. Végh

informs
https://pubsonline.informs.org/journal/moor

MATHEMATICS OF OPERATIONS RESEARCH
Vol. 51, No. 1, February 2026, pp. 456–475
ISSN 0364-765X (print), ISSN 1526-5471 (online)

Mode Connectivity in Auction Design

Christoph Hertrich,[a,*] **Yixin Tao,**[b] **László A. Végh**[c]

[a] Applied Discrete Mathematics Lab, Department Liberal Arts & Social Sciences, University of Technology Nuremberg, 90443 Nuremberg, Germany; [b] Key Laboratory of Interdisciplinary Research of Computation and Economics, Shanghai University of Finance and Economics, Shanghai 200433, China; [c] Hertz Chair in the Transdisciplinary Research Area Modelling, University of Bonn, 53113 Bonn, Germany

*Corresponding author

Contact: christoph.hertrich@utn.de, https://orcid.org/0000-0001-5646-8567 (CH); taoyixin@mail.shufe.edu.cn, https://orcid.org/0000-0002-0573-1369 (YT); lvegh@uni-bonn.de, https://orcid.org/0000-0003-1152-200X (LAV)

Received: May 16, 2024
Revised: December 18, 2024
Accepted: January 25, 2025
Published Online in Articles in Advance: February 26, 2025

MSC2020 Subject Classifications: 91B26, 68T07

https://doi.org/10.1287/moor.2024.0525

Abstract. Optimal auction design is a fundamental problem in algorithmic game theory. This problem is notoriously difficult already in very simple settings. Recent work in differentiable economics showed that neural networks can efficiently learn known optimal auction mechanisms and discover interesting new ones. In an attempt to theoretically justify their empirical success, we focus on one of the first such networks, *RochetNet*, and a generalized version for *affine maximizer auctions*. We prove that they satisfy *mode connectivity*; that is, locally optimal solutions are connected by a simple, piecewise linear path such that every solution on the path is almost as good as one of the two local optima. Mode connectivity has been recently investigated as an intriguing empirical and theoretically justifiable property of neural networks used for prediction problems. Our results give the first such analysis in the context of differentiable economics, where neural networks are used directly for solving nonconvex optimization problems.

Funding: All three authors gratefully acknowledge support by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program [Grant ScaleOpt–757481 (all three authors); Grant ForEFront–615640 (C. Hertrich)]. Y. Tao is partially supported by the National Key R&D Program of China [Grant 2023YFA1009500].

Keywords: differentiable economics • mechanism design • neural network theory • mode connectivity • RochetNet

1. Introduction

Auction design is a core problem in mechanism design, with immense applications in electronic commerce (such as sponsored search auctions) as well as in the public sector (such as spectrum auctions). In a revenue-maximizing auction, the auctioneer needs to design a mechanism to allocate resources to buyers, and set prices in order to maximize the expected revenue. The buyers' preferences are private, and they may behave strategically by misreporting them. For this reason, it is often desirable to devise *dominant strategy incentive compatible (DSIC)* and *individually rational (IR)* mechanisms. By definition, in a DSIC mechanism, it is a dominant strategy for the buyers to report the true valuations; in an IR mechanism, each participating truthful buyer receives a nonnegative payoff.

We focus on DSIC and IR mechanisms that maximize the expected revenue, assuming that the buyers' preferences are drawn from a distribution known to the auctioneer. A classical result of Myerson [32] provides the optimal mechanism for the case of a single item and arbitrary number of buyers. Finding the optimal mechanisms for more general settings is a tantalizingly difficult problem. We refer the reader to the surveys by Manelli and Vincent [31], Rochet and Stole [35], and Daskalakis [13] for partial results and references. In particular, no analytic solution is known even for two items and two buyers. Selling multiple items to a single buyer is computationally intractable (Daskalakis et al. [15]). Already for two items and a single buyer, the description of the optimal mechanism may be uncountable (Daskalakis et al. [14]). Recent work gives a number of important partial characterizations, for example, Daskalakis et al. [16] and Giannakopoulos and Koutsoupias [24], as well as results for weaker notions of Bayesian incentive compatibility, for example, Bhalgat et al. [4], Cai et al. [5], Cai et al. [6], and Cai et al. [7].

Conitzer and Sandholm [9] and Conitzer and Sandholm [10] propose the approach of *automated mechanism design* to use optimization and computational methods to obtain (near) optimal mechanisms for specific problems; see also Sandholm and Likhodedov [36]. An active recent area of research uses machine learning tools. In particular, Dütting et al. [21] design and train neural networks to automatically find optimal auctions. They study two network architectures, and show that several theoretically optimal mechanisms can be recovered using this approach, as well as that interesting new mechanisms can be obtained. The first network they study is *RochetNet*. This is a simple two-layer neural network applicable to the single buyer case, leveraging Rochet's [34] characterization of the optimal mechanism. The second network, *RegretNet*, does not require such a characterization and is applicable for multiple buyers; however, it only provides approximate incentive compatibility.

Dütting et al. [21] coined the term "*differentiable economics*" for this approach, and there has been significant further work in this direction. These include designing auctions for budget constrained buyers (Feng et al. [22]), multifacility location (Golowich et al. [26]), balancing fairness and revenue objectives (Kuo et al. [30]), incorporating nonlinear utility functions and other networks trained from interaction data (Shen et al. [37]),[1] designing revenue-maximizing auctions with differentiable matchings (Curry et al. [12]), contextual auction design (Duan et al. [19]), designing taxation policies (Zheng et al. [41]), and more.

The purpose of this work is to supply theoretical evidence behind the success of neural networks in differentiable economics. The revenue is a highly nonconvex function of the parameters in the neural network. Curiously, gradient approaches seem to recover globally optimal auctions despite this nonconvexity. Similar phenomena have been studied more generally in the context of deep networks, and theoretical explanations have been proposed, in particular, overparametrization (Allen-Zhu et al. [1], Du et al. [18]).

1.1. Mode Connectivity

Recent work has focused on this striking property of the landscape of loss functions of deep neural networks: local optimal solutions (modes) found by gradient approaches are connected by simple paths in the parameter space. We now give an informal definition of *mode connectivity* for general functions.

Definition 1 (ε-Mode-Connected (Informal)). Consider the optimization problem $\min_{x\in K} f(x)$ for $f : K \to \mathbb{R}$. We say that $x, y \in K$ are *ε-mode-connected*, if there is a continuous curve $P \subseteq K$ connecting x and y such that $f(z) \le \max\{f(x), f(y)\} + \varepsilon$ for any $z \in P$.

Mode connectivity of loss functions in deep learning was identified by Garipov et al. [23] and by Draxler et al. [17]. We particularly refer to Garipov et al. [23, figure 1] for an intuitive visualization of the phenomenon. Mode connectivity can help to explain why stochastic gradient descent (SGD), the dominant optimization heuristic used in deep learning, often results in near-zero training loss, regardless of the random starting point, whereas the straight-line interpolation between optima can exhibit high loss.

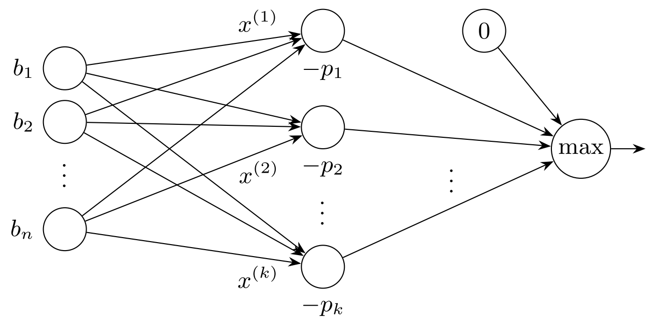
To some extent, mode connectivity prevents a poor local minimal valley region on the function value, from which the SGD method cannot escape easily. However, we would like to emphasize that mode connectivity does not provide a formal proof, but only a useful intuition of the success of SGD.

Kuditipudi et al. [29] gave strong theoretical arguments for mode connectivity. They introduce the notion of *ε-dropout stability*: solutions to a neural network such that in each layer, one can remove at least half the neurons and rescale the remaining units such that the loss function increases by at most ε. Solutions that are ε-dropout stable are then shown to be ε-mode-connected. Moreover, they show that *noise stability* (see, e.g., Arora et al. [2]) implies dropout stability, and hence, mode connectivity. Nguyen [33] showed mode connectivity when there is a hidden layer larger than the training data set. Shevchenko and Mondelli [38] show that stochastic gradient descent solutions to sufficiently overparametrized neural networks are dropout stable even if we only keep a small, randomly sampled set of neurons from each layer.

1.2. Our Contributions

1.2.1. RochetNet. In this paper, we first establish mode connectivity properties of *RochetNet*, the architecture in Dütting et al. [21] for multiple items and a single buyer. These networks have a single hidden layer, corresponding to the menu. Each neuron on the hidden layer represents an *option* in the menu, in the form $(x, p) \in \mathbb{R}^n \times \mathbb{R}$, where x is an allocation of the items and p is a price offered to the buyer (see Figure 1). Given a valuation vector $v \in \mathbb{R}^n$, the buyer has a quasilinear utility $v^\top x - p$. The buyer is assigned a single option (x, p) that maximizes their utility, using some tie-breaking rule if needed. The chosen option is called *active* for the buyer. The revenue of the auctioneer is the expected price paid by the buyer. The loss function in *RochetNet* is the negative revenue. Despite its simplicity, the experiments on *RochetNet* by Dütting et al. [21] gave impressive empirical results in different scenarios. For example, in the experiments with up to six items and uniform value distributions, *RochetNet* achieves almost the same revenue (99.9%) as the Straight-Jacket Auctions (Giannakopoulos and Koutsoupias [25]), which are known to be optimal in this case. This success is not limited to a single example, as *RochetNet* also consistently performs well in other scenarios, including when infinite menu size is necessary. Furthermore,

Figure 1. RochetNet: This architecture maps the bid b to the utility of the buyer.



Dütting et al. [21] demonstrated the usefulness of *RochetNet* in discovering optimal auctions in situations that were previously unexplored from a theoretical perspective.

First, in Theorem 1, we show that for linear utilities, ε-mode connectivity holds between two solutions that are *ε-reducible* in the following sense: out of the $K+1$ menu options (neurons), there exists a subset of at most $\sqrt{K+1}$ containing an active option for the buyer with probability at least $1-\varepsilon$. Assuming that the valuations are normalized such that the maximum valuation of any buyer is at most one, it follows that if we remove all other options from the menu, at most ε of the expected revenue is lost. The assumption of being ε-reducible is stronger than ε-dropout stability that only drops a constant fraction of the neurons. At the same time, experimental results by Dütting et al. [21] show evidence of this property being satisfied in practice. They experimented with different-sized neural networks in a setting when the optimal auction requires infinite menu size. Even with 10,000 neurons available, only 59 options were active, that is, used at least once when tested over a large sample size. We note that this property also highlights an advantage of *RochetNet* over *RegretNet* and other similar architectures: instead of a black-box neural network, it returns a compact, easy-to-understand representation of a mechanism.

Our second main result (Theorem 2) shows that for n items and linear utilities, if the number of menu options K is sufficiently large, namely, $(2/\varepsilon)^{4n}$, then ε-mode connectivity holds between *any* two solutions for *any* underlying distribution. The connectivity property holds pointwise: for any particular valuation profile, the revenue may decrease by at most ε along the path. A key tool in this ε-mode connectivity result is a discretization technique from Dughmi et al. [20]. We note that such a mode connectivity result can be expected to need a large menu size. In Appendix A, we present an example with two disconnected local maxima for $K=1$.

1.2.2. Affine Maximizer Auctions (AMAs). We also extend our results and techniques to neural networks for AMAs studied in Curry et al. [11]. This is a generalization of *RochetNet* for multibuyer scenarios. It can also be seen as a weighted variant of the Vickrey-Clarke-Groves (VCG) mechanism (Clarke [8], Groves [27], Vickrey [39]). AMA offers various allocation options. For a given valuation profile, the auctioneer chooses the allocation with the highest weighted sum of valuations, and computes individual prices for the buyers; the details are described in Section 2.2. AMA is DSIC and IR, but is not rich enough to always represent the optimal auction.

For AMA networks, we show similar results (Theorems 3 and 4) as for *RochetNet*. We first prove ε-mode connectivity holds between two solutions that are *ε-reducible* (see Definition 6). Curry et al. [11, p. 8] provide evidence of this property being satisfied in practice, observing (in section 7.1), "Moreover, we found that starting out with a large number of parameters improves performance, even though by the end of training only a tiny number of these parameters were actually used." Secondly, we also show that if the number of menu options K is sufficiently large, namely, $(16m^3/\varepsilon^2)^{2nm}$, then ε-mode connectivity holds pointwise between *any* two solutions. That is, it is valid for any underlying distribution of valuations, possibly correlated between different buyers.

1.2.3. Relation to Other Network Architectures. The focus of our results is to handle the challenges induced in the specific network architectures of RochetNet and AMA. As such, they are of limited applicability to other network architectures in deep learning, like feedforward neural networks and transformers. On the other hand, our results do not seem to be deducible from previous mode connectivity results for feedforward networks, as we outline as follows. Previous literature on mode connectivity investigated neural networks used for prediction. The results in Kuditipudi et al. [29], Nguyen [33], and Shevchenko and Mondelli [38] and other papers crucially rely on the properties that the networks minimize a convex loss function between the predicted and actual values, and require linear transformations in the final layer. *RochetNet* and AMA networks are fundamentally different. The training data do not come as labeled pairs, and these network architectures are built directly for solving an optimization problem. For an input valuation profile, the loss function is the negative revenue of the auctioneer. In *RochetNet*, this is obtained as the negative price of the utility-maximizing bundle; for AMA it requires an even more intricate calculation. The objective is to find parameters of the neural network such that the expected revenue is as large as possible. The menu options define a piecewise linear surface of utilities, and the revenue in *RochetNet* can be interpreted as the expected bias of the piece corresponding to a randomly chosen input.

Hence, the landscape of the loss function is fundamentally different from those analyzed in the abovementioned works on prediction networks. The weight interpolation argument that shows mode connectivity from dropout stability is not applicable in this context. The main reason is that the loss function is not a simple function of the output of the network, but is defined by choosing the price corresponding to the utility-maximizing option. We thus need a more careful understanding of the piecewise linear surfaces corresponding to the menus.

1.2.4. Significance for Practitioners. We see the main contribution of our paper in *explaining* the empirical success and providing theoretical foundations for already existent practical methods, and not in inventing new methods. Nevertheless, two insights a practitioner could use are as follows: (i) It is worth understanding the structure of the auction in question. If one can, for example, understand whether ε-reducibility holds for a particular auction, this might indicate whether RochetNet or AMA is a good method to apply to this particular case. (ii) Size helps: If one encounters bad local optima, increasing the menu size and rerunning RochetNet or AMA might be a potential fix and will eventually lead to a network satisfying mode connectivity.

2. Auction Settings

We consider the case with m buyers and one seller with n divisible items each in unit supply. Each buyer has an additive valuation function $v_i(S) := \sum_{j\in S} v_{ij}$, where $v_{ij} \in V$ represents the valuation of the buyer i on item j and $V \subseteq \mathbb{R}$ is the set of possible valuations. Throughout the paper, we normalize the range to the unit simplex: we assume $V = [0,1]$, and $\|v_i\|_1 = \sum_j v_{ij} \le 1$ for every buyer i.[2] With slight abuse of notation, we let $v = (v_{11}, v_{12}, \dots, v_{ij}, \dots, v_{mn})^\top$ and $v_i = (v_{i1}, v_{i2}, \dots, v_{in})^\top$. The buyers' valuation profile v is drawn from a distribution $F \in \mathcal{P}(V^{m\times n})$. The seller has access to samples from this distribution. Throughout, we assume that the buyers have *quasilinear utilities*: If a buyer with valuation v_i receives an allocation $x \in [0,1]^n$ at price p, their utility is $v_i^\top x - p$.

In an *auction mechanism*, each bidder i reports a *bid* $b_i \in [0,1]^n$. The entire bid vector $b \in [0,1]^{m\times n}$ will be denoted as $b = (b_1, \dots, b_m) = (b_i, b_{-i})$, where b_{-i} represents all the bids other than buyer i. The mechanism outputs a set of allocations $x(b) = (x_1(b), \dots, x_m(b))$, $x_i(b) \in [0,1]^n$, along with prices $p(b) = (p_1(b), \dots, p_m(b)) \in \mathbb{R}^m$. Because there is unit supply of each item, we require $\sum_i x_{ij}(b) \le 1$, where $x_{ij}(b)$ is the allocation of buyer i of item j.

Definition 2 (DSIC and IR Auction).

(i) An auction is *DSIC* if $v_i^\top x_i(v_i, b_{-i}) - p_i(v_i, b_{-i}) \ge v_i^\top x_i(b_i, b_{-i}) - p_i(b_i, b_{-i})$ for any buyer i and any bid $b = (b_i, b_{-i})$.

(ii) An auction is *IR* if $v_i^\top x_i(v_i, b_{-i}) - p_i(v_i, b_{-i}) \ge 0$ for any buyer i and any bid $b = (b_i, b_{-i})$.

Thus, in a DSIC mechanism, it is a dominant strategy for the agents to report $b_i = v_i$, that is, reveal their true preferences. Therefore, the revenue of a DSIC and IR auction is defined as

$$\texttt{Rev} := \mathbb{E}_{v\sim F}\left[\sum_i p_i(v)\right].$$

2.1. Single Buyer Auctions: RochetNet

Dütting et al. [21] proposed RochetNet as a DSIC and IR auction for the case of a single buyer. We omit the subscript i for buyers in this case. A (possibly infinite-sized) *menu* M comprises a set of *options* offered to the buyer: $M = \{(x^{(k)}, p^{(k)})\}_{k\in\mathcal{K}}$. In each option $(x^{(k)}, p^{(k)})$, $x^{(k)} \in [0,1]^n$ represents the amount of items, and $p^{(k)} \in \mathbb{R}_+$ represents the price. We assume that $0 \in \mathcal{K}$, and $(x^{(0)}, p^{(0)}) = (\mathbf{0}, 0)$ to guarantee IR. We call this the *default option*, whereas all other options are called *regular options*. We will use K to denote the number of regular options; thus, $|\mathcal{K}| = K + 1$.

A buyer submits a bid $b \in [0,1]^n$ representing their valuation, and is assigned to option $k(b) \in \mathcal{K}$ that maximizes the utility

$$k(b) \in \arg\max_{k\in\mathcal{K}} b^\top x^{(k)} - p^{(k)}.$$

This is called the *active option* for the buyer. We break ties in favor of higher prices.[3] Note that option 0 guarantees that the utility is nonnegative, implying the IR property. It is also easy to see that such an auction is DSIC. Therefore, one can assume that $b = v$, that is, the buyer submits their true valuation, or equivalently, the buyer is allowed to directly choose among the menu options one that maximizes their utility. Moreover, it follows from Rochet [34] that every DSIC and IR auction for a single buyer can be implemented with a (possibly infinite-size) menu using an appropriate tie-breaking rule.

Given a menu M, the revenue is defined as

$$\texttt{Rev}(M) := \mathbb{E}_{v\sim F}[p^{(k(v))}].$$

2.1.1. RochetNet. RochetNet (see Figure 1) is a neural network with three layers: an input layer (n neurons), a middle layer (K neurons), and an output layer (one neuron):

1. The input layer takes an n-dimensional bid $b \in V^n$, and sends this information to the middle layer;
2. The middle layer has K neurons. Each neuron represents a regular option in the menu M, which has parameters $x^{(k)} \in [0,1]^n$ and $p^{(k)} \in \mathbb{R}_+$, where $x^{(k)} \in [0,1]^n$ represents the allocation of option k and $p^{(k)}$ represents the price of option k. Neuron k maps from $b \in V^n$ to $b^\top x^{(k)} - p^{(k)}$, that is, the utility of the buyer when choosing option k;
3. The output layer receives all utilities from different options and maximizes over these options and 0: $\max\{\max_k\{(x^{(k)})^\top b - p^{(k)}\}, 0\}$.

We will use $\texttt{Rev}(M)$ to denote the revenue of the auction with menu options $\mathcal{K} = \{0,1,2,\dots,K\}$, where 0 represents the default option $(\mathbf{0},0)$.

The training objective for the RochetNet is to maximize the revenue $\texttt{Rev}(M)$, which is done by stochastic gradient ascent. Note, however, that the revenue is the price of an *argmax* option, which makes it a noncontinuous function of the valuations. For this reason, Dütting et al. [21] use a *softmax* approximation of the *argmax* as their loss function instead; see (16). However, *argmax* is used for testing. In Appendix B, we bound the difference between the revenues computed with these two different activation functions, assuming that the probability density function of the distribution F admits a finite upper bound. Lemma B.1 shows that the difference between the revenues for *softmax* and *argmax* is roughly inverse proportional to the parameter Y of the *softmax* function. This allows the practitioner to interpolate between smoothness of the loss function and provable quality of the softmax approximation by tuning the parameter Y.

2.2. Affine Maximizer Auctions

AMAs also provide a menu M with a set of options $\mathcal{K}$. Each option is of the form $(x^{(k)}, \beta^{(k)}) \in [0,1]^{m\times n} \times \mathbb{R}$, where $x_{ij}^{(k)} \in [0,1]$ represents the allocation of item j to buyer i, with the restriction that $\sum_i x_{ij}^{(k)} \le 1$ for each item j, and $\beta^{(k)}$ represents a "*boost*." We again assume $0 \in \mathcal{K}$, and $(x^{(0)}, \beta^{(0)}) = (\mathbf{0}, 0)$, and call this the *default* option; all other options are called the *regular options*.

Given the bids $b_i \in [0,1]^n$ of the agents, the auctioneer computes a weighted welfare, using weights $w_i \in \mathbb{R}_+$ for the valuations of each agent, and adds the boost $\beta^{(k)}$. Then, the allocation maximizing the weighted boosted welfare is chosen, that is, the option with

$$k(b) \in \arg\max_{k\in\mathcal{K}} \sum_i w_i b_i^\top x_i^{(k)} + \beta^{(k)}.$$

This will also be referred to as the *active option*. The prices collected from the buyers are computed according to the VCG scheme. Namely,

$$p_i(b) := \frac{1}{w_i}\left(\sum_{\ell\neq i} w_\ell b_\ell^\top x_\ell^{(k(b_{-i}))} + \beta^{(k(b_{-i}))}\right) - \frac{1}{w_i}\left(\sum_{\ell\neq i} w_\ell b_\ell^\top x_\ell^{(k(b))} + \beta^{(k(b))}\right). \tag{1}$$

Here, $k(b_{-i})$ represents the option maximizing the weighted boosted welfare when buyer i is omitted, that is, $k(b_{-i}) \in \arg\max_{k\in\mathcal{K}} \sum_{\ell\neq i} w_\ell b_\ell^\top x_\ell^{(k)} + \beta^{(k)}$. It is known that AMA is DSIC and IR [11]. Hence, we can assume that the submitted bids b_i represent the true valuations v_i. We also assume the ties are broken in favor of maximizing the total payment. In the case of unit weights, this is equivalent to choosing the smallest $\beta^{(k)}$ values; see (3) in Section 4. Given the menu M, the revenue of the AMA is

$$\texttt{Rev}(M) := \mathbb{E}_{v\sim F}\left[\sum_i p_i(v)\right].$$

In this paper, we focus on the case when $w_i = 1$ for all buyers. This is also used in the experiments by Curry et al. [11]. For this case, AMA can be implemented by a three-layer neural network similar to *RochetNet*, with $m \times n$ input neurons. For the more general case when the weights w_i can also be adjusted, one can include an additional layer that combines the buyers' allocations.

Note that for a single buyer and $w_1 = 1$, AMA corresponds to *RochetNet*, with price $p^{(k)} = -\beta^{(k)}$ for each menu option. Indeed, in the formula defining the price $p_i(b)$, the first term is 0, as well as the sum in the second term.

As for *RochetNet*, stochastic gradient descent is used with the *softmax* approximation of the revenue $\texttt{Rev}(M)$, in order to avoid the discontinuities introduced by the *argmax*. We bound the difference in the revenue in Appendix C, concluding that it decreases with large-parameter Y as in the RochetNet case.

2.3. Mode Connectivity

One can view the revenue as a function of the menus, that is, the parameters in the mechanism: (i) in *RochetNet*, $\{(x^{(k)}, p^{(k)})\}_{k\in\mathcal{K}}$; (ii) in AMA, $\{(x^{(k)}, \beta^{(k)})\}_{k\in\mathcal{K}}$. We use $\mathcal{M}$ to denote the set of all possible menus.

Definition 3 (Mode Connectivity). Two menus $M_1, M_2 \in \mathcal{M}$ are ε-mode-connected if there is a continuous curve $\pi : [0,1] \to \mathcal{M}$ such that (i) $\pi(0) = M_1$; (ii) $\pi(1) = M_2$; and (iii) for any $t \in [0,1]$, $\texttt{Rev}(\pi(t)) \geq \min\{\texttt{Rev}(M_1), \texttt{Rev}(M_2)\} - \varepsilon$.

3. Mode Connectivity for the RochetNet

In this section we present and prove our main results for the RochetNet. The following definition plays an analogous role to ε-dropout stability in Kuditipudi et al. [29].

Definition 4. A menu M with $|\mathcal{K}| = K+1$ options is called *ε-reducible* if there is a subset $\mathcal{K}' \subseteq \mathcal{K}$ with $0 \in \mathcal{K}'$, $|\mathcal{K}'| \leq \sqrt{K+1}$ such that, with probability at least $1-\varepsilon$ over the distribution of the valuation of the buyer, the active option assigned to the buyer is contained in $\mathcal{K}'$.

As noted in the introduction, such a property can be observed in the experimental results in Dütting et al. [21]. The motivation behind this definition is that if a menu satisfies this property, then all but $\sqrt{K+1}$ options are more or less redundant. In fact, if a menu is ε-reducible, then dropping all but the at-most $\sqrt{K+1}$ many options in $\mathcal{K}'$ results in a menu M' with $\texttt{Rev}(M') \geq \texttt{Rev}(M) - \varepsilon$ because the price of any selected option is bounded by $\|v\|_1 \leq 1$. Our first main theorem is as follows:

Theorem 1. *If two menus M_1 and M_2 for the RochetNet are ε-reducible, then they are ε-mode-connected. Moreover, the curve transforming M_1 into M_2 is piecewise linear with only five pieces.*

The role of the square root in the definition of ε-reducibility becomes apparent in the proof of Theorem 1: intuitively, we can interpolate between two ε-reducible menus M_1 and M_2 by looking at each pair of a nonredundant option in M_1 and a nonredundant option in M_2. By Definition 4, we have $K+1$ such pairs and each of them defines an option in every intermediate menu between M_1 and M_2. The same argument would hold for a stricter definition of ε-reducibility where the number $\sqrt{K+1}$ is replaced with something smaller. For weaker definitions, however, the intermediate menus would need to have more than $K+1$ many options, meaning that we do not obtain mode connectivity with our arguments in those cases.

3.1. Interpolating Between 0-Reducible Menus

As a first step we show Theorem 1 for $\varepsilon = 0$. This will be used both to prove the general case of Theorem 1 as well as to show our second main result, namely, that two large menus are always ε-mode-connected (Theorem 2).

Proposition 1. *If two menus M_1 and M_2 for the RochetNet are 0-reducible, then they are 0-mode-connected. Moreover, the curve transforming M_1 into M_2 is piecewise linear with only three pieces.*

To prove Proposition 1, we introduce two intermediate menus $\widehat{M}_1$ and $\widehat{M}_2$, and show that every menu in the piecewise linear interpolation from M_1 via $\widehat{M}_1$ and $\widehat{M}_2$ to M_2 yields a revenue of at least $\min\{\texttt{Rev}(M_1), \texttt{Rev}(M_2)\}$. Using that menu M_1 has only $\sqrt{K+1}$ nonredundant options, menu $\widehat{M}_1$ will be defined by repeating each of the $\sqrt{K+1}$ options $\sqrt{K+1}$ times. Menu $\widehat{M}_2$ will be derived from M_2 similarly. A technical lemma makes sure that this copying can be done in such a way that each pair of a nonredundant option of M_1 and a nonredundant option of M_2 occurs exactly for one index in $\widehat{M}_1$ and $\widehat{M}_2$.

To make this more formal, we first assume without loss of generality that $K+1$ is a square, such that $\sqrt{K+1}$ is an integer. It is straightforward to verify that the theorem is true for nonsquares $K+1$, too. Suppose the options in M_1 and M_2 are indexed with $k \in \mathcal{K} = \{0, 1, \ldots, K\}$. Because M_1 is 0-reducible, there is a subset $\mathcal{K}_1 \subseteq \mathcal{K}$ with

$0 \in \mathcal{K}_1$, $|\mathcal{K}_1| = \sqrt{K+1}$ such that an option with index in $\mathcal{K}_1$ is selected with probability 1 over the distribution of the possible valuations. Similarly, such a set $\mathcal{K}_2$ exists for M_2. To define the curve that provides mode connectivity, we need the following technical lemma.

Lemma 1. *There exists a bijection $\varphi : \mathcal{K} \to \mathcal{K}_1 \times \mathcal{K}_2$ such that for all $k \in \mathcal{K}_1$ we have that $\varphi(k) \in \{k\} \times \mathcal{K}_2$, and for all $k \in \mathcal{K}_2$ we have that $\varphi(k) \in \mathcal{K}_1 \times \{k\}$.*

Proof. We prove the claim by providing an explicit construction for φ in two different cases.

First, suppose that there is some $k^* \in \mathcal{K}_1 \cap \mathcal{K}_2$. In this case, start by setting $\varphi(k) := (k, k)$ for all $k \in \mathcal{K}_1 \cap \mathcal{K}_2$. Then, for all $k \in \mathcal{K}_1 \setminus \mathcal{K}_2$, set $\varphi(k) = (k, k^*)$, and for all $k \in \mathcal{K}_2 \setminus \mathcal{K}_1$, set $\varphi(k) = (k^*, k)$. So far, we have not assigned any pair twice and the two conditions of the lemma are already satisfied, so we can simply assign the remaining elements in $\mathcal{K} \setminus (\mathcal{K}_1 \cup \mathcal{K}_2)$ arbitrarily.

Second, suppose that $\mathcal{K}_1$ and $\mathcal{K}_2$ are disjoint. Note that for this being possible, $\sqrt{K+1}$ must be at least 2. Pick some distinct $k_1, k_1' \in \mathcal{K}_1$ and $k_2, k_2' \in \mathcal{K}_2$. Set $\varphi(k_1) := (k_1, k_2)$, $\varphi(k_1') := (k_1', k_2')$, $\varphi(k_2) := (k_1', k_2)$, and $\varphi(k_2') := (k_1, k_2')$. Then, for all $k \in \mathcal{K}_1 \setminus \{k_1, k_1'\}$, set $\varphi(k) := (k, k_2)$ and for all $k \in \mathcal{K}_2 \setminus \{k_2, k_2'\}$, set $\varphi(k) := (k_1, k)$. Again, we have not assigned any pair twice and the two conditions of the lemma are already satisfied, so we can simply assign the remaining elements in $\mathcal{K} \setminus (\mathcal{K}_1 \cup \mathcal{K}_2)$ arbitrarily. □

With this lemma, we can define $\widehat{M}_1$ and $\widehat{M}_2$. Let φ be the bijection from Lemma 1 and suppose $M_1 = \{(x^{(k)}, p^{(k)})\}_{k \in \mathcal{K}}$. We then define $\widehat{M}_1 = \{(x^{(\varphi_1(k))}, p^{(\varphi_1(k))})\}_{k \in \mathcal{K}}$, where $\varphi_1(k)$ is the first component of $\varphi(k)$. Similarly, $\widehat{M}_2$ is derived from M_2 by using the second component $\varphi_2(k)$ of $\varphi(k)$ instead of $\varphi_1(k)$. It remains to show that all menus on the three straight-line segments from M_1 via $\widehat{M}_1$ and $\widehat{M}_2$ to M_2 yield a revenue of at least $\min\{\mathtt{Rev}(M_1), \mathtt{Rev}(M_2)\}$, which is established by the following two propositions.

Proposition 2. *Let $M = \lambda M_1 + (1-\lambda)\widehat{M}_1$ be a convex combination of the menus M_1 and $\widehat{M}_1$. Then $\mathtt{Rev}(M) \geq \mathtt{Rev}(M_1)$. Similarly, every convex combination of the menus M_2 and $\widehat{M}_2$ has revenue at least $\mathtt{Rev}(M_2)$.*

Proof. We only prove the first statement on M_1; the statement on M_2 follows analogously. The idea is that, on the whole line segment from M_1 to $\widehat{M}_1$, the only active options are those in $\mathcal{K}'$, implying that the revenue does not decrease.

We show that for each possible valuation v of the buyer, the price paid to the seller for menu M is at least as high as in menu M_1. Suppose for valuation v that the buyer chooses the k-th option in menu M_1. Note that we may assume $k \in \mathcal{K}_1$ because of 0-reducibility of M_1. By construction of φ, it follows that $\varphi_1(k) = k$. Therefore, the k-th option in M is exactly equal to the k-th option in M_1. Making use of the fact that ties are broken in favor of larger prices, it suffices to show that the k-th option is utility-maximizing in M, too.

To this end, let $k' \in \mathcal{K}$ be an arbitrary index. If $M_1 = \{(x^{(k)}, p^{(k)})\}_{k \in \mathcal{K}}$, then the utility of option k' in M is

$$\begin{aligned}
&v^\top(\lambda x^{(k')} + (1-\lambda)x^{(\varphi_1(k'))}) - (\lambda p^{(k')} + (1-\lambda)p^{(\varphi_1(k'))}) \\
&\quad = \lambda(v^\top x^{(k')} - p^{(k')}) + (1-\lambda)(v^\top x^{(\varphi_1(k'))} - p^{(\varphi_1(k'))}) \\
&\quad \leq \lambda(v^\top x^{(k)} - p^{(k)}) + (1-\lambda)(v^\top x^{(k)} - p^{(k)}) \\
&\quad = v^\top x^{(k)} - p^{(k)},
\end{aligned}$$

where the inequality follows because the k-th option is utility-maximizing for menu M_1. This shows that it is utility-maximizing for menu M, completing the proof. □

Proposition 3. *Let $M = \lambda\widehat{M}_1 + (1-\lambda)\widehat{M}_2$ be a convex combination of the menus $\widehat{M}_1$ and $\widehat{M}_2$. Then, $\mathtt{Rev}(M) \geq \lambda\mathtt{Rev}(\widehat{M}_1) + (1-\lambda)\mathtt{Rev}(\widehat{M}_2)$.*

The idea to prove Proposition 3 is that, because of the special structure provided by Lemma 1, a linear interpolation between the menus also provides a linear interpolation between the revenues. Note that without the construction of Lemma 1, such a linear relation may not hold, as demonstrated by an example in Appendix A.

Proof of Proposition 3. The claim is trivial for $\lambda = 0$ or $\lambda = 1$. Therefore, assume $0 < \lambda < 1$ for the remainder of the proof. Again, we show that the claim holds pointwise for each possible valuation and therefore also for the revenue. For valuation v, let k_1 and k_2 be the active option assigned to the buyer in $\widehat{M}_1$ and $\widehat{M}_2$, respectively. Note that by construction of the menus $\widehat{M}_1$ and $\widehat{M}_2$, we may assume without loss of generality that $k_1 \in \mathcal{K}_1$ and $k_2 \in \mathcal{K}_2$. Let $k^* := \varphi^{-1}(k_1, k_2)$.

We show that option k^* is utility-maximizing in $M = \{(x^{(k)}, p^{(k)})\}_{k\in\mathcal{K}}$. To this end, we use the notation $\widehat{M}_1 = \{(\hat{x}^{(k)}, \hat{p}^{(k)})\}_{k\in\mathcal{K}}$ and $\widehat{M}_2 = \{(\hat{y}^{(k)}, \hat{q}^{(k)})\}_{k\in\mathcal{K}}$. Let $k' \in \mathcal{K}$ be an arbitrary index. The utility of option k' in menu M can be bounded as follows:

$$\begin{aligned} v^\top x^{(k')} - p^{(k')} &= \lambda(v^\top \hat{x}^{(k')} - \hat{p}^{(k')}) + (1-\lambda)(v^\top \hat{y}^{(k')} - \hat{q}^{(k')}) \\ &\le \lambda(v^\top \hat{x}^{(k_1)} - \hat{p}^{(k_1)}) + (1-\lambda)(v^\top \hat{y}^{(k_2)} - \hat{q}^{(k_2)}) \\ &= \lambda(v^\top \hat{x}^{(k^*)} - \hat{p}^{(k^*)}) + (1-\lambda)(v^\top \hat{y}^{(k^*)} - \hat{q}^{(k^*)}) \\ &= v^\top x^{(k^*)} - p^{(k^*)}, \end{aligned}$$

where the inequality in the second line follows because k_1 and k_2 are utility-maximizing for $\widehat{M}_1$ and $\widehat{M}_2$, respectively, and the equality in the third line follows because, by construction, in menu $\widehat{M}_1$ option k^* is equivalent to option $k_1 = \varphi_1(k^*) \in \mathcal{K}_1$, and similarly in menu $\widehat{M}_2$ option k^* is equivalent to option $k_2 = \varphi_2(k^*) \in \mathcal{K}_2$. This concludes the proof that k^* is utility-maximizing. □

With the same reasoning as above, we obtain $p^{(k^*)} = \lambda\hat{p}^{(k_1)} + (1-\lambda)\hat{q}^{(k_2)}$, from which we conclude that the price achieved by the seller in menu M for valuation v is at least as high as the convex combination of the achieved prices for menus $\widehat{M}_1$ and $\widehat{M}_2$. Proposition 1 directly follows from Proposition 2 and Proposition 3. Based on Proposition 1, we can show our two main theorems for the RochetNet.

3.2. Mode Connectivity for ε-Reducible Menus

We are now ready to derive Theorem 1.

Proof of Theorem 1. We prove this result by showing that every ε-reducible menu M can be linearly transformed into a 0-reducible menu $\widetilde{M}$ such that each convex combination of M and $\widetilde{M}$ achieves a revenue of at least $\texttt{Rev}(M) - \varepsilon$. This transformation converting M_1 and M_2 to $\widetilde{M}_1$ and $\widetilde{M}_2$, respectively, yields the first and the fifth of the linear pieces transforming M_1 to M_2. Together with Proposition 1 applied to $\widetilde{M}_1$ and $\widetilde{M}_2$ serving as the second to fourth linear piece, the theorem then follows.

To this end, let M be an ε-reducible menu with options indexed by $k \in \mathcal{K}$. By definition, there is a subset $\mathcal{K}' \subseteq \mathcal{K}$ of at most $\sqrt{K+1}$ many options such that, with probability at least $1-\varepsilon$, the assigned active option is contained in $\mathcal{K}'$. Let $\widetilde{M}$ consist of the same allocations as M, but with modified prices. For an option $k \in \mathcal{K}'$, the price $\widetilde{p}^{(k)} = p^{(k)}$ in $\widetilde{M}$ is the same as in M. However, for an option $k \in \mathcal{K} \setminus \mathcal{K}'$, we set the price $\widetilde{p}^{(k)} > 1$ in $\widetilde{M}$ to be larger than the largest possible valuation of any option $\|v\|_1 \le 1$. It follows that such an option will never be selected and $\widetilde{M}$ is 0-reducible.

To complete the proof, let us look at the reward of a convex combination $M' = \lambda M + (1-\lambda)\widetilde{M}$. If for a particular valuation v the selected option in M was in $\mathcal{K}'$, then the same option will be selected in M'. This happens with probability at least $1-\varepsilon$. In any other case, anything can happen, but the revenue cannot worsen by more than the maximum possible valuation, which is $\|v\|_1 \le 1$. Therefore, $\texttt{Rev}(M) - \texttt{Rev}(M') \le \varepsilon \cdot 1 = \varepsilon$, completing the proof. □

3.3. Mode Connectivity for Large Menus

In this subsection we prove our second main result about mode connectivity for the RochetNet, namely that large menus are ε-mode-connected.

Theorem 2. *If two menus M_1 and M_2 for the RochetNet have size at least $\lceil \frac{4}{\varepsilon^2} \rceil^{2n}$, then they are ε-connected. Moreover, the curve transforming M_1 into M_2 is piecewise linear with only five pieces.*

The intuition behind this theorem is that if menus are large, then they should contain many redundant options. Indeed, as in the previous theorem, the strategy is as follows. We show that every menu M of size at least $\lceil \frac{4}{\varepsilon^2} \rceil^{2n}$ can be linearly transformed into a 0-reducible menu $\widetilde{M}$ such that each convex combination of M and $\widetilde{M}$ achieves a revenue of at least $\texttt{Rev}(M) - \varepsilon$. This transformation converting M_1 and M_2 to $\widetilde{M}_1$ and $\widetilde{M}_2$, respectively, yields the first and the fifth of the linear pieces transforming M_1 to M_2. Together with Proposition 1 applied to $\widetilde{M}_1$ and $\widetilde{M}_2$ serving as the second to fourth linear piece, the theorem then follows.

However, a more careful linear transformation is needed from M to $\widetilde{M}$ than in the previous theorem. It is not sufficient to only adapt the prices, but we also need to change the allocations of the menu options by rounding them to discretized values. Our approach is inspired by Dughmi et al. [20], but there are notable differences. Because the rounding may also modify the active option for each valuation, we have to carefully adapt the prices in order to make sure that for each valuation, the newly selected option is not significantly worse than the originally selected one. Finally, this property has to be proven not only for $\widetilde{M}$, but for every convex combination of M and $\widetilde{M}$.

After the above rounding, the number of possible allocations for any option is bounded by $\lceil \frac{4}{\varepsilon^2} \rceil^n$. Out of several options with the same allocation, the buyer would always choose the cheapest one, implying that the resulting menu $\widetilde{M}$ is 0-reducible.

To proceed with the formal proof, we will show how to convert any menu M of size at least $\lceil \frac{4}{\varepsilon^2} \rceil^{2n}$ into a 0-reducible menu $\widetilde{M}$ such that each convex combination of M and $\widetilde{M}$ achieves a revenue of at least $\texttt{Rev}(M) - \varepsilon$. Without loss of generality, we assume that M has size exactly $K+1 = \lceil \frac{4}{\varepsilon^2} \rceil^{2n}$.

To construct the menu $\widetilde{M}$ satisfying these requirements, we adapt techniques from Dughmi et al. [20].[4] In general, the idea is to discretize the allocations in the menu by a finite allocation set S (see Definition 5) of size at most $\sqrt{K+1} = \lceil \frac{4}{\varepsilon^2} \rceil^n$. However, because of the discretization, the buyer may choose an option with a much smaller price, providing a lower revenue compared with the original menu. To deal with this, we proportionally decrease the prices on the menu. Intuitively, this incentivizes the buyer to choose the option with an originally high price. We show, after this modification, the menu achieves a revenue of at least $\texttt{Rev}(M) - \varepsilon$.

For ease of notation, we will use $\tilde{\varepsilon} := \frac{\varepsilon^2}{4}$ and, therefore, $2\sqrt{\tilde{\varepsilon}} = \varepsilon$.

Definition 5. Let S be a (finite) set of allocations. We say that S is an $\tilde{\varepsilon}$-cover if, for every possible allocation x, there exists an allocation $\tilde{x} \in S$ such that for every possible valuation vector v we have that $v^\top x \geq v^\top \tilde{x} \geq v^\top x - \tilde{\varepsilon}$.

The following proposition shows that one can construct an $\tilde{\varepsilon}$-cover S with size at most $\lceil \frac{4}{\varepsilon^2} \rceil^n$.

Proposition 4. *If* $\|v\|_1 \leq 1$, *then*

$$S = \underbrace{\{\tilde{\varepsilon}s\}_{s=0}^{\lfloor \frac{1}{\tilde{\varepsilon}} \rfloor} \times \{\tilde{\varepsilon}s\}_{s=0}^{\lfloor \frac{1}{\tilde{\varepsilon}} \rfloor} \times \cdots \times \{\tilde{\varepsilon}s\}_{s=0}^{\lfloor \frac{1}{\tilde{\varepsilon}} \rfloor}}_{n\,terms}$$

is an $\tilde{\varepsilon}$*-cover, and* $|S| = \lceil \frac{1}{\tilde{\varepsilon}} \rceil^n = \lceil \frac{4}{\varepsilon^2} \rceil^n$.

Proof. For any allocation x, we can round it down to $\tilde{x}$, such that $\tilde{x}_j = \lfloor \frac{x_j}{\tilde{\varepsilon}} \rfloor \cdot \tilde{\varepsilon}$. It is not hard to see that $v^\top x \geq v^\top \tilde{x}$. Additionally, the inequality $v^\top \tilde{x} \geq v^\top x - \tilde{\varepsilon}$ follows as the total loss is at most $v^\top(\tilde{x} - x) \leq \|v\|_1 \|\tilde{x} - x\|_\infty \leq \tilde{\varepsilon}$. □

3.3.1. Construction of $\widetilde{M}$. Given S, we can construct $\widetilde{M}$ as follows. Each option $(x^{(k)}, p^{(k)})$ in menu M is modified to $(\tilde{x}^{(k)}, \widetilde{p}^{(k)})$ in menu $\widetilde{M}$, where $\tilde{x}^{(k)}$ is the corresponding allocation of $x^{(k)}$ in S and the price is set to $\widetilde{p}^{(k)} = (1 - \sqrt{\tilde{\varepsilon}})p^{(k)}$:

$$\widetilde{M} = \{(\tilde{x}^{(k)}, \widetilde{p}^{(k)})\}_{k \in \mathcal{K}}.$$

The following lemma shows that this construction indeed ensures that the reward decreases by at most ε.

Lemma 2. *It holds that* $\texttt{Rev}(\widetilde{M}) \geq \texttt{Rev}(M) - 2\sqrt{\tilde{\varepsilon}} = \texttt{Rev}(M) - \varepsilon$.

Proof. The following inequalities demonstrate the buyer who chooses option k in menu M will not choose option k' in menu $\widetilde{M}$ such that $p^{(k')} < p^{(k)} - \sqrt{\tilde{\varepsilon}}$.

$$\begin{aligned} v^\top \tilde{x}^{(k)} - (1 - \sqrt{\tilde{\varepsilon}})p^{(k)} &\geq v^\top x^{(k)} - p^{(k)} - \tilde{\varepsilon} + \sqrt{\tilde{\varepsilon}}p^{(k)} \\ &\geq v^\top x^{(k')} - p^{(k')} - \tilde{\varepsilon} + \sqrt{\tilde{\varepsilon}}p^{(k)} \\ &\geq v^\top \tilde{x}^{(k')} - (1 - \sqrt{\tilde{\varepsilon}})p^{(k')} - \tilde{\varepsilon} + \sqrt{\tilde{\varepsilon}}(p^{(k)} - p^{(k')}) \\ &> v^\top \tilde{x}^{(k')} - (1 - \sqrt{\tilde{\varepsilon}})p^{(k')}. \end{aligned} \tag{2}$$

The first and third inequalities hold by Definition 5, and the second inequality holds as the buyer will choose option k in menu M_1.

Therefore, the total loss on the revenue is upper bounded by $\sqrt{\tilde{\varepsilon}}p^{(k)} + \sqrt{\tilde{\varepsilon}} \leq 2\sqrt{\tilde{\varepsilon}}$, as the price satisfies $p^{(k)} \leq 1$. □

In addition to this property of $\widetilde{M}$ itself, we also need to show the revenue does not drop more than ε for any menu on the line segment connecting M to $\widetilde{M}$.

Lemma 3. *Let* $M' = \lambda M + (1 - \lambda)\widetilde{M}$ *be a convex combination of the menus* M *and* $\widetilde{M}$. *Then,* $\texttt{Rev}(M') \geq \texttt{Rev}(M) - 2\sqrt{\tilde{\varepsilon}} = \texttt{Rev}(M) - \varepsilon$.

Proof. Let $M = \{(x^{(k)}, p^{(k)})\}_{k \in \mathcal{K}}$ and $M' = \{(x'(k), p'(k))\}_{k \in \mathcal{K}}$. Similar to the proof of Lemma 2, we show that the buyer who chooses option k in menu M will not choose option k' in menu M' such that $p^{(k')} < p^{(k)} - \sqrt{\tilde{\varepsilon}}$. This is true by the following (in)equalities. For any $k' \in \mathcal{K}$, we have that

$$\begin{aligned} v^\top x'(k) - p'(k) &= \lambda(v^\top x^{(k)} - p^{(k)}) + (1-\lambda)(v^\top \tilde{x}^{(k)} - \widetilde{p}^{(k)}) \\ &> \lambda(v^\top x^{(k')} - p^{(k')}) + (1-\lambda)(v^\top \tilde{x}^{(k')} - \widetilde{p}^{(k')}). \end{aligned}$$

The inequality follows by combining (i) $v^\top x^{(k)} - p^{(k)} \geq v^\top x^{(k')} - p^{(k')}$, which is true as the buyer will choose option k in menu M, and (ii) $v^\top \tilde{x}^{(k)} - \widetilde{p}^{(k)} > v^\top \tilde{x}^{(k')} - \widetilde{p}^{(k')}$ from (2).

Similar to the proof of Lemma 2, it follows than that the total loss on the revenue is upper bounded by $2\sqrt{\tilde{\varepsilon}}$. □

With these lemmas at hand, we can finally prove Theorem 2.

Proof of Theorem 2. Applying the transformation described in this section to convert M_1 and M_2 results in two menus, $\widetilde{M}_1$ and $\widetilde{M}_2$, respectively. Because $\widetilde{M}_1$ and $\widetilde{M}_2$ contain at most $\sqrt{K+1} = \lceil \frac{4}{\varepsilon^2} \rceil^n$ different allocations and a buyer would always choose the cheapest out of several options with the same allocation, they are 0-reducible. Applying Proposition 1 to them implies that they are 0-mode-connected with three linear pieces. Combining these observations with Lemmas 2 and 3 implies that M_1 and M_2 are ε-connected with five linear pieces. □

4. Mode Connectivity for the Affine Maximizer Auctions

Throughout this section, we focus on AMAs with fixed weights $w_i = 1$ for all buyers i. Similarly to *RochetNet*, we have the following definition for AMAs.

Definition 6. A menu M with $K+1$ options is ε-reducible if and only if there exists a subset $\mathcal{K}' \subseteq \mathcal{K}$, $0 \in \mathcal{K}'$, $|\mathcal{K}'| \leq \sqrt{K+1}$ such that, with probability at least $1 - \frac{\varepsilon}{m}$ over the distribution of the valuation of the buyers, (i) $k(v_{-i}) \in \mathcal{K}'$ for any buyer i and (ii) $k(v) \in \mathcal{K}'$.

Such phenomena are observed in the experiments in Curry et al. [11, section 6.3].

4.1. Interpolating Between 0-Reducible Menus

Analogously to Proposition 1, the following proposition will be the key tool to derive our two main results, namely that two ε-reducible menus are always ε-connected (Theorem 3), and that two large menus are always ε-connected (Theorem 4).

Proposition 5. *If two AMA menus M_1 and M_2 are 0-reducible, then they are 0-connected. Moreover, the curve transforming M_1 into M_2 is piecewise linear with only three pieces.*

The proof idea is similar to the proof of Proposition 1 in *RochetNet*, but requires additional arguments because of the more intricate price structure.

Similar to *RochetNet*, we introduce two intermediate menus $\widehat{M}_1$ and $\widehat{M}_2$, and show that every menu in the piecewise linear interpolation from M_1 via $\widehat{M}_1$ and $\widehat{M}_2$ to M_2 yields a revenue of at least $\min\{\texttt{Rev}(M_1), \texttt{Rev}(M_2)\}$. Using that menu M_1 has only $\sqrt{K+1}$ nonredundant options, menu $\widehat{M}_1$ will be defined by repeating each of the $\sqrt{K+1}$ options $\sqrt{K+1}$ times. Menu $\widehat{M}_2$ will be derived from M_2 similarly.

To make this more formal, let $\mathcal{K}'_1$ (and $\mathcal{K}'_2$) denote the set of the indexes of options in M_1 (and M_2) in definition of ε-reducibility, respectively. Let $\varphi : \mathcal{K} \to \mathcal{K}'_1 \times \mathcal{K}'_2$ as in Lemma 1. We define $\widehat{M}_1$ and $\widehat{M}_2$ as $\widehat{M}_1 := \{(x^{(\varphi_1(k))}, \beta^{(\varphi_1(k))})\}_{k\in\mathcal{K}}$, where $\varphi_1(k)$ is the first component of $\varphi(k)$, and, similarly, $\widehat{M}_2$ is derived from M_2 by using the second component $\varphi_2(k)$ of $\varphi(k)$ instead of $\varphi_1(k)$.

It remains to show that all menus on the three straight-line segments from M_1 via $\widehat{M}_1$ and $\widehat{M}_2$ to M_2 yield revenue of at least $\min\{\texttt{Rev}(M_1), \texttt{Rev}(M_2)\}$.

Proposition 6. *Let $M = \lambda M_1 + (1-\lambda)\widehat{M}_1$ be a convex combination of the menus M_1 and $\widehat{M}_1$. Then $\texttt{Rev}(M) \geq \texttt{Rev}(M_1)$. Similarly, every convex combination of the menus M_2 and $\widehat{M}_2$ has revenue at least $\texttt{Rev}(M_2)$.*

Proof. We only prove the first statement; the second one follows analogously. We show that for each possible valuation $v \in V^{m\times n}$ (with $\|v_i\| \leq 1$ for all i) of the buyers, the total payment paid to the auctioneer for menu M is at least as high as in menu $M_1 = \{(x^{(k)}, \beta^{(k)})\}_{k\in\mathcal{K}}$. Suppose for valuation $v \in V^{m\times n}$ that the auctioneer chooses the $k(v)$-th option in menu M_1 in maximizing the boosted welfare. Note that we may assume $k(v) \in \mathcal{K}'_1$ because of 0-reducibility of M_1. By construction of φ, it follows that $\varphi_1(k(v)) = k(v)$. Therefore, the $k(v)$-th option in M exactly equals the $k(v)$-th option in M_1. Because ties are broken in favor of larger total payments, it suffices to show that the $k(v)$-th option is the one with the highest boosted welfare also in M.[5]

Let $k' \in \mathcal{K}$ be an arbitrary index. The boosted welfare of option k' in M is

$$\begin{aligned}
&\sum_i v_i^\top(\lambda x_i^{(k')} + (1-\lambda)x_i^{(\varphi_1(k'))})) + (\beta^{(k')} + (1-\lambda)\beta^{(\varphi_1(k'))}) \\
&\quad = \lambda\left(\sum_i v_i^\top x_i^{(k')} + \beta^{(k')}\right) + (1-\lambda)\left(\sum_i v_i^\top x_i^{(\varphi_1(k'))} + \beta^{(\varphi_1(k'))}\right) \\
&\quad \le \lambda\left(\sum_i v_i^\top x_i^{(k(v))} + \beta^{(k(v))}\right) + (1-\lambda)\left(\sum_i v_i^\top x_i^{(k(v))} + \beta^{(k(v))}\right) \\
&\quad = \sum_i v_i^\top x_i^{(k(v))} + \beta^{(k(v))},
\end{aligned}$$

where the inequality follows because the $k(v)$-th option is boosted welfare maximizing for menu M_1. This shows that $k(v)$ is also a boosted welfare maximizer for menu M, completing the proof. □

Proposition 7. *Let $M = \lambda\widehat{M}_1 + (1-\lambda)\widehat{M}_2$ be a convex combination of the menus $\widehat{M}_1$ and $\widehat{M}_2$. Then $\mathtt{Rev}(M) \ge \lambda\mathtt{Rev}(\widehat{M}_1) + (1-\lambda)\mathtt{Rev}(\widehat{M}_2)$.*

Proof. The claim is trivial for $\lambda = 0$ or $\lambda = 1$. Therefore, assume $0 < \lambda < 1$ for the remainder of the proof. For possible valuation $v \in V^{m\times n}$ such that $\|v_i\| \le 1$ for all i, let $k_1(v)$ and $k_2(v)$ be the boosted welfare maximizing options in $\widehat{M}_1$ and $\widehat{M}_2$, respectively. Note that by the construction of the menus $\widehat{M}_1$ and $\widehat{M}_2$, we may assume without loss of generality that $k_1(v) \in \mathcal{K}'_1$ and $k_2(v) \in \mathcal{K}'_2$. Let $k^*(v) := \varphi^{-1}(k_1(v), k_2(v))$.

We show that option $k^*(v)$ is boosted welfare maximizing in $M = \{(x^{(k)}, \beta^{(k)})\}_{k\in\mathcal{K}}$ with valuation v. To this end, we use the notation $\widehat{M}_1 = \{(\hat{x}^{(1,k)}, \hat{\beta}^{(1,k)})\}_{k\in\mathcal{K}}$ and $\widehat{M}_2 = \{(\hat{x}^{(2,k)}, \hat{\beta}^{(2,k)})\}_{k\in\mathcal{K}}$. Let $k' \in \mathcal{K}$ be an arbitrary index. Then, the boosted welfare of option k' can be bounded as follows:

$$\begin{aligned}
\sum_i v_i^\top x_i^{(k')} + \beta^{(k')} &= \lambda\left(\sum_i v_i^\top \hat{x}_i^{(1,k')} + \hat{\beta}^{(1,k')}\right) + (1-\lambda)\left(\sum_i v_i^\top \hat{x}_i^{(2,k')} + \hat{\beta}^{(2,k')}\right) \\
&\le \lambda\left(\sum_i v_i^\top \hat{x}_i^{(1,k_1(v))} + \hat{\beta}^{(1,k_1(v))}\right) + (1-\lambda)\left(\sum_{ij} v_i \hat{x}_{ij}^{(2,k_2(v))} + \hat{\beta}^{(2,k_2(v))}\right) \\
&= \lambda\left(\sum_i v_i^\top \hat{x}_i^{(1,k^*(v))} + \hat{\beta}^{(1,k^*(v))}\right) + (1-\lambda)\left(\sum_i v_i^\top \hat{x}_i^{(1,k^*(v))} + \hat{\beta}^{(1,k^*(v))}\right) \\
&= \sum_i v_i^\top x_i^{(k^*(v))} + \beta^{(k^*(v))},
\end{aligned}$$

where the inequality in the second line follows because $k_1(v)$ and $k_2(v)$ are boosted welfare maximizers for $\widehat{M}_1$ and $\widehat{M}_2$, respectively. The equality in the third line follows because, by construction, in menu $\widehat{M}_1$ option $k^*(v)$ is equivalent to option $k_1(v) = \varphi_1(k^*(v)) \in \mathcal{K}'_1$. Similarly, in menu $\widehat{M}_2$ option $k^*(v)$ is equivalent to option $k_2(v) = \varphi_2(k^*(v)) \in \mathcal{K}'_2$. This concludes the proof that $k^*(v)$ is a boosted welfare maximizer in M.

Recall that given the valuation v, and the selected menu option $k(v)$, the total revenue of the auctioneer is

$$\sum_i p_i(v) = \sum_i \left(\sum_{l\neq i} v_l^\top x_l^{(k(v_{-i}))} + \beta^{(k(v_{-i}))}\right) - \sum_i \left(\sum_{l\neq i} v_l^\top x_l^{(k(v))} + \beta^{(k(v))}\right).$$

We have already shown that $k^*(v)$ is among the boosted welfare maximizers. As ties are broken in favor of larger total payments, it follows that the total revenue of the auctioneer is

$$\sum_i p_i(v) \ge \sum_i \left(\sum_{l\neq i} v_l^\top x_l^{(k^*(v_{-i}))} + \beta^{(k^*(v_{-i}))}\right) - \sum_i \left(\sum_{l\neq i} v_l^\top x_l^{(k^*(v))} + \beta^{(k^*(v))}\right).$$

Because $k^*(v)$ is fixed for different λ and, by linear combination, it holds that $\hat{x}^{(k^*(\cdot))} = \lambda\hat{x}^{(1,k^*(\cdot))} + (1-\lambda)\hat{x}^{(2,k^*(\cdot))}$ and $\hat{\beta}^{(k^*(\cdot))} = \lambda\hat{\beta}^{(1,k^*(\cdot))} + (1-\lambda)\hat{\beta}^{(2,k^*(\cdot))}$,

$$\begin{aligned}
&\sum_i \left(\sum_{l\neq i} v_l^\top x_l^{(k^*(v_{-i}))} + \beta^{(k^*(v_{-i}))}\right) - \sum_i \left(\sum_{l\neq i} v_l^\top x_l^{(k^*(v))} + \beta^{(k^*(v))}\right)\\
&= \lambda\left[\sum_i \left(\sum_{l\neq i} v_l^\top \hat{x}_l^{(1,k^*(v_{-i}))} + \hat{\beta}^{(1,k^*(v_{-i}))}\right) - \sum_i \left(\sum_{l\neq i} v_l^\top \hat{x}_l^{(1,k^*(v))} + \hat{\beta}^{(1,k^*(v))}\right)\right]\\
&\quad + (1-\lambda)\left[\sum_i \left(\sum_{l\neq i} v_l^\top \hat{x}_l^{(2,k^*(v_{-i}))} + \hat{\beta}^{(2,k^*(v_{-i}))}\right) - \sum_i \left(\sum_{l\neq i} v_l^\top \hat{x}_l^{(2,k^*(v))} + \hat{\beta}^{(2,k^*(v))}\right)\right]\\
&= \lambda \texttt{Rev}(\widehat{M}_1) + (1-\lambda)\texttt{Rev}(\widehat{M}_2).
\end{aligned}$$

This completes the proof. □

By construction of $\widehat{M}_1$ and $\widehat{M}_2$, Proposition 5 now follows immediately from Propositions 6 and 7. Based on this, we are able to show our two main results about AMAs.

4.2. Mode Connectivity for ε-Reducible Menus

Our first result achieves ε-connectivity from ε-reducibility.

Theorem 3. *If two AMAs M_1 and M_2 are ε-reducible, then they are ε-mode-connected. Moreover, the curve transforming M_1 to M_2 is piecewise linear with only five pieces.*

Before the proof, we recall how the total payment is calculated for a valuation profile v. We choose $k(v)$ as the option which maximizes the boosted welfare, $\sum_i v_i^\top x_i^{(k)} + \beta^{(k)}$. According to (1), the total revenue can be written as

$$\sum_i p_i(v) = \sum_i \underbrace{\left(\sum_{\ell\neq i} v_\ell^\top x_\ell^{(k(v_{-i}))} + \beta^{(k(v_{-i}))}\right)}_{\text{boosted welfare of } v_{-i}} - (m-1)\underbrace{\left(\sum_i v_i^\top x_i^{(k(v))} + \beta^{(k(v))}\right)}_{\text{boosted welfare of } v} - \beta^{(k(v))}. \tag{3}$$

Proof of Theorem 3. Similar to the proof of Theorem 1, it is sufficient to show that every ε-reducible menu M can be linearly transformed into a 0-reducible menu $\widetilde{M}$ such that each convex combination of M and $\widetilde{M}$ achieves a revenue of at least $\texttt{Rev}(M) - \varepsilon$. This can then be used as the first and fifth linear piece of the curve connecting M_1 and M_2, whereas the middle three pieces are provided by Proposition 5.

We construct $\widetilde{M}$ by (i) keeping all options in $\mathcal{K}'$ unchanged; (ii) for the options $k \in \mathcal{K}\setminus\mathcal{K}'$, we decrease $\beta^{(k)}$ to be smaller than $-m$, which implies such an option will never be selected (recall that $0 \in \mathcal{K}'$ is assumed, and the option $(\mathbf{0}, 0)$ is better than any such option). Consequently, $\widetilde{M}$ is 0-reducible.

To complete the proof, let us look at the revenue of $M' = \{(x'(k), \beta'^{(k)})\}_{k\in\mathcal{K}}$, which is a convex combination of M and $\widetilde{M}$: $M' = \lambda M + (1-\lambda)\widetilde{M}$ for $0 \le \lambda < 1$. Let $k'(v) = \arg\max_k \sum_i v_i^\top x_i'^{(k)} + \beta'^{(k)}$. As we decrease $\beta^{(k)}$ for $k \notin \mathcal{K}'$, $k(v) \in \mathcal{K}'$ implies $k'(v) \in \mathcal{K}'$ and, additionally, option $k'(v)$ and option $k(v)$ achieve the same boosted welfare and same β. Therefore, because M is ε-reducible, with probability at least $1 - \frac{\varepsilon}{m}$, the boosted welfare of v as well as the boosted welfare of v_{-i} for all buyers i is the same for M and for M'. According to Equation (3), the total payment for the profile v is the same for M and M'. Therefore, the loss on the revenue can only appear with probability at most $\frac{\varepsilon}{m}$, and the maximum loss is at most m, which implies an ε loss in total. □

4.3. Mode Connectivity for Large Menus

Our second result shows that mode connectivity also holds for those AMAs with large menu sizes, namely for $K+1 \ge \lceil\frac{16m^3}{\varepsilon^2}\rceil^{2nm}$.

Theorem 4. *For any $0 < \varepsilon \le \frac{1}{4}$, if two AMAs M_1 and M_2 have at least $K+1 \ge \lceil\frac{16m^3}{\varepsilon^2}\rceil^{2nm}$ options, then they are ε-mode-connected. Moreover, the curve transforming M_1 to M_2 is piecewise linear with only five pieces.*

Similar to *RochetNet*, the idea of proving Theorem 4 is to discretize the allocations in the menu; then one can use Proposition 5 to construct the low-loss transformation from M_1 to M_2 by five linear pieces. To do this, one wants the loss of revenue to be small during the discretization. Consider Equation (3) of the total payment. The first two terms do not change much by a small change of the discretization. However, the last term $\beta^{(k(v))}$ might be significantly affected by discretization, which may cause a notable decrease in the total payment. To avoid this, we perform a proportional discount on β, incentivizing the auctioneer to choose an allocation with a small β. By this approach, the original revenue will be approximately maintained. Furthermore, we show a linear path, connecting the original menu and the menu after discretizing, which will suffer a small loss.

Lemma 4. *Consider an AMA M_1 with at least $K+1=\lceil\frac{16m^3}{\varepsilon^2}\rceil^{2nm}$ options. There exists a 0-reducible menu $\widetilde{M}_1$, such that, for any linear combination of M_1 and $\widetilde{M}_1$, $M=\lambda M_1+(1-\lambda)\widetilde{M}_1$ for $\lambda\in[0,1]$, $\texttt{Rev}(M)\geq\texttt{Rev}(M_1)-\varepsilon$.*

Theorem 4 simply follows by combining Lemma 4 and Proposition 5.

We now formally define $\widetilde{M}_1$. We introduce parameters $\tilde{\varepsilon}$ and δ, which will be specified later.

4.3.1. Construction of $\widetilde{M}_1$. For $x^{(k)}$, we round it to $\tilde{x}^{(k)}$ in which $\tilde{x}^{(k)}_{ij}=\frac{\tilde{\varepsilon}}{m}\left\lfloor\frac{mx^{(k)}_{ij}}{\tilde{\varepsilon}}\right\rfloor$. With this rounding, for any $v\in V^{m\times n}$ such that $\|v_i\|\leq 1$ for i,

$$\sum_i v_i^\top x_i^{(k)}\geq\sum_i v_i^\top\tilde{x}_i^{(k)}\geq\sum_i v_i^\top x_i^{(k)}-\tilde{\varepsilon}\,, \tag{4}$$

where the second inequality uses $v_i^\top(x_i^{(k)}-\tilde{x}_i^{(k)})\leq\|v_i\|_1\|x_i^{(k)}-\tilde{x}_i^{(k)}\|_\infty\leq\frac{\tilde{\varepsilon}}{m}$. Further, let $\tilde{\beta}^{(k)}:=(1-\delta)\beta^{(k)}$.

Note that the payments and allocations only depend on those boosted welfare maximizing options. Therefore, to show that $\widetilde{M}_1$ is 0-reducible, it suffices to show $\widetilde{M}_1$ has at most $\sqrt{K+1}$ different allocations.

Lemma 5. *For any given $0<\varepsilon\leq\frac{1}{4}$, let $\delta=\frac{\sqrt{\tilde{\varepsilon}}}{m}$ and $\tilde{\varepsilon}=\frac{\varepsilon^2}{16m^2}$. Then,*

$$\texttt{Rev}(\widetilde{M}_1)\geq\texttt{Rev}(M_1)-\varepsilon.$$

The number of different allocations in $\widetilde{M}_1$ is at most $\lceil\frac{16m^3}{\varepsilon^2}\rceil^{nm}$. Additionally, for any linear combination of M_1 and $\widetilde{M}_1$, $M=\lambda M_1+(1-\lambda)\widetilde{M}_1$, $\texttt{Rev}(M)\geq\texttt{Rev}(M_1)-\varepsilon$.

Lemma 5 proves Lemma 4 and therefore Theorem 4. The only remaining part is to prove Lemma 5.

Proof of Lemma 5. We first demonstrate $\texttt{Rev}(\widetilde{M}_1)\geq\texttt{Rev}(M_1)-m\tilde{\varepsilon}-\frac{m^2\delta}{1-\delta}-\frac{\tilde{\varepsilon}}{\delta}$. The result follows by picking $\delta=\frac{\sqrt{\tilde{\varepsilon}}}{m}$ and $\tilde{\varepsilon}=\frac{\varepsilon^2}{16m^2}$. The proof of the bound on the linear combination of M_1 and $\widetilde{M}_1$ is analogous. We use the notation $M_1=\{x^{(k)},\beta^{(k)}\}_{k\in\mathcal{K}}$ and $\widetilde{M}_1=\{\tilde{x}^{(k)},\tilde{\beta}^{(k)}\}_{k\in\mathcal{K}}$.

We fix the valuation v. Let $k(v)=\arg\max_k\sum_i v_i^\top x_i^{(k)}+\beta^{(k)}$ and satisfy the tie-breaking rule. The total payment using M_1 can be expressed as follows:

$$\begin{aligned}&\sum_i\left(\sum_{l\neq i}v_l^\top x_l^{(k(v_{-i}))}+\beta^{(k(v_{-i}))}\right)-\sum_i\left(\sum_{l\neq i}v_l^\top x_l^{(k(v))}-\beta^{(k(v))}\right)\\&=\underbrace{\sum_i\left(\sum_{l\neq i}v_l^\top x_l^{(k(v_{-i}))}+\beta^{(k(v_{-i}))}\right)}_{A_1}\underbrace{-(m-1)\left(\sum_i v_i^\top x_i^{(k(v))}+\beta^{(k(v))}\right)}_{B_1}-\beta^{(k(v))}.\end{aligned}\tag{5}$$

Similarly, let $\tilde{k}(v)=\arg\max_k\sum_i v_i^\top\tilde{x}_i^{(k)}+\tilde{\beta}^{(k)}$; then, the total payment with $\widetilde{M}_1$ is

$$\underbrace{\sum_i\left(\sum_{l\neq i}v_l^\top\tilde{x}_l^{(\tilde{k}(v_{-i}))}+\tilde{\beta}^{(\tilde{k}(v_{-i}))}\right)}_{A_2}\underbrace{-(m-1)\left(\sum_i v_i^\top\tilde{x}_i^{(\tilde{k}(v))}+\tilde{\beta}^{(\tilde{k}(v))}\right)}_{B_2}-\tilde{\beta}^{(\tilde{k}(v))}.\tag{6}$$

We bound the differences between A_1 and A_2, B_1 and B_2, and $\beta^{(k(v))}$ and $\tilde{\beta}^{(\tilde{k}(v))}$ separately.

First, for the difference between A_1 and A_2, we can use the following inequalities: for any possible v such that $\|v_i\|\leq 1$ for all i,

$$\begin{aligned}\sum_i v_i^\top x_i^{(k(v))}+\beta^{(k(v))}&\leq\sum_i v_i^\top\tilde{x}_i^{(k(v))}+\tilde{\beta}^{(k(v))}+\tilde{\varepsilon}+\delta\beta^{(k(v))}\\&\leq\sum_i v_i^\top\tilde{x}_i^{(\tilde{k}(v))}+\tilde{\beta}^{(\tilde{k}(v))}+\tilde{\varepsilon}+\delta\beta^{(k(v))}.\end{aligned}$$

This implies

$$A_1 \le A_2 + m\tilde{\varepsilon} + \sum_i \delta\beta^{(k(v_{-i}))}. \tag{7}$$

Second, for the difference between B_1 and B_2, we can use the following inequalities: for any possible v such that $\|v_i\| \le 1$ for all i,

$$\begin{aligned}\sum_i v_i^\top x_i^{(k(v))} + \beta^{(k(v))} &\ge \sum_i v_i^\top x_i^{(\tilde{k}(v))} + \beta^{(\tilde{k}(v))}\\ &\ge \sum_i v_i^\top \tilde{x}_i^{(\tilde{k}(v))} + \tilde{\beta}^{(\tilde{k}(v))} + \delta\beta^{(\tilde{k}(v))}.\end{aligned}$$

This implies

$$B_1 \le B_2 - (m-1)\delta\beta^{(\tilde{k}(v))}. \tag{8}$$

Finally, we want to claim

$$\beta^{(\tilde{k}(v))} \le \beta^{(k(v))} + \frac{\tilde{\varepsilon}}{\delta}, \text{ which implies } \tilde{\beta}^{(\tilde{k}(v))} \le \beta^{(k(v))} + \frac{\tilde{\varepsilon}}{\delta} - \delta\beta^{(\tilde{k}(v))}; \tag{9}$$

as otherwise if $\beta^{(\tilde{k}(v))} > \beta^{(k(v))} + \frac{\tilde{\varepsilon}}{\delta}$, then this implies

$$\begin{aligned}\sum_i v_i^\top \tilde{x}_i^{(k(v))} + \tilde{\beta}^{(k(v))} &\ge \sum_i v_i^\top x_i^{(k(v))} + \beta^{(k(v))} - \tilde{\varepsilon} - \delta\beta^{(k(v))}\\ &\ge \sum_i v_i^\top x_i^{(\tilde{k}(v))} + \beta^{(\tilde{k}(v))} - \tilde{\varepsilon} - \delta\beta^{(k(v))}\\ &\ge \sum_i v_i^\top \tilde{x}_i^{(\tilde{k}(v))} + \tilde{\beta}^{(\tilde{k}(v))} - \tilde{\varepsilon} + \delta(\beta^{(\tilde{k}(v))} - \beta^{(k(v))})\\ &> \sum_i v_i^\top \tilde{x}_i^{(\tilde{k}(v))} + \tilde{\beta}^{(\tilde{k}(v))},\end{aligned} \tag{10}$$

which contradicts the fact that $\tilde{k}(v) = \arg\max_k \sum_i v_i^\top \tilde{x}_i^{(k)} + \tilde{\beta}^{(k)}$.

By combining the formula of total payment with M_1, (5), the formula of total payment with $\widetilde{M}_1$, (6), and Inequalities (7, 8), and (9), the loss on the total payment is at most $m\tilde{\varepsilon} + \sum_i \delta\beta^{(k(v_{-i}))} - m\delta\beta^{(\tilde{k}(v))} + \frac{\tilde{\varepsilon}}{\delta}$. Note that $\tilde{\beta}^{(k(v_{-i}))} \le \tilde{\beta}^{(\tilde{k}(v))} + m$.[6] Therefore, the total loss on the payment is at most $m\tilde{\varepsilon} + \frac{m^2\delta}{1-\delta} + \frac{\tilde{\varepsilon}}{\delta}$, which is $\mathtt{Rev}(\widetilde{M}_1) \ge \mathtt{Rev}(M_1) - m\tilde{\varepsilon} - \frac{m^2\delta}{1-\delta} - \frac{\tilde{\varepsilon}}{\delta}$.

Now, we prove a similar result for M, which is a linear combination of M_1 and $\widetilde{M}_1$: $M = \lambda M_1 + (1-\lambda)\widetilde{M}_1$. Let $M = (x'(k), \beta'^{(k)})_{k\in\mathcal{K}}$ and $k'(v) = \arg\max_k \sum_i v_i^\top x_i'^{(k)} + \beta'^{(k)}$. The total payment of M is

$$\underbrace{\sum_i\left(\sum_{l\ne i} v_l^\top x_l'^{(k(v_{-i}))} + \beta'^{(k(v_{-i}))}\right)}_{A_3} \underbrace{- (m-1)\left(\sum_i v_i^\top x_i'^{(k(v))} + \beta'^{(k(v))}\right)}_{B_3} - \beta'^{(k(v))}. \tag{11}$$

Note that, by linear combination, for any k, $\sum_i v_i^\top x_i^{(k)} \ge \sum_i v_i^\top x_i'^{(k)} \ge \sum_i v_i^\top x_i^{(k)} - \tilde{\varepsilon}$ and $\beta'^{(k)} = (1-\delta')\beta^{(k)}$ such that $\delta' = (1-\lambda)\delta$. With similar proofs as (7) and (8), the following two inequalities hold:

$$A_1 \le A_3 + m\tilde{\varepsilon} + \sum_i \delta'\beta^{(k(v_{-i}))}; \tag{12}$$

$$B_1 \le B_3 - (m-1)\delta'\beta^{(k'(v))}. \tag{13}$$

And, similarly,

$$\beta^{(k'(v))} \le \beta^{(k(v))} + \frac{\tilde{\varepsilon}}{\delta}, \tag{14}$$

as otherwise $\beta^{(k'(v))} > \beta^{(k(v))} + \frac{\tilde{\varepsilon}}{\delta}$ implies

$$\begin{aligned}
&\sum_i v_i^\top x_i'^{(k(v))} + \beta'^{(k(v))} \\
&\quad = \lambda\left(\sum_i v_i^\top x_i^{(k(v))} + \beta^{(k(v))}\right) + (1-\lambda)\left(\sum_i v_i^\top \tilde{x}_i^{(k(v))} + \tilde{\beta}^{(k(v))}\right) \\
&\quad > \lambda\left(\sum_i v_i^\top x_i^{(\hat{k}(v))} + \beta^{(\hat{k}(v))}\right) + (1-\lambda)\left(\sum_i v_i^\top \hat{x}_i^{(k(v))} + \hat{\beta}^{(k(v))}\right) \\
&\quad = \sum_i v_i^\top x_i'^{(k'(v))} + \beta'^{(k'(v))}.
\end{aligned}$$

The strict inequality follows from (10) and $\lambda < 1$.

Therefore, combining inequalities (12, 13), and (14), the total payment with M_1, (5), and the total payment with M, (11), the total loss for M is at most $m\tilde{\varepsilon} + \frac{m^2\delta'}{1-\delta'} + \frac{\tilde{\varepsilon}}{\delta} \le m\tilde{\varepsilon} + \frac{m^2\delta}{1-\delta} + \frac{\tilde{\varepsilon}}{\delta}$.

The result follows by picking $\delta = \frac{\sqrt{\tilde{\varepsilon}}}{m}$ and $\tilde{\varepsilon} = \frac{\varepsilon^2}{16m^2}$. □

5. Conclusion

We have given theoretical evidence of mode connectivity in neural networks designed to learn auction mechanisms. Our results show that, for a sufficiently wide hidden layer, ε-mode connectivity holds in the strongest possible sense. Perhaps more practically, we have shown ε-mode connectivity under ε-reducibility, that is, the assumption that there is a sufficiently small subset of neurons that preserve most of the revenue. There is evidence for this assumption in previous work in differentiable economics. A systematic experimental study that verifies this assumption under various distributions and network sizes is left for future work.

Our results make a first step in providing theoretical arguments underlying the success of neural networks in mechanism design. Our focus was on some of the most basic architectures. A natural next step is to extend the arguments for AMA networks with variable weights w_i. Such a result will need to analyze a four-layer network, and thus could make headway into understanding the behavior of deep networks. Besides *RochetNet*, Dütting et al. [21] also proposed *RegretNet*, based on minimizing a regret objective. This network is also applicable to multiple buyers, but only provides approximate incentive compatibility, and has been extended in subsequent work, for example, Duan et al. [19], Feng et al. [22], and Golowich et al. [26]. The architecture is, however, quite different from *RochetNet*: it involves two deep neural networks in conjunction, an allocation and a payment network, and uses expected ex post regret as the loss function. Recently, Wang et al. [40] introduced GemNet, a deep learning-based framework for multibuyer auctions that is both menu-based and incentive compatible. The framework employs a menu network that determines the allocation and pricing for each buyer, using the valuations of other buyers as input. To ensure feasibility—specifically, that no item is allocated more than once—the authors incorporate a mixed-integer linear program and design a specialized loss function. Given these innovations, we expect mode connectivity analyses for *RegretNet* and *GemNet* to require a considerable extension of the techniques used by us. We believe that such analyses would be a significant next step in the theoretical analysis of neural networks in differentiable economics.

Acknowledgments

This work was partially completed while all three authors were affiliated with the London School of Economics and Political Science, United Kingdom, and while Christoph Hertrich was affiliated with Université Libre de Bruxelles, Belgium. A preliminary conference version (Hertrich et al. [28]) appeared in the proceedings of the Conference on Neural Information Processing Systems (NeurIPS 2023).

Appendix A. Example: Disconnected Local Maxima

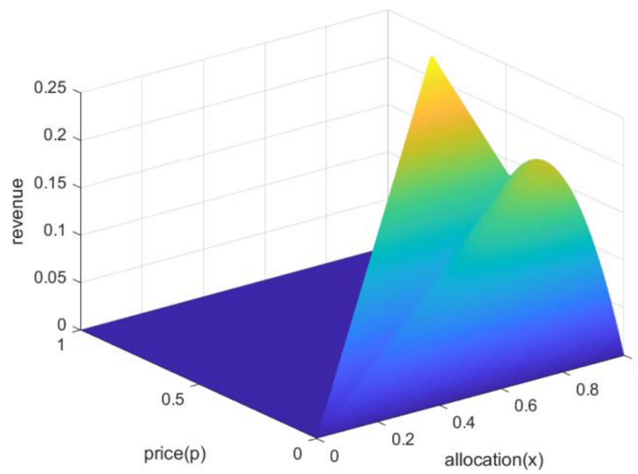
This section shows that the revenue is not quasiconcave on M, and in fact it might have disconnected local maxima. It therefore shows that the assumptions of our theorems (reducibility or sufficient width) are really necessary.

Recall that a function g is quasiconcave if and only if, for any x, y, and $\lambda \in [0,1]$,

$$g(\lambda x + (1-\lambda)y) \ge \min\{g(x), g(y)\}.$$

Hence, quasiconcavity implies 0-mode connectivity with a single straight-line segment.

Figure A.1. (Color online) Revenue of the mechanism $M = \{(x, p)\}$ when the value distribution is f.



We consider the case that there is only one buyer, one item, and one regular option on the menu. Consider the following value distribution f:

$$f(x) = \begin{cases} 1.5 & 0 < x \le \frac{1}{3} + 0.15 \\ 0 & \frac{1}{3} + 0.15 < x \le \frac{2}{3} + 0.15 \\ 1.5 & \frac{2}{3} + 0.15 < x \le 1. \end{cases} \tag{A.1}$$

With this probability distribution, we show the following result. As Figure A.1 shows, there are two local maxima so that any continuous curve connecting them has lower revenue than either endpoint. Hence, mode connectivity fails between these two points. We only give a formal proof of the fact that the revenue is not quasiconcave.

Lemma A.1. $\texttt{Rev}(M)$ *is not quasiconcave on* M.

Proof. We consider the case where $n = 1$ (single item case); $K = 1$ (menu with single options). The value distribution, f, is defined in (A.1).

We consider two menus, M_1 and M_2, where $M_1 = \{(0,0),(1,0.36)\}$ and $M_2 = \{(0,0),(1,0.84)\}$. Then, $\texttt{Rev}(M_1) = 0.1656$ and $\texttt{Rev}(M_2) = 0.2016$.

However, if we consider $M_3 = \frac{1}{2}(M_1 + M_2) = \{(0,0),(1,0.6)\}$, then this provides a revenue of 0.165, which is strictly smaller than $\texttt{Rev}(M_1)$ and $\texttt{Rev}(M_2)$. More intuitively, Figure A.1 shows the revenue for $x \in [0,1]$ and $p \in [0,1]$. □

Appendix B. Bounds on the Error of the Softmax Approximation for RochetNet

In the RochetNet, to ensure that the objective is a smooth function, a softmax operation is used instead of the argmax during the training process:

$$\texttt{Rev}^{\texttt{softmax}}(M) := \int \sum_{k=1}^{K} p_i \frac{e^{Y(x^{(k)\top} v - p^{(k)})}}{\sum_{k'=1}^{K} e^{Y(x^{(k')\top} v - p^{(k')})}} \mathrm{d}F(v). \tag{B.1}$$

Here, Y is a sufficiently large constant. In this section, we will look at the difference between the actual revenue and this softmax revenue.

We would like to assume the density of the valuation distribution is upper bounded by $\mathcal{X} = \max_{v \in [0,1]^n \text{ and } \|v\|_1 \le 1} f(v)$, which is a finite value. Given this assumption, the following lemma shows that, for any menu M of size K, the difference between the actual revenue and the softmax revenue is bounded.

Lemma B.1. *For any* M *and* $Y \ge 1$,

$$|\texttt{Rev}^{\texttt{softmax}}(M) - \texttt{Rev}(M)| \le \frac{K+1}{Y}\left(\left(n\mathcal{X} + 1 + \frac{\mathcal{X}}{Y}\right)\log\frac{Y}{\mathcal{X}} + \mathcal{X}\right).$$

Proof. We prove $\texttt{Rev}^{\texttt{softmax}}(M) - \texttt{Rev}(M) \le \frac{K}{Y}\left(\left(n\mathcal{X} + 1 + \frac{\mathcal{X}}{Y}\right)\log\frac{Y}{\mathcal{X}} + \mathcal{X}\right)$. $\texttt{Rev}(M) - \texttt{Rev}^{\texttt{softmax}}(M) \le \frac{K}{Y}\left(\left(n\mathcal{X} + 1 + \frac{\mathcal{X}}{Y}\right)\log\frac{Y}{\mathcal{X}} + \mathcal{X}\right)$ follows by a similar argument.

Let $k(v)$ be the option chosen in menu M when the buyer's valuation is v. Then, the difference between these two can be bounded as follows:

$$\begin{aligned}\texttt{Rev}^{\texttt{softmax}}(M) - \texttt{Rev}(M) &\le \int \sum_{k=0}^{K} (p^{(k)} - p^{(k(v))})^+ \cdot \frac{e^{Y(v^\top x^{(k)} - p^{(k)})}}{\sum_{k'=1}^{K} e^{Y(v^\top x^{(k')} - p^{(k')})}}\,\mathrm{d}F(v) \\ &\le \int \sum_{k=0}^{K} (p^{(k)} - p^{(k(v))})^+ e^{Y(v^\top x^{(k)} - p^{(k)} - v^\top x^{(k(v))} + p^{(k(v))})}\,\mathrm{d}F(v).\end{aligned}$$

Here, $(\cdot)^+ \triangleq \max\{\cdot, 0\}$. Now, we focus on one option k, and we will give an upper bound on

$$\int (p^{(k)} - p^{(k(v))})^+ 1_{v^\top x^{(k)} - p^{(k)} + \sigma \ge v^\top x^{(k(v))} - p^{(k(v))} \ge v^\top x^{(k)} - p^{(k)}}\,\mathrm{d}F(v) \tag{B.2}$$

for the nonnegative parameter σ, which will be specified later. Note that it is always true that $v^\top x^{(k(v))} - p^{(k(v))} \ge v^\top x^{(k)} - p^{(k)}$. If $v^\top x^{(k)} - p^{(k)} + \sigma \ge v^\top x^{(k(v))} - p^{(k(v))}$ is not satisfied, then $e^{Y(v^\top x^{(k)} - p^{(k)} - v^\top x^{(k(v))} + p^{(k(v))})} \le e^{-Y\sigma}$. Therefore, if (B.2) is upper bounded by $\mathcal{C}(\sigma)$, then $\texttt{Rev}_M^{\texttt{softmax}} - \texttt{Rev}(M) \le (K+1)(\mathcal{C}(\sigma) + (1+\sigma)e^{-Y\sigma})$.[7]

Note that

$$\int (p^{(k)} - p^{(k(v))})^+ 1_{v^\top x^{(k)} - p^{(k)} + \sigma \ge v^\top x^{(k(v))} - p^{(k(v))} \ge v^\top x^{(k)} - p^{(k)}}\,\mathrm{d}F(v) \le \sigma + \int \sum_{j=1}^{n} v_j (x_j^{(k)} - x_j^{(k(v))})^+ 1_{v^\top x^{(k)} - p^{(k)} + \sigma \ge v^\top x^{(k(v))} - p^{(k(v))} \ge v^\top x^{(k)} - p^{(k)}}\,\mathrm{d}F(v).$$

The inequality follows as we consider the region of v such that $v^\top x^{(k)} - p^{(k)} + \sigma \ge v^\top x^{(k(v))} - p^{(k(v))}$. Additionally, because $v_j \in [0,1]$,

$$\int v_j (x_j^{(k)} - x_j^{(k(v))})^+ 1_{v^\top x^{(k)} - p^{(k)} + \sigma \ge v^\top x^{(k(v))} - p^{(k(v))} \ge v^\top x^{(k)} - p^{(k)}}\,\mathrm{d}F(v) \le \int (x_j^{(k)} - x_j^{(k(v))})^+ 1_{v^\top x^{(k)} - p^{(k)} + \sigma \ge v^\top x^{(k(v))} - p^{(k(v))} \ge v^\top x^{(k)} - p^{(k)}}\,\mathrm{d}F(v).$$

Now we fix all coordinates of valuation v other than coordinate j. Note that the function $v^\top x^{(k(v))} - p^{(k(v))} - v^\top x^{(k)} - p^{(k)}$ is a convex function on v_j and $x_j^{(k)} - x_j^{(k(v))}$ is the negative gradient of this convex function. Because we are looking at the region such that the function $v^\top x^{(k(v))} - p^{(k(v))} - v^\top x^{(k)} - p^{(k)}$ is bounded in $[0, \sigma]$, this directly implies

$$\int_{v_j \in [0,1]} (x_j^{(k)} - x_j^{(k(v))})^+ 1_{v^\top x^{(k)} - p^{(k)} + \sigma \ge v^\top x^{(k(v))} - p^{(k(v))} \ge v^\top x^{(k)} - p^{(k)}}\,\mathrm{d}F(v) \le \mathcal{X}\sigma.$$

This implies $\texttt{Rev}_M^{\texttt{softmax}} - \texttt{Rev}(M) \le (K+1)(\sigma + n\mathcal{X}\sigma + (1+\sigma)e^{-Y\sigma})$, which is upper bounded by $\frac{K+1}{Y}\left((n\mathcal{X} + 1 + \frac{\mathcal{X}}{Y})\log\frac{Y}{\mathcal{X}} + \mathcal{X}\right)$ by setting $\sigma = \frac{1}{Y}\log\frac{Y}{\mathcal{X}}$. □

Appendix C. Bounds on the Error of the Softmax Approximation for AMA

Similar to the case of the *RochetNet*, in the training process of AMA, the softmax operation is used instead of argmax. Recall that the revenue of AMA is

$$\texttt{Rev} = \int_v \sum_i p_i(v)\mathrm{d}F(v),$$

where

$$p_i(v) = \left(\sum_{l \ne i} v_l^\top x_l^{(k(v_{-i}))} + \beta^{(k(v_{-i}))}\right) - \left(\sum_{l \ne i} v_l^\top x_l^{(k(v))} + \beta^{(k(v))}\right),$$

and

$$k(v) = \arg\max_k \sum_i v_i^\top x_i^{(k)} + \beta^{(k)}.$$

For the softmax version, instead of using $k(v)$, which exactly maximizes the boosted social welfare, now $k^{\texttt{softmax}}(v)$ is a random variable:

$$k^{\texttt{softmax}}(v) = k \text{ with probability } \frac{e^{Y(v_i^\top x_i^{(k)} + \beta^{(k)})}}{\sum_{k'} e^{Y(v_i^\top x_i^{(k')} + \beta^{(k')})}};$$

and the price is the expectation on $k(v)$

$$p_i^{\texttt{softmax}}(v) = \mathbb{E}\left[\left(\sum_{l \ne i} v_l^\top x_l^{(k^{\texttt{softmax}}(v_{-i}))} + \beta^{(k^{\texttt{softmax}}(v_{-i}))}\right) - \left(\sum_{l \ne i} v_l^\top x_l^{(k^{\texttt{softmax}}(v))} + \beta^{(k^{\texttt{softmax}}(v))}\right)\right];$$

and the revenue is

$$\texttt{Rev}^{\texttt{softmax}}(M) = \int_v \sum_i p_i^{\texttt{softmax}}(v)\mathrm{d}F(v).$$

We show the following result. Note that we also assume the maximal density of a valuation type is $\mathcal{X}$.

Theorem C.1. *It holds that*

$$|\mathtt{Rev}^{\mathtt{softmax}}(M) - \mathtt{Rev}(M)| \le \frac{m(K+1)}{eY} + \frac{nm\mathcal{X}(K+1)}{Y}\left(1 + \log\frac{mY}{m\mathcal{X}}\right).$$

To prove this theorem, we need the following lemma, which provides one of the basic properties of the softmax.

Lemma C.1. *Given L values, $a_1 \ge a_2 \ge a_3 \ge \cdots \ge a_L$, then $0 \le a_1 - \sum_k a_k \frac{e^{Ya_k}}{\sum_{k'} e^{Ya_{k'}}} \le \frac{L}{eY}$.*

Proof. It's clear that $0 \le a_1 - \sum_k a_k \frac{e^{Ya_k}}{\sum_{k'} e^{Ya_{k'}}}$. On the other direction,

$$\begin{aligned} a_1 - \sum_k a_k \frac{e^{Ya_k}}{\sum_{k'} e^{Ya_{k'}}} &\le \sum_k (a_1 - a_k)\frac{e^{Y(a_k - a_1)}}{\sum_{k'} e^{Y(a_{k'} - a_1)}} \\ &\le \frac{1}{Y}\sum_k Y(a_1 - a_k)\frac{e^{Y(a_k - a_1)}}{\sum_{k'} e^{Y(a_{k'} - a_1)}} \\ &\le \frac{1}{Y}\sum_k e^{Y(a_1 - a_k) - 1}\frac{e^{Y(a_k - a_1)}}{\sum_{k'} e^{Y(a_{k'} - a_1)}} \\ &\le \frac{L}{eY}\frac{1}{\sum_{k'} e^{Y(a_{k'} - a_1)}} \\ &\le \frac{L}{eY}. \quad \square \end{aligned}$$

Now, we can prove Theorem C.1.

Proof of Theorem C.1. We first give the upper bound on $\mathtt{Rev}(M) - \mathtt{Rev}^{\mathtt{softmax}}(M)$.

Let $k(v) = \arg\max_k \sum_i v_i^\top x_i^{(k)} + \beta^{(k)}$ be the rule used in $\mathtt{Rev}(M)$. Recall that

$$\sum_i p_i(v) = \sum_i \left(\sum_{l \ne i} v_l^\top x_l^{(k(v_{-i}))} + \beta^{(k(v_{-i}))}\right) - \sum_i \left(\sum_{l \ne i} v_l^\top x_l^{(k(v))} + \beta^{(k(v))}\right)$$

and

$$\sum_i p_i^{\mathtt{softmax}}(v) = \mathbb{E}\left[\sum_i \left(\sum_{l \ne i} v_l^\top x_l^{(k^{\mathtt{softmax}}(v_{-i}))} + \beta^{(k^{\mathtt{softmax}}(v_{-i}))}\right) - \sum_i \left(\sum_{l \ne i} v_l^\top x_l^{(k^{\mathtt{softmax}}(v))} + \beta^{(k^{\mathtt{softmax}}(v))}\right)\right].$$

Then,

$$\begin{aligned} &\mathtt{Rev}(M) - \mathtt{Rev}^{\mathtt{softmax}}(M) \\ &= \mathbb{E}_v\left[\sum_i \left(\sum_{l \ne i} v_l^\top x_l^{(k(v_{-i}))} + \beta^{(k(v_{-i}))}\right) - \mathbb{E}\left[\sum_i \left(\sum_{l \ne i} v_l^\top x_l^{(k^{\mathtt{softmax}}(v_{-i}))} + \beta^{(k^{\mathtt{softmax}}(v_{-i}))}\right)\right] - m\left(\sum_i v_i^\top x_i^{(k(v))} + \beta^{(k(v))}\right)\right. \\ &\quad \left. + \mathbb{E}\left[m\left(\sum_i v_i^\top x_i^{(k^{\mathtt{softmax}}(v))} + \beta^{(k^{\mathtt{softmax}}(v))}\right)\right] + \sum_i v_i^\top x_i^{(k(v))} - \mathbb{E}\left[\sum_i v_i^\top x_i^{(k^{\mathtt{softmax}}(v))}\right]\right] \\ &\le \frac{m(K+1)}{eY} + \mathbb{E}\left[\sum_i v_i^\top x_i^{(k(v))}\right] - \mathbb{E}\left[\sum_i v_i^\top x_i^{(k^{\mathtt{softmax}}(v))}\right]. \end{aligned}$$

The inequality follows by Lemma C.1. Then, we will bound the difference between $\mathbb{E}[\sum_i v_i^\top x_i^{(k(v))}]$ and $\mathbb{E}[\sum_i v_i^\top x_i^{(k^{\mathtt{softmax}}(v))}]$. More specifically,

$$\begin{aligned} &\mathbb{E}[v_{ij} x_{ij}^{(k(v))}] - \mathbb{E}[v_{ij} x_{ij}^{(k^{\mathtt{softmax}}(v))}] \\ &= \int_v v_{ij} x_{ij}^{(k(v))} - \sum_k v_{ij} x_{ij}^{(k)} \frac{e^{Y\left(\sum_{i'} v_{i'}^\top x_{i'}^{(k)} + \beta^{(k)}\right)}}{\sum_{k'} e^{Y\left(\sum_{i'} v_{i'}^\top x_{i'}^{(k')} + \beta^{(k')}\right)}} \mathrm{d}F(v) \\ &\le \int_v \sum_k (x_{ij}^{(k)} - x_{ij}^{(k(v))})^+ \frac{e^{Y\left(\sum_{i'} v_{i'}^\top x_{i'}^{(k)} + \beta^{(k)}\right)}}{\sum_{k'} e^{Y\left(\sum_{i'} v_{i'}^\top x_{i'}^{(k')} + \beta^{(k')}\right)}} \mathrm{d}F(v) \\ &\le \int_v \sum_k (x_{ij}^{(k)} - x_{ij}^{(k(v))})^+ e^{Y\left(\sum_{i'} v_{i'}^\top x_{i'}^{(k)} + \beta^{(k)} - \sum_{i'} v_{i'}^\top x_{i'}^{(k(v))} - \beta^{(k(v))}\right)} \mathrm{d}F(v). \end{aligned}$$

Recall that $(\cdot)^+ \triangleq \max\{\cdot, 0\}$. Let's define $\mathrm{BW}(k) = \sum_{i'} v_{i'}^\top x_{i'}^{(k)} + \beta^{(k)}$ to be the boosted welfare of option k for simplicity. Now, we focus on one option k, and we will give an upper bound on

$$\int (x_{ij}^{(k)} - x_{ij}^{(k(v))})^+ 1_{\mathrm{BW}(k)+\sigma \geq \mathrm{BW}(k(v)) \geq \mathrm{BW}(k)} \mathrm{d}F(v) \tag{C.1}$$

for the nonnegative σ. The value of σ will be determined later. Note that it is always true that $\mathrm{BW}(k(v)) \geq \mathrm{BW}(k)$ by the definition of $k(v)$. Additionally, if $\mathrm{BW}(k) + \sigma \geq \mathrm{BW}(k(v))$ is not satisfied, then $e^{Y\left(\sum_{i'} v_{i'}^\top x_{i'}^{(k)} + \beta^{(k)} - \sum_{i'} v_{i'}^\top x_{i'}^{(k(v))} - \beta^{(k(v))}\right)} \leq e^{-Y\sigma}$. Therefore, if (C.1) is upper bounded by $\mathcal{C}_{ij}(\sigma)$, then $\mathtt{Rev}^{\mathtt{softmax}}(M) - \mathtt{Rev}(M) \leq \frac{mK}{eY} + (K+1)(\sum_{ij} \mathcal{C}_{ij}(\sigma) + nme^{-Y\sigma})$.

Note that

$$\int (x_{ij}^{(k)} - x_{ij}^{(k(v))})^+ 1_{\mathrm{BW}(k)+\sigma \geq \mathrm{BW}(k(v)) \geq \mathrm{BW}(k)} \,\mathrm{d}F(v) = \int (x_{ij}^{(k)} - x_{ij}^{(k(v))})^+ 1_{\mathrm{BW}(k(v)) - \mathrm{BW}(K) \in [0,\sigma]} \,\mathrm{d}F(v). \tag{C.2}$$

Now we fix all coordinates of valuation v other than coordinate ij. Note that the function $\mathrm{BW}(k(v)) - \mathrm{BW}(k) = \sum_{i'} v_{i'}^\top x_{i'}^{(k(v))} + \beta^{(k(v))} - \sum_{i'} v_{i'}^\top x_{i'}^{(k)} - \beta^{(k)}$ is a convex function on v_{ij} and $(x_{ij}^{(k(v))} - x_{ij}^{(k)})$ is the gradient. Because we are looking at the region such that the function $\mathrm{BW}(k(v)) - \mathrm{BW}(k)$ is bounded in $[0, \sigma]$, this directly implies

$$\int (x_{ij}^{(k)} - x_{ij}^{(k(v))})^+ 1_{\mathrm{BW}(k)+\sigma \geq \mathrm{BW}(k(v)) \geq \mathrm{BW}(k)} \,\mathrm{d}F(v) \leq \mathcal{X}\sigma$$

as the maximal density is at most $\mathcal{X}$. This implies $\mathtt{Rev}(M) - \mathtt{Rev}^{\mathtt{softmax}}(M) \leq \frac{m(K+1)}{eY} + (K+1)nm\mathcal{X}\sigma + (K+1)nme^{-Y\sigma}$, which is upper bounded by $\frac{m(K+1)}{eY} + \frac{nm\mathcal{X}(K+1)}{Y}\left(1 + \log\frac{Y}{\mathcal{X}}\right)$ by setting $\sigma = \frac{1}{Y}\log\frac{Y}{\mathcal{X}}$.

The upper bound on $\mathtt{Rev}^{\mathtt{softmax}}(M) - \mathtt{Rev}(M)$ follows by a similar argument. □

Endnotes

[1] MenuNet, developed by Shen et al. [37], also encodes menu items as RochetNet. However, unlike RochetNet's approach of repeatedly sampling valuations from the underlying distribution, MenuNet discretizes the buyer's valuation space into discrete values.

[2] Instead of making this normalization, one could also explicitly carry through $\|v_i\|_1$ in all our statements. For example, Theorems 1 and 2 would prove $(\varepsilon \cdot \max_i \|v_i\|_1)$-connectivity instead of ε-connectivity.

[3] It is not hard to see that our results transfer to other tie-breaking rules, too. See also the discussion in section B.2 of Babaioff et al. [3].

[4] In their paper, they construct a menu with a finite number of options to approximate the optimal mechanism. The approximation is based on the multiplicative error, and they assume the buyer's valuation is no less than one.

[5] Recall that the auctioneer will choose the option k to maximize the $\beta^{(k)}$ among all boosted welfare maximizing options given the formula of the total payment (3).

[6] This is true because $v^\top \tilde{x}^{(k(v_{-i}))} + \tilde{\beta}^{(k(v_{-i}))} \leq v^\top \tilde{x}^{(\tilde{k}(v))} + \tilde{\beta}^{(\tilde{k}(v))}$ and $v^\top \tilde{x}^{(k(v))} \leq m$.

[7] Note that if $p^{(k)} \geq 1 + \sigma$, then $v^\top x^{(k)} - p^{(k)} + (p^{(k)} - 1) \leq v^\top x^{(k(v))} - p^{(k(v))}$ as $\mathtt{LHS} \leq 0$ and $\mathtt{RHS} \geq 0$. Therefore, if $v^\top x^{(k)} - p^{(k)} + \sigma \geq v^\top x^{(k(v))} - p^{(k(v))}$ is not satisfied, then $(p^{(k)} - p^{(k(v))})^+ e^{Y(v^\top x^{(k)} - p^{(k)} - v^\top x^{(k(v))} + p^{(k(v))})} \leq \max_{\sigma' \geq \sigma} \{(1+\sigma')e^{-Y\sigma'}\}$. Note that $\max_{\sigma' \geq \sigma}\{(1+\sigma')e^{-Y\sigma'}\} \leq (1+\sigma)e^{-Y\sigma}$ when $Y \geq 1$.

References

[1] Allen-Zhu Z, Li Y, Song Z (2019) A convergence theory for deep learning via over-parameterization. *Proc. 36th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 97 (PMLR, New York), 242–252.

[2] Arora S, Ge R, Neyshabur B, Zhang Y (2018) Stronger generalization bounds for deep nets via a compression approach. *Proc. 35th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 80 (PMLR, New York), 254–263.

[3] Babaioff M, Gonczarowski YA, Nisan N (2022) The menu-size complexity of revenue approximation. *Games Econom. Behav.* 134(C):281–307.

[4] Bhalgat A, Gollapudi S, Munagala K (2013) Optimal auctions via the multiplicative weight method. *Proc. 14th ACM Conf. Electronic Commerce* (Association for Computing Machinery, New York), 73–90.

[5] Cai Y, Daskalakis C, Weinberg SM (2012) An algorithmic characterization of multi-dimensional mechanisms. *Proc. 44th Annual ACM Sympos. Theory Comput.* (Association for Computing Machinery, New York), 459–478.

[6] Cai Y, Daskalakis C, Weinberg SM (2012) Optimal multi-dimensional mechanism design: Reducing revenue to welfare maximization. *2012 IEEE 53rd Annual Sympos. Foundations Comput. Sci.* (IEEE, New York), 130–139.

[7] Cai Y, Daskalakis C, Weinberg SM (2013) Understanding incentives: Mechanism design becomes algorithm design. *2013 IEEE 54th Annual Sympos. Foundations Comput. Sci.* (IEEE, New York), 618–627.

[8] Clarke EH (1971) Multipart pricing of public goods. *Public Choice* 11:17–33.

[9] Conitzer V, Sandholm T (2002) Complexity of mechanism design. Darwiche A, Friedman N, eds. *Proc. 18th Conf. Uncertainty Artificial Intelligence* (Morgan Kaufmann Publishers Inc., San Francisco), 103–110.

[10] Conitzer V, Sandholm T (2004) Self-interested automated mechanism design and implications for optimal combinatorial auctions. *Proc. 5th ACM Conf. Electronic Commerce* (Association for Computing Machinery, New York), 132–141.

[11] Curry M, Sandholm T, Dickerson J (2022) Differentiable economics for randomized affine maximizer auctions. Preprint, submitted February 6, https://arxiv.org/abs/2202.02872.

[12] Curry MJ, Lyi U, Goldstein T, Dickerson JP (2022) Learning revenue-maximizing auctions with differentiable matching. *Proc. 25th Internat. Conf. Artificial Intelligence Statist.*, Proceedings of Machine Learning Research, vol. 151 (PMLR, New York), 6062–6073.

[13] Daskalakis C (2015) Multi-item auctions defying intuition? *ACM SIGecom Exchanges* 14(1):41–75.
[14] Daskalakis C, Deckelbaum A, Tzamos C (2013) Mechanism design via optimal transport. *Proc. 14th ACM Conf. Electronic Commerce* (Association for Computing Machinery, New York), 269–286.
[15] Daskalakis C, Deckelbaum A, Tzamos C (2014) The complexity of optimal mechanism design. Chekuri C, ed. *Proc. 2014 Annual ACM-SIAM Sympos. Discrete Algorithms* (SIAM, Philadelphia), 1302–1318.
[16] Daskalakis C, Deckelbaum A, Tzamos C (2015) Strong duality for a multiple-good monopolist. *Proc. 16th ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 449–450.
[17] Draxler F, Veschgini K, Salmhofer M, Hamprecht F (2018) Essentially no barriers in neural network energy landscape. *Proc. 35th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 80 (PMLR, New York), 1309–1318.
[18] Du S, Lee J, Li H, Wang L, Zhai X (2019) Gradient descent finds global minima of deep neural networks. *Proc. 36th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 97 (PMLR, New York), 1675–1685.
[19] Duan Z, Tang J, Yin Y, Feng Z, Yan X, Zaheer M, Deng X (2022) A context-integrated transformer-based neural network for auction design. *Proc. 39th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 162 (PMLR, New York), 5609–5626.
[20] Dughmi S, Han L, Nisan N (2014) Sampling and representation complexity of revenue maximization. Liu TY, Qi Q, Ye Y, eds. *Web Internet Econom. WINE 2014* (Springer, Cham, Switzerland), 277–291.
[21] Dütting P, Feng Z, Narasimhan H, Parkes D, Ravindranath SS (2019) Optimal auctions through deep learning. *Proc. 36th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 97 (PMLR, New York), 1706–1715.
[22] Feng Z, Narasimhan H, Parkes DC (2018) Deep learning for revenue-optimal auctions with budgets. *Proc. 17th Internat. Conf. Autonomous Agents Multiagent Systems* (International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC), 354–362.
[23] Garipov T, Izmailov P, Podoprikhin D, Vetrov DP, Wilson AG (2018) Loss surfaces, mode connectivity, and fast ensembling of DNNs. Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R, eds. *Advances in Neural Information Processing Systems*, vol. 31 (Curran Associates, Red Hook, NY).
[24] Giannakopoulos Y, Koutsoupias E (2014) Duality and optimality of auctions for uniform distributions. *Proc. 15th ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 259–276.
[25] Giannakopoulos Y, Koutsoupias E (2018) Duality and optimality of auctions for uniform distributions. *SIAM J. Comput.* 47(1):121–165.
[26] Golowich N, Narasimhan H, Parkes DC (2018) Deep learning for multi-facility location mechanism design. Lang J, ed. *Proc. 27th Internat. Joint Conf. Artificial Intelligence* (International Joint Conferences on Artificial Intelligence), 261–267.
[27] Groves T (1973) Incentives in teams. *Econometrica* 41(4):617–631.
[28] Hertrich C, Tao Y, Végh LA (2023) Mode connectivity in auction design. Oh A, Naumann T, Globerson A, Saenko K, Hardt M, Levine S, eds. *Advances in Neural Information Processing Systems*, vol. 36 (Curran Associates, Red Hook, NY), 52957–52968.
[29] Kuditipudi R, Wang X, Lee H, Zhang Y, Li Z, Hu W, Ge R, Arora S (2019) Explaining landscape connectivity of low-cost solutions for multilayer nets. Wallach H, Larochelle H, Beygelzimer A, d'Alché-Buc F, Fox E, Garnett R, eds. *Advances in Neural Information Processing Systems*, vol. 32 (Curran Associates, Red Hook, NY).
[30] Kuo K, Ostuni A, Horishny E, Curry MJ, Dooley S, Chiang P-Y, Goldstein T, Dickerson JP (2020) ProportionNet: Balancing fairness and revenue for auction design with deep learning. Preprint, submitted October 13, https://arxiv.org/abs/2010.06398.
[31] Manelli AM, Vincent DR (2007) Multidimensional mechanism design: Revenue maximization and the multiple-good monopoly. *J. Econom. Theory* 137(1):153–185.
[32] Myerson RB (1981) Optimal auction design. *Math. Oper. Res.* 6(1):58–73.
[33] Nguyen Q (2019) On connected sublevel sets in deep learning. *Proc. 36th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research, vol. 97 (PMLR, New York), 4790–4799.
[34] Rochet J-C (1987) A necessary and sufficient condition for rationalizability in a quasi-linear context. *J. Math. Econom.* 16(2):191–200.
[35] Rochet J-C, Stole LA (2003) The economics of multidimensional screening. *Econometric Soc. Monographs* 35:150–197.
[36] Sandholm T, Likhodedov A (2015) Automated design of revenue-maximizing combinatorial auctions. *Oper. Res.* 63(5):1000–1025.
[37] Shen W, Tang P, Zuo S (2019) Automated mechanism design via neural networks. *Proc. 18th Internat. Conf. Autonomous Agents Multiagent Systems* (International Foundation for Autonomous Agents Multiagent Systems, Richland, SC), 215–223.
[38] Shevchenko A, Mondelli M (2020) Landscape connectivity and dropout stability of SGD solutions for over-parameterized neural networks. Daumé H, Singh A, eds. *Proc. 37th Internat. Conf. Machine Learn.*, Proceedings of Machine Learning Research (PMLR, New York), 8773–8784.
[39] Vickrey W (1961) Counterspeculation, auctions, and competitive sealed tenders. *J. Finance* 16(1):8–37.
[40] Wang T, Jiang Y, Parkes DC (2024) GemNet: Menu-based, strategy-proof multi-bidder auctions through deep learning. *Proc. 25th ACM Conf. Electronic Commerce* (Association for Computing Machinery, New York), 1100.
[41] Zheng S, Trott A, Srinivasa S, Parkes DC, Socher R (2022) The AI economist: Taxation policy design via two-level deep multiagent reinforcement learning. *Sci. Adv.* 8(18):1–17.