Recommender Systems meet Mechanism Design

YANG CAI*, Yale University, USA
CONSTANTINOS DASKALAKIS†, Massachusetts Institute of Technology, USA

Machine learning has developed a variety of tools for learning and representing high-dimensional distributions with structure. Recent years have also seen big advances in designing multi-item mechanisms. Akin to overfitting, however, these mechanisms can be extremely sensitive to the Bayesian prior that they target, which becomes problematic when that prior is only approximately known. At the same time, even if access to the exact Bayesian prior is given, it is known that optimal or even approximately optimal multi-item mechanisms run into sample, computational, representation and communication intractability barriers.

We consider a natural class of multi-item mechanism design problems with very large numbers of items, but where the bidders' value distributions can be well-approximated by a topic model akin to those used in recommendation systems with very large numbers of possible recommendations. We propose a mechanism design framework for this setting, building on a recent robustification framework by Brustle et al., which disentangles the statistical challenge of estimating a multi-dimensional prior from the task of designing a good mechanism for it, and robustifies the performance of the latter against the estimation error of the former. We provide an extension of this framework appropriate for our setting, which allows us to exploit the expressive power of topic models to reduce the effective dimensionality of the mechanism design problem and remove the dependence of its computational, communication and representation complexity on the number of items.

CCS Concepts: • **Theory of computation → Algorithmic mechanism design**; **Algorithmic game theory**; • **Information systems →** *Recommender systems.*

ACM Reference Format:
Yang Cai and Constantinos Daskalakis. 2022. Recommender Systems meet Mechanism Design. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC '22), July 11–15, 2022, Boulder, CO, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3490486.3538354

1 INTRODUCTION

Mechanism Design has found important applications in the design of offline and online markets. One of its main applications is the design of auctions, where a common goal is to maximize the seller's revenue from the sale of one or multiple items to one or multiple bidders. This is challenging because bidders are strategic and interact with the auction in a way that benefits themselves rather than the seller. It is well-understood that, without any information about the bidders' willingness to pay for different bundles of items, there is no meaningful way to optimize revenue. As such, a classical approach in Economics is to assume that bidders' *types* – which determine their values for different bundles and thus their willingness to pay for different bundles – are not arbitrary but

*Supported by the NSF Award CCF-1942583 (CAREER) and a Sloan Foundation Research Fellowship.
†Supported by NSF Awards CCF-1901292, DMS-2022448 and DMS-2134108, by a Simons Investigator Award, by the Simons Collaboration on the Theory of Algorithmic Fairness, by a DSTA grant, and by the DOE PhILMs project (No. DE-AC05-76RL01830).

EC '22, July 11–15, 2022, Boulder, CO, USA

ACM ISBN 978-1-4503-9150-4/22/07…$15.00
https://doi.org/10.1145/3490486.3538354

randomly drawn from a joint distribution D that is common knowledge, i.e. known to all bidders and the auctioneer. With such a Bayesian prior, the revenue of different mechanisms is compared on the basis of what revenue they achieve in expectation with respect to bidder type vectors drawn from D, and assuming that bidders play according to some (Bayesian) Nash equilibrium strategies, or some other type of (boundedly) rational behavior, e.g. no-regret learning.

Even with a Bayesian prior, however, revenue maximization is quite a challenging task. While Myerson's celebrated work showed that a relatively simple mechanism is optimal in single-item settings [38], characterizing the structure of optimal multi-item mechanisms has been notoriously difficult both analytically and computationally. Indeed, it is known that (even approximately) optimal multi-item mechanisms may require description complexity that scales exponentially in the number of items, even when there is a single buyer [3, 24, 27, 34], they might be computationally intractable, even in simple settings [10, 20, 23], and they may exhibit several counter-intuitive properties which do not arise in single-item settings; see survey [21]. Nevertheless, recent years have seen substantial progress on various fronts: analytical characterizations of optimal multi-item mechanisms [22, 24, 30, 35]; computational frameworks for computing near-optimal multi-item mechanisms [2, 9–12]; approximate multi-item revenue optimization via simple mechanisms [1, 4, 13–15, 17–19, 25, 33, 36, 40, 46]; and (approximate) multi-item revenue optimization using sample access to the type distribution [7, 8, 31, 32, 37, 44], including via the use of deep learning [28, 29, 42].

The afore-described progress on multi-item revenue optimization provides a diversity of tools that can be combined to alleviate the analytical and computational intractability of optimal mechanisms. Yet, there still remains an important challenge in applying those tools, which is that they typically require that the type distribution D is either known or can be sampled. However, this is too strong an assumption. It is common that D is *estimated* through market research or econometric analysis in related settings, involving similar items or a subset of the items. In this case, we would only hope to know some approximate distribution $\hat{D}$ that is close to D. In other settings, we may have sample access to the true distribution D but there might be errors in measuring or recording those samples. Again, we might hope to *estimate* an approximate distribution $\hat{D}$ that is close to D. Unfortunately, it is well understood that designing a mechanism for $\hat{D}$ and using it for D might be a bad idea, as optimal mechanisms tend to overfit the details of the type distribution. This has motivated a strand of recent literature to study how to robustify mechanisms to errors in the distribution [5, 6, 8].

There is, in fact, another important reason why one might want to design mechanisms for some approximate type distribution. Multi-dimensional data is complex and one would want to leverage the extensive statistical and machine learning toolkit that allows approximating such high-dimensional distributions with more structured models. Indeed, while the true type distribution D might not conform to a simple model, it might be close to a distribution $\hat{D}$ that does. We would like to leverage the simple structure in $\hat{D}$ to (i) alleviate the computational intractability of multi-item mechanisms, and (ii) reduce the amount of communication that the bidders and the auctioneer need to exchange. While the structured model $\hat{D}$ might allow (i) and (ii), we need the guarantee that the revenue of our mechanism be robust when we apply it to the true distribution D.

Motivated by the discussion above, in this work we build a multi-item mechanism design framework that combines matrix factorization models developed for recommendation systems with mechanism design, targeting two issues: (1) the intractability of Mechanism Design with respect to the number of items (arising from the exponential dependence of the number of types on the number of items if no further assumptions are placed); (2) the lack of exact access to the Bayesian priors. In particular, we assume that each bidder draws their type – specifying their values for a very large universe of N items (think all restaurants in a city or all items on Amazon) – from a distribution D_i that is close to a Matrix Factorized model $\hat{D}_i$, whose latent dimension is $k << N$. Targeting these

approximate distributions $\hat{D}_i$ allows us to reduce the effective dimensionality of bidder types to k, which has huge advantages in terms of the computational/representation/communication/sample complexity of mechanism design. We develop tools that allow us to (a) use the mechanism constructed for the approximate $\hat{D}_i$'s under the true D_i's without sacrificing much revenue; and (b) interact with the bidders who are unaware of the latent codes (they only understand their values for the N items and are oblivious to the matrix factorized model) yet exploit the factorized model for efficiently communicating with them without the impractical burden of having them communicate their N-dimensional type to the mechanism. In sum, our results are as follows:

- With a query protocol Q that learns an approximate latent representation of a bidder's type, Theorem 1 shows how to combine it with any mechanism $\widehat{M}$ that is designed only for the Matrix Factorization model to produce a mechanism that generates comparable revenue but with respect to the true distribution. The result is obtained via a refinement of the robustification result in [7], where the loss in revenue, as well as the violation in incentive compatibility now only depend on the effective dimension of the Matrix Factorization model, k, but not the total number of items, N (Lemma 2).

- We show how to obtain communication-efficient query protocols in several natural settings (Theorem 2) when the valuations are constrained-additive (Definition 5). The queries we consider ask a bidder whether they are willing to purchase an item at a given price. In the first setting, the design matrix of the Matrix Factorization model contains a diagonally dominant matrix – a generalization of the well-known separability assumption by Donoho and Stodden [26]. In two other settings, we assume that the design matrix is generated from a probablistic model and show that a simple query protocol succeeds with high probability.

- Combining Theorems 1 and 2, we show that, given any mechanism $\widehat{M}$ that is designed only for the Matrix Factorization model, we can design a mechanism that achieves comparable revenue and only requires the bidders to answer a small number of simple queries. In particular, for several natural settings, we show that the number of queries scales quasi-linearly in the effective dimension of the Matrix Factorization model and independent of the total number of items (Proposition 1).

2 PRELIMINARIES

2.1 Brief Introduction to Mechanism Design

We provide a brief introduction to mechanism design. To avoid a very long introduction, we only define the concepts in the context of multi-item auctions, which will be the focus of this paper. See Chapter 9 of [39] and the references therein for a more detailed introduction to mechanism design.

Multi-item Auctions. The seller is selling N **heterogenous items** to m **bidders**. Each bidder i is assumed to have a **private type** t_i that encodes their preference over the items and bundles of items. We assume that t_i lives in the N-dimensional Euclidean space. For each bidder, there is a publicly known valuation function $v_i(\cdot, \cdot)$, where $v_i(t_i, S) \in \mathbb{R}$ is bidder i's value for bundle $S \subseteq [N]$ when i's private type is t_i. In this paper, we consider the *Bayesian setting with private types*, that is, each bidder's type t_i is drawn *privately* and *independently* from a publicly known distribution D_i.

Mechanism. The seller designs a mechanism to sell the items to bidders. A mechanism consists of an allocation rule and a payment rule, where the allocation rule decides a way to allocate the items to the bidders, and the payment rule decides how much to charge each bidder.

Direct Mechanism: In a direct mechanism, the mechanism directly solicits types from the bidders and apply the allocation and payment rules on the reported types. More specifically, for any reported type profile $b = (b_1, \ldots, b_m)$, a direct mechanism $M := (x(\cdot), p(\cdot))$ selects $x(b) \in \{0,1\}^{m \times N}$ as the allocation and charges each bidder i payment $p_i(b)$.[1] We slightly abuse notation to allow the allocation rule to be randomized, so $x(b) \in \Delta\left(\{0,1\}^{m \times N}\right)$. We assume that bidders have quasi-linear utilities. If bidder i's private type is t_i, her utility under reported bid profile b is $u_i(t_i, M(b)) = \mathbb{E}[v_i(t_i, x(b)) - p_i(b)]$, where the expectation is over the randomness of the allocation and payment rule.

Expected Revenue: In this paper, our goal is to design mechanisms with high expected revenue. For a direct mechanism M, we use $\textsc{Rev}(M, D)$ to denote $\mathbb{E}_{t \sim D}[\sum_{i \in [m]} p_i(t)]$, where $t = (t_1, \ldots, t_m)$ is the type profile and is drawn from $D = \bigtimes_{i \in [m]} D_i$.

Incentive Compatibility and Individual Rationality. Since the bidders' types are private, unless the mechanism *incentivizes* the bidders to report truthfully, there is no reason to expect that the reported types correspond to the true types. The notion of incentive compatibility is defined to capture this.

- **ε-Bayesian Incentive Compatible (ε-BIC):** if bidders draw their types from some distribution $D = \bigtimes_{i=1}^{m} D_i$, then a direct mechanism M is *ε-BIC with respect to D* if for each bidder $i \in [m]$

$$\mathbb{E}_{t_{-i} \sim D_{-i}}[u_i(t_i, M(t_i, t_{-i}))] \geq \mathbb{E}_{t_{-i} \sim D_{-i}}[u_i(t_i, M(t_i', t_{-i}))] - \varepsilon,$$

  for all potential misreports t_i', in expectation over all other bidders bid t_{-i}. A mechanism is BIC if it is 0-BIC.

- **(ε, δ)-Bayesian Incentive Compatible ((ε, δ)-BIC):** if bidders draw their types from some distribution $D = \bigtimes_{i=1}^{m} D_i$, then a direct mechanism M is *(ε, δ)-BIC with respect to D* if for each bidder $i \in [m]$:

$$\Pr_{t_i \sim D_i}\left[\mathbb{E}_{t_{-i} \sim D_{-i}}[u_i(t_i, M(t_i, t_{-i}))] \geq \mathbb{E}_{t_{-i} \sim D_{-i}}[u_i(t_i, M(t_i', t_{-i}))] - \varepsilon\right] \geq 1 - \delta.$$

- **Individually Rational (IR):** A direct mechanism M is *IR* if for all type profiles $t = (t_1, \ldots, t_m)$,

$$u_i(t_i, M(t_i, t_{-i})) \geq 0$$

  for all bidders $i \in [m]$.

Indirect Mechanism: An *indirect mechanism* does not directly solicit the bidders' types. After interacting with the bidders, the mechanism selects an allocation and payments. The notions of ε-Bayesian Incentive Compatibility and Individual Rationality can be extended to indirect mechanisms using the solution concept of ε-Bayes Nash equilibrium. The notion of (ε, δ)-Bayesian Incentive Compatibility can be extended to indirect mechanisms using the new solution concept, which we call *(ε, δ)-weak approximate Bayes Nash equilibrium.* In an incomplete information game, a strategy profile is an (ε, δ)-weak approximate Bayes Nash equilibrium if for every bidder, with probability no more than δ (over the randomness of their own type), unilateral deviation from the Bayesian Nash strategy can increase the deviating bidder's expected utility (with respect to the randomness of the other bidders' types and assuming those follow their Bayesian Nash equilibrium strategies) by more than ε.

[1] Note that $p(b) = (p_1(b), \ldots, p_m(b))$.

Remark 1. *For a (ε, δ)-weak approximate Bayes Nash equilibrium, its expected revenue computation is made in this paper using the convention that all bidders follow their (ε, δ)-weak approximate Bayes Nash equilibrium strategies. At a cost of an additive $m^2\delta H$ loss in revenue (where H is the highest possible value of any bidder), we can assume that only the $(1-\delta)$-fraction of types of each bidder who have no more than ε incentive to deviate from the weak approximate Bayes Nash equilibrium strategies follow these strategies while the remaining δ fraction use arbitrary strategies. Similarly, we can interpret the (ε, δ)-weak approximate Bayes Nash equilibrium definition as requiring that at least $(1-\delta)$-fraction of the types of each bidder have at most $O(\varepsilon + m\delta H)$ incentive to deviate from the Bayes Nash strategies assuming that for every other bidder at most δ fraction of their types deviate from their Bayes Nash strategies.*

2.2 Further Preliminaries

Definition 1. *Let (U, d) be a metric space and $\mathcal{B}$ be a σ-algebra on U. For all $A \in \mathcal{B}$, let $A^\varepsilon = \{x : \exists y \in A \ s.t \ \ d(x,y) < \varepsilon\}$. Two probability measure P and Q on $\mathcal{B}$ have* Prokhorov distance

$$\inf \{\varepsilon > 0 : P(A) \le Q(A^\varepsilon) + \varepsilon \text{ and } \ Q(A) \le P(A^\varepsilon) + \varepsilon, \ \forall A \in \mathcal{B}\}.$$

We consider distributions supported on some Euclidean Space, and we choose d to be the ℓ_∞-distance. We denote the ℓ_∞-Prokhorov distance between distributions F, $\widehat{F}$ by $d_P(F, \widehat{F})$.

We will also make use of the following characterization of the Prokhorov metric by [43].

Lemma 1 (Characterization of the Prokhorov Metric [43]). *Let F and $\widehat{F}$ be two distributions supported on $\mathbb{R}^n$. $d_P(F, \widehat{F}) \le \varepsilon$ if and only if there exists a coupling γ of F and $\widehat{F}$, such that $\Pr_{(x,y)\sim\gamma}\left[\|x-y\|_\infty > \varepsilon\right] \le \varepsilon$.*

Definition 2 (Influence Matrix and Weak Dependence). *For any d-dimensional random vector $X = (X_1, \ldots, X_d)$, we define the influence of variable j on variable i as*

$$\alpha_{i,j} := \sup_{\substack{x_{-i-j} \\ x_j \neq x_j'}} d_{TV}\left(F_{X_i|X_j=x_j, X_{-i-j}=x_{-i-j}}, F_{X_i|X_j=x_j', X_{-i-j}=x_{-i-j}}\right),$$

where $F_{X_i|X_{-i}=x_{-i}}$ denotes the conditional distribution of X_i given $X_{-i} = x_{-i}$, and $d_{TV}(D, D')$ denotes the total variational distance between distribution D and D'. Also, let $\alpha_{i,i} := 0$ for each i, and we use $\textsc{Inf}(X)$ to denote the $d \times d$ matrix $(\alpha_{i,j})_{i\in[d], j\in[d]}$. In this paper, we consider the coordinates of X to be ***weakly dependent*** *if $\|\textsc{Inf}(X)\|_2 < 1$.*

3 OUR MODEL AND MAIN RESULTS

Setting and Goal: We consider a classical mechanism design problem, wherein a seller is selling N items to m buyers, where buyer i's type t_i is drawn from a distribution D_i over $\mathbb{R}^N$ independently. The goal is to design a mechanism that maximizes the seller's revenue. In this paper, we operate in a setting where D_i is unknown, but we are given access to the following components: (I) For each bidder i, we are given a machine learning model $\widehat{D}_i$ — of the matrix factorization type as described below, which approximates D_i. (II) We are given a good mechanism $\widehat{M}$ for the approximate type distributions; in its design this mechanism can exploit the low effective dimensionality, k, of types in the approximate model. Our goal is (III) to use (I) and (II) to obtain a good mechanism for the true type distributions.

(I) The Machine Learning Component: We assume that each bidder's type distribution D_i can be well-approximated by a known *Matrix Factorization (MF)* model $\widehat{D}_i$. In particular:

- We use $A \in \mathbb{R}^{N \times k}$ to denote the design matrix of the model, where each column can be viewed as the type (over N items) of an "archetype." As described in the following two bullets, types are sampled by each $\widehat{D}_i$ as linear combinations over archetypes.

- We use $\widehat{D}_{z,i}$ to denote a distribution over $[0,1]^k$. The subscript z is not a parameter of the distribution — it serves to remind us that this distribution samples in the latent space $[0,1]^k$ and distinguish it from the distribution $\widehat{D}_i$ defined next.

- If F is a distribution over $\mathbb{R}^k$, we use $A \circ F$ to denote the distribution of the random variable Az, where $z \sim F$. With this notation, we use $\widehat{D}_i$ to denote $A \circ \widehat{D}_{z,i}$.

- We assume that, for each bidder, the matrix factorization model is not far away from the true type distribution, that is, for some $\varepsilon_1 > 0$ we have that $d_P(D_i, \widehat{D}_i) \leq \varepsilon_1$ for all $i \in [m]$.

Remark 2. *In the above description we assumed that all $\widehat{D}_i$'s share the same design matrix A. This is done to avoid overloading notation but all our results would hold if each $\widehat{D}_i$ had its own design matrix A_i.*

(II) The Mechanism Design Component: We assume that we are given a direct mechanism $\widehat{M}$ for types drawn from the Machine Learning model. In particular, we assume that this mechanism makes use of the effective dimension k of the Machine Learning model, accepting "latent types" (of dimension k) as input from the bidders. Specifically:

- Recall that, for each bidder i, their valuation function $v_i(\cdot,\cdot) : \mathbb{R}^N \times 2^{[N]} \to \mathbb{R}$ is common knowledge. (Recall that v_i takes as input the bidder's type and a subset of items so how the bidder values different subsets of items depends on their private type.)

- The designer is given A and $\widehat{D}_{z,i}$ for each bidder i, and treats bidder i's type as drawn from $\widehat{D}_{z,i}$, i.e. in the latent space $[0,1]^k$. With respect to such "latent types," there is an induced valuation function. In particular, for each bidder i, we use $v_i^A : \mathbb{R}^k \times 2^{[N]} \to \mathbb{R}$ to denote the valuation function defined as follows $v_i^A(z_i, S) := v_i(Az_i, S)$, where $z_i \in \mathbb{R}^k$.

- With the above as setup, we assume that the designer designs a mechanism $\widehat{M}$ that is BIC and IR w.r.t. $\widehat{D}_z = \bigtimes_{i=1}^m \widehat{D}_{z,i}$ and valuation functions $\{v_i^A(\cdot,\cdot)\}_{i\in[m]}$.

(III) The New Component: We consider the regime where $N \gg k$, and our goal is to combine the Machine Learning component with the Mechanism Design component to produce a mechanism which generates revenue comparable to $\textsc{Rev}(\widehat{M}, \widehat{D}_z)$ when used for bidders whose types are drawn from $D = \bigtimes_{i=1}^m D_i$. There are two challenges: (i) $\widehat{M}$ takes as input the latent representation of a bidder's type under $\widehat{D}_z$, however under D a bidder is simply ignorant about any latent representation of their type so they cannot be asked about it; (ii) $\widehat{M}$'s revenue is evaluated with respect to $\widehat{D}_z$ and valuation functions $\{v_i^A(\cdot,\cdot)\}_{i\in[m]}$ and our goal is to obtain a mechanism whose revenue is similar under D and valuation functions $\{v_i(\cdot,\cdot)\}_{i\in[m]}$. We show how to use a communication efficient query protocol together with a robustification procedure to combine the Machine Learning and Mechanism Design components.

To state our results, we first need to formally define query protocols and some of their properties.

Definition 3 ((ε,δ)-query protocol). *Let $\mathcal{Q}$ be a **query protocol**, i.e., some communication protocol that exchanges messages with a bidder over possibly several rounds and outputs a vector in*

$\mathbb{R}^k$. *We say that a bidder interacts with the query protocol truthfully, if whenever the protocol asks the bidder to evaluate some function on their type the bidder evaluates the function and returns the result truthfully. We use* $Q(t) \in \mathbb{R}^k$ *to denote the output of* Q *when interacting with a truthful bidder whose type is* $t \in \mathbb{R}^N$. Q *is called a* (ε, δ)*-query protocol, if for any* $t \in \mathbb{R}^N$ *and* $z \in \mathbb{R}^k$ *satisfying* $\|t - Az\|_\infty \le \varepsilon$, *we have that* $\|z - Q(t)\|_\infty \le \delta$.

We also need the notion of Lipschitz valuations to formally state our result.

Definition 4 (Lipschitz Valuations). $v(\cdot,\cdot) : \mathbb{R}^N \times 2^{[N]} \to \mathbb{R}$ *is a* $\mathcal{L}$*-Lipschitz valuation, if for any two types* $t, t' \in \mathbb{R}^N$ *and any bundle* $S \subseteq [N]$, $|v(t,S) - v(t',S)| \le \mathcal{L}\,\|t - t'\|_\infty$.

This includes familiar settings, for example if the bidder is c-demand, the Lipschitz constant $\mathcal{L} = c$.[2]

We are now ready to state our first main result.

Theorem 1. *Let* $D = \bigtimes_{i=1}^m D_i$ *be the bidders' type distributions and* $v_i : \mathbb{R}^N \times 2^{[N]} \to \mathbb{R}$ *be a* $\mathcal{L}$*-Lipschitz valuation for each bidder* $i \in [m]$. *Also, let* $A \in \mathbb{R}^{N\times k}$ *be a design matrix and* $\widehat{D}_{z,i}$ *be a distribution over* $\mathbb{R}^k$ *for each* $i \in [m]$.

Suppose we are given query access to a mechanism $\widehat{M}$ *that is BIC and IR w.r.t.* $\widehat{D}_z = \bigtimes_{i=1}^m \widehat{D}_{z,i}$ *and valuations* $\{v_i^A\}_{i\in[m]}$ *(as defined in the second bullet of the Mechanism Design component above), and there exists* $\varepsilon_1 > 0$ *such that* $d_P(D_i, A \circ \widehat{D}_{z,i}) \le \varepsilon_1$ *for all* $i \in [m]$. *Given any* $(\varepsilon_1, \varepsilon)$*-query protocol with* $\varepsilon \ge \varepsilon_1$, *we can construct mechanism* M *using only query access to* $\widehat{M}$ *and obliviously with respect to* D, *such that for any possible* D *that satisfies the above conditions of Prokhorov distance closeness the following hold:*

(1) M *only interacts with every bidder using* Q *once;*

(2) M *is* (κ, ε_1)*-BIC w.r.t.* D *and IR, where* $\kappa = O\left(\mathcal{L}\varepsilon_1 + \|A\|_\infty \mathcal{L} m\varepsilon + \|A\|_\infty \mathcal{L}\sqrt{m\varepsilon}\right)$;

(3) The expected revenue of M *is at least* $\textsc{Rev}(\widehat{M}, \widehat{D}_z) - O(m\kappa)$.

Remark 3. *The mechanism* M *will be an indirect mechanism. We are slightly imprecise here to call the mechanism* (κ, ε_1)*-BIC. Formally what we mean is that interacting with* Q truthfully *is a* (κ, ε_1)*-weak approximate Bayes Nash equilibrium. We compute the expected revenue assuming all bidders interacting with* Q truthfully. *As we discussed in Remark 1, with an additional additive* $\|A\|_\infty \mathcal{L} m^2 \varepsilon_1$ *loss in revenue, we can assume that only the* $(1-\delta)$*-fraction of types of each bidder who have no more than* ε *incentive to deviate from the Bayes Nash strategies interact with* Q *truthfully while the remaining* δ *fraction uses arbitrary strategies.*

Why isn't [7] sufficient? One may be tempted to prove Theorem 1 using [7]. However, there are two subtle issues with this approach: (i) The violation of the incentive compatibility constraints and the revenue loss of the robustification process in [7] depend linearly in N, rather than in $\|A\|_\infty$ as in Theorem 1. Note that $\|A\|_\infty = \max_{i\in[N]} \sum_{j=1}^k |A_{ij}|$, which only depends on k and the largest value an archetype can have for a single item and thus could be significantly smaller than N. (ii) The robustification process involves sampling from the conditional distribution of $A \circ \widehat{D}_{z,i}$ on an N-dimensional cube, which is equivalent to sampling from the conditional distribution of $\widehat{D}_{z,i}$ on a set S whose image after the linear transformation A is the N-dimensional cube. However, S may be difficult to sample from if A is not a well-conditioned.

[2] A bidder is c-demand if for any set S of items, the bidder picks their favorite bundle with size no more than c in S evaluating the value of each such bundle additively, with values as determined by the bidder's type t. Formally, $v(t, S) = \max_{B\subseteq S, |B|\le c} \sum_{j\in B} t_j$.

In the following lemma, we refine the robustification result in [7] (Theorem 3 in that paper) and show that given an approximate distribution $\widehat{F}$ in the latent space and a BIC and IR mechanism $\widehat{M}$ w.r.t. $\widehat{F}$, we can *robustify* $\widehat{M}$ with *negligible revenue loss* so that it is an approximately BIC and exactly IR mechanism w.r.t. F for any distribution F that is within the ε-Prokhorov ball around $\widehat{F}$. Importantly, we exploit the effective dimension of the matrix factorization model to replace the dependence on N with $\|A\|_\infty$ in both the violation of the incentive compatibility constraints and the revenue loss. Additionally, we only need to be able to sample from the conditional distribution of $\widehat{D}_{z,i}$ on a k-dimensional cube. We postpone the proof of Lemma 2 to the Appendix A.

Lemma 2. *Let $A \in \mathbb{R}^{N\times k}$ be the design matrix. Suppose we are given a collection of distributions over latent types $\{\widehat{F}_{z,i}\}_{i\in[m]}$, where the support of each $\widehat{F}_{z,i}$ lies in $[0,1]^k$, and a BIC and IR mechanism $\widehat{M}$ w.r.t. $\widehat{F} = \bigtimes_{i=1}^m \widehat{F}_{z,i}$ and valuations $\{v_i^A\}_{i\in[m]}$, where each v_i is an $\mathcal{L}$-Lipschitz valuation. Let $F = \bigtimes_{i=1}^m F_{z,i}$ be any distribution such that $d_P(F_{z,i}, \widehat{F}_{z,i}) \le \varepsilon$ for all $i \in [m]$. Given access to a **sampling algorithm** $\mathcal{S}_i$ for each $i \in [m]$, where $\mathcal{S}_i(x,\delta)$ draws a sample from the conditional distribution of $\widehat{F}_{z,i}$ on the k-dimensional cube $\bigtimes_{j\in[k]}[x_j, x_j+\delta)$, we can construct a randomized mechanism $\widetilde{M}$ using only query access to $\widehat{M}$ and obliviously with respect to F, such that for any F satisfying the above conditions of Prokhorov distance closeness the following hold:*

(1) M is κ-BIC and IR w.r.t. F and valuations $\{v_i^A\}_{i\in[m]}$, where $\kappa = O\left(\|A\|_\infty \mathcal{L} m\varepsilon + \|A\|_\infty \mathcal{L}\left(\delta + \frac{m\varepsilon}{\delta}\right)\right)$;

(2) The expected revenue of $\widetilde{M}$ is $\textsc{Rev}\left(\widetilde{M}, F\right) \ge \textsc{Rev}(\widehat{M}, \widehat{F}) - O(m\kappa)$.

Equipped with Lemma 2, we proceed to prove Theorem 1.

Proof of Theorem 1: Consider the following mechanism:

ALGORITHM 1: Query-based Indirect Mechanism M

1: Construct mechanism $\widetilde{M}$ using Lemma 2 by choosing $\widehat{F}_{z,i}$ to be $\widehat{D}_{z,i}$ for each $i \in [m]$ and δ to be $\sqrt{m\varepsilon}$.
2: Query each agent i using Q. Let $Q(b_i)$ be the output after interacting with bidder i. (For any possible output produced by Q, there exists a type $b \in \mathbb{R}^N$, so this is w.l.o.g..)
3: Execute mechanism $\widehat{M}$ on bid profile $(Q(b_1), \ldots, Q(b_m))$.

Let t_i be bidder i's type and z_i be a random variable distributed according to $\widehat{D}_{z,i}$. Since $d_P(D_i, \widehat{D}_i) \le \varepsilon_1$, Lemma 1 guarantees a coupling between t_i and Az_i such that their ℓ_∞ distance is more than ε_1 with probability no more than ε_1. As Q is a $(\varepsilon_1, \varepsilon)$-query protocol, when t_i and Az_i are not ε_1 away, we have $\|Q(t_i) - z_i\|_\infty \le \varepsilon$. Hence, there exists a coupling between $Q(t_i)$ and z_i so that their ℓ_∞ distance is more than ε with probability no more than ε (recall $\varepsilon_1 \le \varepsilon$). If we choose $F_{z,i}$ to be the distribution of $Q(t_i)$, $\widehat{F}_{z,i}$ to be $\widehat{D}_{z,i}$, and δ to be $\sqrt{m\varepsilon}$, Lemma 2 states that $\widetilde{M}$ is a $O\left(\|A\|_\infty \mathcal{L} m\varepsilon + \|A\|_\infty \mathcal{L}\sqrt{m\varepsilon}\right)$-BIC mechanism if bidder i has valuation $v_i^A(\cdot)$ and type $Q(t_i)$. Consider two cases: (a) When $\|t_i - Az_i\|_\infty \le \varepsilon_1$, then $\|t_i - AQ(t_i)\|_\infty \le \varepsilon_1 + \|A\|_\infty \varepsilon$. Since $v_i(\cdot)$ is $\mathcal{L}$-Lipschitz, deviating from interacting with Q truthfully can increase the expected utility by at most $O\left(\mathcal{L}\varepsilon_1 + \|A\|_\infty \mathcal{L} m\varepsilon + \|A\|_\infty \mathcal{L}\sqrt{m\varepsilon}\right)$. (b) When $\|t_i - Az_i\|_\infty > \varepsilon_1$, the bidder may substantially improve their expected utility by deviating. Luckily, such case happens with probability no more than ε_1. □

In Theorem 2, we show how to obtain (ε, δ)-queries under various settings. We further assume that the bidders' valuations are all constrained-additive.

Definition 5 (Constrained-Additive valuation). *A valuation function $v : \mathbb{R}^N \times 2^{[N]} \to \mathbb{R}$ is constrained additive if $v(t, S) = \max_{T \in \mathcal{I} \cap 2^S} \sum_{j \in T}(\mu_j + t_j)$, where $\mathcal{I}$ is a downward-closed set system, and $\mu = (\mu_1, \ldots, \mu_N)$ is a fixed vector.*[3] *For example, unit-demand valuation is when $\mathcal{I}$ includes all subsets with size no more than 1. If all elements of $\mathcal{I}$ have size no more than $\mathcal{L}$, then v is a $\mathcal{L}$-Lipschitz valuation.*

Theorem 2. *Let all bidders' valuations be constrained-additive. We consider queries of the form: $e_j^T t \overset{?}{\geq} p$, where e_j is the j-th standard unit vector in $\mathbb{R}^N$. The query simply asks whether the bidder is willing to pay at least $p + \mu_j$ for winning item j. The bidder provides a* Yes/No *answer. We obtain communicationally efficient protocols in the following settings:*

- ***Deterministic Structure:*** *If A^T can be expressed as $[C^T H^T]\Pi_N$, where $\Pi_N \in \mathbb{R}^N$ is a permutation matrix, H is an arbitrary $(N-k) \times k$ matrix, and $C \in \mathbb{R}^{k \times k}$ is diagonally dominant both by rows and by columns. This is a relaxation of the well-known* ***separability assumption*** *by Donoho and Stodden [26], that is, A^T can be expressed as $[I_k H^T]\Pi_N$, where I_k is the k-dimensional identity matrix. Let $\alpha = \min_{i \in [k]} \left(|C_{ii}| - \sum_{j \neq i} |C_{ij}|\right)$ and $\beta = \min_{j \in [k]} \left(|C_{jj}| - \sum_{i \neq j} |C_{ij}|\right)$. We have a $\left(\varepsilon, \frac{4 \cdot \max_{j \in [k]} C_{jj}}{\alpha\beta} \cdot \varepsilon\right)$-query protocol using $O\left(k \cdot \log\left(\frac{\|A\|_\infty}{\varepsilon}\right)\right)$ queries for any $\varepsilon > 0$.*

- ***Ex-ante Analysis:*** *If A is generated from a distribution, where each archetype is an independent copy of a N-dimensional random vector θ.*
  - ***Multivariate Gaussian Distributions:*** *θ is distributed according to a multivariate Gaussian distribution $\mathcal{N}(0, \Sigma)$. If there exists a subset $S \subseteq [N]$ such that $\frac{\mathrm{Tr}(\Sigma_S)}{\rho(\Sigma_S)} > 64k$, where $\Sigma_S = \mathbb{E}[\theta_S \theta_S^T]$ is the covariance matrix for items in S and $\rho(\Sigma_S)$ is the largest eigenvalue of Σ_S,*[4] *then with probability at least $1 - 2\exp\left(-\frac{\mathrm{Tr}(\Sigma_S)}{16 \cdot \rho(\Sigma_S)}\right)$, we have a $\left(\varepsilon, \frac{64\sqrt{|S|k}}{\sqrt{\mathrm{Tr}(\Sigma_S)}} \cdot \varepsilon\right)$-query protocol using $O\left(|S| \cdot \log\left(\frac{\|A\|_\infty}{\varepsilon}\right)\right)$ queries for any $\varepsilon > 0$. Note that when the entries of θ are i.i.d., any S with size at least $64k$ satisfies the condition.*
  - ***Bounded Distributions with Weak Dependence:*** *Let θ_i be supported on $[-c, c]$ and has mean 0 for each $i \in [N]$. If there exists a subset $S \subseteq [N]$ such that $\|\textsc{Inf}(\theta_S)\|_2 < 1$, and $\sum_{i \in S} v_i^2 > \frac{16c^2 k\sqrt{|S|}}{1 - \|\textsc{Inf}(\theta_S)\|_2}$, where $v_i^2 := \mathrm{Var}[\theta_i]$, then with probability at least $1 - 2\exp\left(-\frac{(1 - \|\textsc{Inf}(\theta_S)\|_2) \cdot (\sum_{i \in S} v_i^2)^2}{64c^4 k|S|}\right)$, we have a $\left(\varepsilon, \frac{64\sqrt{|S|k}}{\sqrt{\sum_{i \in S} v_i^2}} \cdot \varepsilon\right)$-query protocol using $O\left(|S| \cdot \log\left(\frac{\|A\|_\infty}{\varepsilon}\right)\right)$ queries for any $\varepsilon > 0$. Note that when the entries of θ are independent, $\|\textsc{Inf}(\theta_S)\|_2 = 0$ for any set S. If each θ_i has variance $\Omega(c^2)$, then any set with size at least αk^2 suffices for some absolute constant α.*

Remark 4. *In the ex-ante analysis, the success probabilities depend on the parameters of the distributions, but note that they are both at least $1 - 2\exp(-4k)$.*

Before we prove Theorem 2, we combine it with Theorem 1 to derive results for a few concrete settings.

Proposition 1. *Under the same setting as in Theorem 1 with the extra assumption that every valuation v_i is constrained-additive, we can construct mechanism M using only query access to the*

[3] One can interpret μ as the common based values for the items that are shared among all types.

[4] θ_S is the $|S|$-dimensional vector that contains all θ_i with $i \in S$.

given mechanism $\widehat{M}$ and oblivious to the true type distribution D, such that for any possible D, M is (η, ε_1)-BIC and IR, where $\eta = O\left(\mathcal{L}\varepsilon_1 + \|A\|_\infty \mathcal{L}mf(\varepsilon_1) + \|A\|_\infty \mathcal{L}\sqrt{mf(\varepsilon_1)}\right)$, and has revenue at least $\textsc{Rev}(\widehat{M}, \widehat{D}_z) - O\left(\|A\|_\infty \mathcal{L}m^2 f(\varepsilon_1) + \|A\|_\infty \mathcal{L}m^{3/2} f(\varepsilon_1)^{1/2}\right)$. Recall that ε_1 satisfies $d_P(D_i, A \circ \widehat{D}_{z,i}) \le \varepsilon_1$ for all $i \in [m]$. We compute the function $f(\cdot)$ and the number of queries for the following three concrete settings (one for each of the three assumptions in Theorem 2).

(1) ***Deterministic Structure: Separability.*** *If the design matrix A satisfies the **separability assumption** by Donoho and Stodden [26], that is, A^T can be expressed as $[I_k H^T]\Pi_N$, where $\Pi_N \in \mathbb{R}^N$ is a permutation matrix, $f(\varepsilon_1) = 4\varepsilon_1$ for all $\varepsilon > 0$. The number of queries each bidder needs to answer is $O\left(k \cdot \log\left(\frac{\|A\|_\infty}{\varepsilon_1}\right)\right)$.*

(2) ***Multivariate Gaussian Distributions: Well-Conditioned Covariance Matrix.*** *Let A be generated from a distribution, where each archetype is an independent draw from a N-dimensional normal distribution $\mathcal{N}(0, \Sigma)$. Let $\kappa(\Sigma)$ be the condition number of Σ.[5] For any set S with size $64\kappa(\Sigma)k$, if we query each bidder about items in S, with probability at least $1 - 2\exp(-4k)$, $f(\varepsilon_1) = O\left(\frac{k\sqrt{\kappa(\Sigma)}}{\sqrt{\mathrm{Tr}(\Sigma_S)}} \cdot \varepsilon_1\right)$, and each bidder needs to answer $O\left(\kappa(\Sigma)k \cdot \log\left(\frac{\|A\|_\infty}{\varepsilon_1}\right)\right)$ queries.*

(3) ***Weak Dependence: Sufficient Variance per Item.*** *Let A be generated from a distribution, where each archetype is an independent copy of an N-dimensional random vector θ. Assuming (i) $\|\textsc{Inf}(\theta)\|_2 < 1$, (ii) θ_i lies in $[-c, c]$, and (iii) $\mathrm{Var}[\theta_i] \ge a^2$ for each $i \in [N]$, then for any set S with size $\frac{256c^4k^2}{a^4(1-\|INF(\theta)\|_2)^2}$, if we query each bidder about items in S, with probability at least $1 - 2\exp(-4k)$, $f(\varepsilon_1) = O\left(\frac{\sqrt{k}}{a} \cdot \varepsilon_1\right)$ and each bidder needs to answer $O\left(\frac{c^4k^2}{a^4(1-\|INF(\theta)\|_2)^2} \cdot \log\left(\frac{\|A\|_\infty}{\varepsilon_1}\right)\right)$ queries.*[6]

Proof. The results in the first and last setting follows directly from Theorem 2. For the second setting, notice that by the eigenvalue interlacing theorem, $\kappa(\Sigma_S) \le \kappa(\Sigma)$, as Σ_S is a principal submatrix of Σ. Therefore, $\frac{\mathrm{Tr}(\Sigma_S)}{\rho(\Sigma_S)} \ge \frac{|S|}{\kappa(\Sigma_S)} \ge 64k$. Now, the result follows from Theorem 2. □

Proof of Theorem 2: Instead of directly studying the query complexity under our query model. We first consider the query complexity under a seemingly stronger query model, where we directly query the bidder about their value of $e_j^T t$, and their answer will be within $e_j^T t \pm \eta$ for some $\eta > 0$. We refer to this type of queries as noisy value queries. Since for each item j, $|e_j^T Az| \le \|A\|_\infty$ for all $z \in [0,1]^k$ and we only care about types in $\mathbb{R}^N$ that are close to some Az, we can use our queries to perform binary search on p to simulate noisy value queries. In particular, we only need $\log\|A\|_\infty + \log 1/\eta + \log 1/\varepsilon$ many queries to simulate one noisy value queries. From now on, the plan is to first investigate the query complexity for noisy value queries, then convert the result to query complexity in the original model.

We first fix the notation. Let ℓ be the number of noisy value queries, and $Q \in \mathbb{R}^{\ell \times N}$ be the query matrix, where, each row of Q is a standard unit vector. We use $\hat{y} \in \mathbb{R}^\ell$ to denote the bidder's answers to the queries and $y \in \mathbb{R}^\ell$ to true answers to the queries. Note that $\|\hat{y} - y\|_\infty \le \eta$. Given $\hat{y}$, we solve the following least squares problem: $\min_{z \in \mathbb{R}^k} \|QAz - \hat{y}\|_2^2$.

[5] Σ is well-conditioned if $\kappa(\Sigma)$ is small. When $\Sigma = I_N$, $\kappa(\Sigma) = 1$.

[6] Clearly, we can weaken condition (i),(ii) and (iii). The result still holds if we can find a set S, so that for vector θ_S, condition (i), (ii), and (iii) hold, and $|S|$ is at least $\frac{256c^4k^2}{a^4(1-\|INF(\theta_S)\|_2)^2}$.

The problem has a closed form solution: $\hat{z} = \left(A^T Q^T Q A\right)^{-1} A^T Q^T \hat{y}$. Let $B := QA$, and $z(t) \in \mathbb{R}^k$ be a vector that satisfies $\|t - Az(t)\|_\infty \le \varepsilon$. We are interested in upper bounding $\|\hat{z} - z(t)\|_\infty$. Note that

$$\begin{aligned}\hat{z} - z(t) =& (B^T B)^{-1} B^T (\hat{y} - Bz(t)) \\ =& (B^T B)^{-1} B^T \left((\hat{y} - y) + (y - Bz(t))\right) \\ =& (B^T B)^{-1} B^T (\hat{y} - y) + (B^T B)^{-1} B^T Q(t - Az(t))\end{aligned}$$

Since the rows of Q are all standard unit vectors, $\|Q\|_\infty = 1$.

$$\begin{aligned}\|\hat{z} - z(t)\|_\infty &\le \left\|(B^T B)^{-1} B^T (\hat{y} - y)\right\|_\infty + \left\|(B^T B)^{-1} B^T Q(t - Az(t))\right\|_\infty \\ &\le \left\|(B^T B)^{-1}\right\|_\infty \left\|B^T\right\|_\infty \left(\eta + \|Q(t - Az(t))\|_\infty\right) \\ &\le \left\|(B^T B)^{-1}\right\|_\infty \left\|B^T\right\|_\infty (\eta + \varepsilon).\end{aligned}$$

Next, we bound $\left\|(B^T B)^{-1}\right\|_\infty \left\|B^T\right\|_\infty$ under the different assumptions.

Deterministic Structure: We choose $\ell = k$ and Q so that $QA = B = C$. Since C is diagonally dominant, C is non-singular, and $(C^T C)^{-1} = C^{-1}(C^T)^{-1}$.

Lemma 3 (Adapted from Theorem 1 and Corollary 1 of [45]). *If a matrix $U \in \mathbb{R}^{n \times n}$ is diagonally dominant both by rows and by columns, and $\alpha = \min_{i \in [n]} \left(|U_{ii}| - \sum_{j \ne i} |U_{ij}|\right)$ and $\beta = \min_{j \in [n]} \left(|U_{jj}| - \sum_{i \ne j} |U_{ij}|\right)$, then $\left\|U^{-1}\right\|_\infty \le 1/\alpha$ and $\left\|(U^T)^{-1}\right\|_\infty \le 1/\beta$.*

By Lemma 3, $\left\|(C^T C)^{-1}\right\|_\infty \left\|C^T\right\|_\infty \le \frac{\left\|C^T\right\|_\infty}{\alpha\beta}$. Note that $\left\|C^T\right\|_\infty = \max_{j \in [k]} \sum_{i \in [k]} |C_{ij}| \le 2 \max_{j \in [k]} C_{jj}$. The last inequality is because C is diagonally dominant by columns. To sum up, if we choose Q so that $QA = C$,

$$\|\hat{z} - z(t)\|_\infty \le \frac{(\varepsilon + \eta) \cdot \left\|C^T\right\|_\infty}{\alpha\beta} \le \frac{2(\varepsilon + \eta) \cdot \max_{j \in [k]} C_{jj}}{\alpha\beta}.$$

Ex-ante Analysis: Since $\left\|(B^T B)^{-1}\right\|_\infty \le \sqrt{k} \left\|(B^T B)^{-1}\right\|_2$ and $\left\|B^T\right\|_\infty \le \sqrt{\ell} \|B\|_2$,

$$\|\hat{z} - z(t)\|_\infty \le \frac{\sqrt{\ell k} \cdot \sigma_{max}(B)}{\sigma_{min}(B)^2} \cdot (\eta + \varepsilon),$$

where $\sigma_{max}(B)$ (or $\sigma_{min}(B)$) is B's largest (or smallest) singular value.

Multivariate Gaussian distribution: When θ is distributed according to a multivariate Gaussian distribution, we choose $\ell = |S|$ and Q so that each row corresponding to an e_j with $j \in S$. Now, B is a $\ell \times k$ random matrix where each column is an independent copy of θ_S. We use Lemma 4 to bound B's largest singular value $\sigma_{max}(B)$ and smallest singular value $\sigma_{min}(B)$. The proof of Lemma 4 is postponed to Section 4.1.

Lemma 4. *[Concentration of Singular Values under multivariate Gaussian distributions] Let $U = [X^{(1)}, \ldots, X^{(n)}]$ be a $m \times n$ random matrix, where each column of U is an independent copy of a m-dimensional random vector X distributed according to a multivariate Gaussian distribution $\mathcal{N}(0, \Lambda^T D \Lambda)$. In particular, $\Lambda \in \mathbb{R}^{m \times m}$ is an orthonormal matrix, and $D \in \mathbb{R}^{m \times m}$ is a diagonal matrix. We have $\sigma_{max}(U) \le 2\sqrt{\mathrm{Tr}(D)}$ and $\sigma_{min}(U) \ge \frac{\sqrt{\mathrm{Tr}(D)}}{4}$, with probability at least $1 - 2\exp\left(-\frac{\mathrm{Tr}(D)}{8 \cdot d_{max}} + 4n\right)$, where d_{max} is the largest entry in D.*

Since $\frac{\mathrm{Tr}(\Sigma_S)}{\rho(\Sigma_S)} > 64k$, by Lemma 4, $\sigma_{max}(B) \le 2\sqrt{\mathrm{Tr}(\Sigma_S)}$ and $\sigma_{min}(B) \ge \sqrt{\mathrm{Tr}(\Sigma_S)}/4$ with probability at least $1-2\exp\left(-\frac{\mathrm{Tr}(\Sigma_S)}{16\cdot\rho(\Sigma_S)}\right) \ge 1-2\exp(-4k)$. Hence, $\|\hat{z}-z(t)\|_\infty \le \frac{32\sqrt{|S|k}}{\sqrt{\mathrm{Tr}(\Sigma_S)}}\cdot(\eta+\varepsilon)$ with probability at least $1-2\exp\left(-\frac{\mathrm{Tr}(\Sigma_S)}{16\cdot\rho(\Sigma_S)}\right)$.

Weakly Dependent Distributions: When the coordinates of θ_S are weakly dependent, i.e., $\|\textsc{Inf}(\theta_S)\|_2 < 1$, we choose $\ell = |S|$ and Q so that each row corresponding to an e_j with $j \in S$. Now, B is a $\ell \times k$ random matrix where each column is an independent copy of θ_S. We use Lemma 5 to bound B's largest singular value $\sigma_{max}(B)$ and smallest singular value $\sigma_{min}(B)$. The proof of Lemma 5 is postponed to Section 4.2.

Lemma 5. *[Concentration of Singular Values under Weak Dependence] Let $U = [X^{(1)},\ldots,X^{(n)}]$ be a $m \times n$ random matrix, where each column of U is an independent copy of a m-dimensional random vector X. We assume that the coordinates of X are weakly dependent, i.e., $\|\textsc{Inf}(X)\|_2 < 1$, and each coordinate of X lies in $[-c,c]$ and has mean 0 and variance v_i^2. Let $v = \sqrt{\sum_{i\in[m]} v_i^2}$. We have $\sigma_{max}(U) \le 2v$ and $\sigma_{min}(U) \ge \frac{v}{4}$, with probability at least $1-2\exp\left(-\frac{(1-\|\textsc{Inf}(X)\|_2)v^4}{32c^4nm}+4n\right)$.*

Since $\sum_{i\in S} v_i^2 > \frac{16c^2k\sqrt{|S|}}{1-\|\textsc{Inf}(\theta_S)\|_2}$, by Lemma 5, we have $\sigma_{max}(B) \le 2\sqrt{\sum_{i\in S} v_i^2}$ and $\sigma_{min}(B) \ge \sqrt{\sum_{i\in S} v_i^2}/4$ with probability at least $1-2\exp\left(-\frac{(1-\|\textsc{Inf}(\theta_S)\|_2)\cdot(\sum_{i\in S} v_i^2)^2}{64c^4k|S|}\right) \ge 1-2\exp(-4k)$. Therefore, $\|\hat{z}-z(t)\|_\infty \le \frac{32\sqrt{|S|k}}{\sqrt{\sum_{i\in S} v_i^2}}\cdot(\eta+\varepsilon)$ with probability at least $1-2\exp\left(-\frac{(1-\|\textsc{Inf}(\theta_S)\|_2)\cdot(\sum_{i\in S} v_i^2)^2}{64c^4k|S|}\right)$.

Query Complexity in Different Models: We set η to be ε.

- **Deterministic structure:** we have a $\left(\varepsilon, \frac{4\cdot\max_{j\in[k]} C_{jj}}{\alpha\beta}\cdot\varepsilon\right)$-query protocol using $k(\log\|A\|_\infty + 2\log(1/\varepsilon))$ queries.
- **Multivariate Gaussian distributions:** with probability at least $1-2\exp\left(-\frac{\mathrm{Tr}(\Sigma_S)}{16\cdot\rho(\Sigma_S)}\right)$ (no less than $1-2\exp(-4k)$ by our choice of S), we have a $\left(\varepsilon, \frac{64\sqrt{|S|k}}{\sqrt{\mathrm{Tr}(\Sigma_S)}}\cdot\varepsilon\right)$-query protocol using $|S|(\log\|A\|_\infty + 2\log(1/\varepsilon))$ queries.
- **Weakly dependent distributions:** with probability at least $1-2\exp\left(-\frac{(1-\|\textsc{Inf}(\theta_S)\|_2)\cdot(\sum_{i\in S} v_i^2)^2}{64c^4k|S|}\right)$ (no less than $1-2\exp(-4k)$ by our choice of S), we have a $\left(\varepsilon, \frac{64\sqrt{|S|k}}{\sqrt{\sum_{i\in S} v_i^2}}\cdot\varepsilon\right)$-query protocol using $|S|(\log\|A\|_\infty + 2\log(1/\varepsilon))$ queries.

□

4 BOUNDING THE LARGEST AND SMALLEST SINGULAR VALUES

We prove both Lemma 4 and 5 using an ε-net argument. We first state a lemma that says that for any matrix M, if we can bound the maximum value of $\|Mx\|_2$ over all points x in the ε-net, then we also bound the largest and smallest singular values of M.

Lemma 6 (Adapted from [41]). *For any $\varepsilon < 1$, there exists an ε-net $\mathcal{K} \subseteq S^{n-1}$, i.e., $\forall x \in S^{n-1}\ \exists y \in \mathcal{K}\ \|x-y\|_2 < \varepsilon$, such that $|\mathcal{K}| \le (3/\varepsilon)^n$. For any matrix $M \in \mathbb{R}^{m\times n}$, let $a = max_{x\in\mathcal{K}} \|Mx\|_2$ and $b = min_{x\in\mathcal{K}} \|Mx\|_2$, then $\sigma_{max}(M) \le \frac{a}{1-\varepsilon}$ and $\sigma_{min}(M) \ge b - \frac{\varepsilon}{1-\varepsilon}\cdot a$.*

Proof of Lemma 6: Let $x^* \in S^{n-1}$ be a vector that satisfies $\|Mx^*\|_2 = \sigma_{max}(M)$. Let x be a vector in $\mathcal{K}$ such that $\|x-x^*\|_2 \le \varepsilon$. Then $\sigma_{max}(M) = \|Mx^*\|_2 \le \|Mx\|_2 + \|M(x-x^*)\|_2 \le a + \varepsilon\sigma_{max}(M)$, which implies that $\sigma_{max}(M) \le \frac{a}{1-\varepsilon}$. On the other hand, for any $y \in S^{n-1}$, let $y' \in \mathcal{K}$ satisfies $\|y-y'\|_2 \le \varepsilon$, then $\|My\|_2 \ge \|My'\|_2 - \|M(y-y')\|_2 \ge b - \varepsilon\cdot\sigma_{max}(M) \ge b - \frac{\varepsilon}{1-\varepsilon}\cdot a$. □

4.1 Multivariate Gaussian Distributions

In this section, we prove the case where the columns of the random matrix are drawn from a multivariate Gaussian distribution. The key is again to prove that for every unit-vector, $\|Ux\|_2$ lies between $[c_1\cdot\mathbb{E}[\|Ux\|_2], c_2\cdot\mathbb{E}[\|Ux\|_2]]$ with high probability for some absolute constant c_1 and c_2 (Lemma 7). Lemma 4 follows from the combination of Lemma 7, 6, and the union bound.
Proof of Lemma 4: Let $Y^{(1)},\dots,Y^{(n)}$ be n i.i.d. samples from the distribution $\mathcal{N}(0,I_m)$, and $V := D^{1/2}[Y^{(1)},\dots,Y^{(s)}]$.

Proposition 2. *$\mathcal{N}(0,\Sigma) \overset{d}{=} \Lambda^T \circ \mathcal{N}(0,D)$ and $U \overset{d}{=} \Lambda^T V$.*

Proof. $\mathbb{E}[\Lambda^T D^{1/2} Y^{(i)} (Y^{(i)})^T D^{1/2}\Lambda] = \Lambda^T D^{1/2}\mathbb{E}[Y^{(i)}(Y^{(i)})^T]D^{1/2}\Lambda = \Lambda^T D\Lambda = \Sigma$. □

Since Λ is an orthonormal matrix, $\sigma_{max}(U) = \sigma_{max}(V)$ and $\sigma_{min}(U) = \sigma_{min}(V)$. We will proceed to show that both $\sigma_{max}(V)$ and $\sigma_{max}(V)$ concentrate around their means. We do so via an ε-net argument.

Lemma 7. *For any fix $x \in S^{n-1}$, $\mathbb{E}[\|Vx\|_2^2] = \mathrm{Tr}(D)$. Moreover,*

$$\Pr\left[\|Vx\|_2^2 \le \frac{\mathrm{Tr}(D)}{4}\right] \le \exp\left(-\frac{\mathrm{Tr}(D)}{8\cdot d_{max}}\right),$$

and

$$\Pr\left[\|Vx\|_2^2 \ge 2\mathrm{Tr}(D)\right] \le \exp\left(-\frac{\mathrm{Tr}(D)}{4\cdot d_{max}}\right).$$

Proof of Lemma 7: Let $g_1,\dots,g_n$ to be n i.i.d. samples from $\mathcal{N}(0,1)$. It is not hard to see that $Vx \overset{d}{=} (\sqrt{d_1}g_1,\dots,\sqrt{d_n}g_n)^T$, so we need to prove that $\sum_{i\in[n]} d_i g_i^2$ concentrates around its mean $\mathrm{Tr}(D)$.

$$\begin{aligned}
&\Pr\left[\sum_{i\in[n]} d_i g_i^2 \le \mathrm{Tr}(D) - t\right] \\
=&\Pr\left[\exp\left(\lambda\cdot(\mathrm{Tr}(D) - \sum_{i\in[n]} d_i g_i^2)\right) \ge \exp(\lambda t)\right] \qquad (\lambda>0 \text{ and will be specified later}) \\
\le&\frac{\exp(\lambda\mathrm{Tr}(D))\mathbb{E}\left[\exp\left(-\lambda\cdot\sum_{i\in[n]} d_i g_i^2\right)\right]}{\exp(\lambda t)} = \frac{\exp(\lambda\mathrm{Tr}(D))\prod_{i\in[n]}\mathbb{E}\left[\exp\left(-\lambda\cdot d_i g_i^2\right)\right]}{\exp(\lambda t)}
\end{aligned}$$

Since g_i^2 distributes according to a chi-square distribution, its moment generating function

$$\mathbb{E}\left[\exp\left(-\lambda\cdot d_i g_i^2\right)\right] = \frac{1}{\sqrt{1+2\lambda d_i}}.$$

If we choose λ to be no more than $1/2d_{max}$, since for any $a \in [0,1]$, $1+2a \geq e^a$, we have that

$$\frac{1}{\sqrt{1+2\lambda d_i}} \leq \exp(-\lambda d_i/2).$$

Putting everything together, we have that

$$\Pr\left[\sum_{i\in[n]} d_i g_i^2 \leq \mathrm{Tr}(D) - t\right] \leq \exp\left(-\lambda \cdot (t - \mathrm{Tr}(D)/2)\right).$$

When we choose $\lambda = 1/2d_{max}$ and $t = 3/4 \cdot \mathrm{Tr}(D)$, the RHS of the inequality becomes $\exp\left(-\frac{\mathrm{Tr}(D)}{8\cdot d_{max}}\right)$.

Next, we upper bound $\Pr\left[\sum_{i\in[n]} d_i g_i^2 \geq \mathrm{Tr}(D) + t\right]$ via a similar approach.

$$\begin{aligned}
&\Pr\left[\sum_{i\in[n]} d_i g_i^2 \geq \mathrm{Tr}(D) + t\right] \\
&= \Pr\left[\exp\left(\lambda \cdot \left(\sum_{i\in[n]} d_i g_i^2 - \mathrm{Tr}(D)\right)\right) \geq \exp(\lambda t)\right] \qquad (\lambda > 0 \text{ and will be specified later}) \\
&\leq \frac{\prod_{i\in[n]} \mathbb{E}\left[\exp\left(\lambda \cdot (d_i g_i^2 - d_i)\right)\right]}{\exp(\lambda t)}
\end{aligned}$$

Note that $\mathbb{E}\left[\exp\left(\lambda \cdot (d_i g_i^2 - d_i)\right)\right] = \frac{\exp(-\lambda d_i)}{\sqrt{1-2\lambda d_i}}$.

Proposition 3. *For any $x \in [0, 1/4]$, $\frac{\exp(-x)}{\sqrt{1-2x}} \leq \sqrt{1+2x}$.*

Proof of Proposition 3: We first state a few inequalities that are not hard to verify. First, for all $x > 0$, $e^{-x} \leq 1 - x + x^2$. Second, $\sqrt{1-4x^2} \geq 1 - 2x^2 - 8x^4$ if $x \in [0, 1/2)$. Finally, $1 - 2x^2 - 8x^4 \geq 1 - x + x^2$ if $x \in [0, 1/4]$. Combining all three inequalities, we have that

$$e^{-x} \leq \sqrt{1-4x^2} = \sqrt{1-2x}\sqrt{1+2x}, \text{ for all } x \in [0, 1/4].$$

□

If we choose λ to be no more than $1/4d_{max}$, then by Proposition 3, $\frac{\exp(-\lambda d_i)}{\sqrt{1-2\lambda d_i}} \leq \sqrt{1+2\lambda d_i}$, which is upper bounded by $\exp(\lambda d_i)$. Putting everything together, we have that

$$\Pr\left[\sum_{i\in[n]} d_i g_i^2 \geq \mathrm{Tr}(D) + t\right] \leq \exp\left(-\lambda(t - \mathrm{Tr}(D))\right).$$

When we choose $\lambda = 1/4d_{max}$ and $t = 2\mathrm{Tr}(D)$, the RHS of the inequality becomes $\exp\left(-\frac{\mathrm{Tr}(D)}{4\cdot d_{max}}\right)$.
□

Next, we only consider when the good event happens, that is, for all points x in the ε-net, $\|Vx\|_2 \in \left[\frac{\sqrt{\mathrm{Tr}(D)}}{2}, \sqrt{2\mathrm{Tr}(D)}\right]$. Combining Lemma 7 and the union bound, we know that the good event happens with probability at least $1 - 2\exp\left(-\frac{\mathrm{Tr}(D)}{8\cdot d_{max}} + \ln(3/\varepsilon)\cdot n\right)$. According to Lemma 6, $\sigma_{max}(V) \leq \frac{\sqrt{2\mathrm{Tr}(D)}}{1-\varepsilon}$ and $\sigma_{min}(V) \geq \frac{\sqrt{\mathrm{Tr}(D)}}{2} - \frac{\varepsilon}{1-\varepsilon}\cdot\sqrt{2\mathrm{Tr}(D)}$. If we choose $\varepsilon = 1/7$, then $\sigma_{max}(V) \leq 2\sqrt{\mathrm{Tr}(D)}$ and $\sigma_{min}(V) \geq \frac{\sqrt{\mathrm{Tr}(D)}}{4}$. □

4.2 Bounded Distributions with Weak Dependence

In this section, we prove the case where the columns of the random matrix are drawn from a m-dimensional distribution that satisfies weak dependence. The overall plan is similar to the one for multivariate Gaussian distributions. The key is again to prove that for every unit-vector, $\|Ux\|_2$ lies between $[c_1 \cdot \mathbb{E}[\|Ux\|_2], c_2 \cdot \mathbb{E}[\|Ux\|_2]]$ with high probability for some absolute constant c_1 and c_2 (Lemma 8). Lemma 5 then follows from the combination of Lemma 8, 6, and the union bound.

Proof of Lemma 5:

We first show that for each fix $x \in S^{n-1}$, $\|Ux\|_2$ is concentrates around its mean. Then, we apply Lemma 6 to bound $\sigma_{max}(U)$ and $\sigma_{min}(U)$.

Lemma 8. *Let $U = [X^{(1)}, \ldots, X^{(n)}]$ be a $m \times n$ random matrix, where each column of U is an independent copy of a m-dimensional random vector X. We assume that the coordinates of X are weakly dependent, i.e., $\|\textsc{Inf}(X)\|_2 < 1$, and each coordinate of X lies in $[-c, c]$ and has mean 0 and variance v_i^2. Let $v = \sqrt{\sum_{i\in[m]} v_i^2}$. For any fix $x \in S^{n-1}$, $\mathbb{E}[\|Ux\|_2^2] = v^2$ and*

$$\Pr\left[\left|\, \|Ux\|_2^2 - v^2\right| > t\right] \le 2\exp\left(-\frac{(1 - \|\textsc{Inf}(X)\|_2)\, t^2}{16c^4 nm}\right)$$

Proof of Lemma 8: We first expand $\|Ux\|_2^2$.

$$\|Ux\|_2^2 = \sum_{i\in[m]} \left(\sum_{j\in[n]} u_{ij}x_j\right)^2 = \sum_{i\in[m]} \left(\sum_{j\in[n]} u_{ij}^2 x_j^2 + 2\sum_{k\neq j} u_{ij}u_{ik}x_j x_k\right).$$

Therefore, $\mathbb{E}\left[\|Ux\|_2^2\right] = \sum_{i\in[m]} v_i^2 = v^2$. To prove that $\|Ux\|_2^2$ concentrates, we first need a result by Chatterjee [16].

Lemma 9 (Adapted from Theorem 4.3 in [16]). *Let X be a d-dimensional random vector. Suppose function f satisfies the following generalized Lipschitz condition:*

$$|f(x) - f(y)| \le \sum_{i\in[d]} c_i \mathbb{1}[x_i \neq y_i],$$

for any x and y in the support of X. If $\textsc{Inf}(X) < 1$, we have

$$\Pr\left[|f(X) - \mathbb{E}[f(X)]| \ge t\right] \le 2\exp\left(-\frac{(1 - \|\textsc{Inf}(X)\|_2)\, t^2}{\sum_{i\in[d]} c_i^2}\right).$$

The function we care about is $\|Ux\|_2^2$, where the variables are $\{u_{ij}\}_{i\in[m], j\in[n]}$. If U and U' only differs at the (i, j) entry, then

$$\begin{aligned}
&\left|\, \|Ux\|_2^2 - \|U'x\|_2^2 \,\right| \\
=&\left|u_{ij}^2 x_j^2 + 2\sum_{k\neq j} u_{ij}u_{ik}x_j x_k - (u'_{ij})^2 x_j^2 - 2\sum_{k\neq j} u'_{ij}u_{ik}x_j x_k\right| \\
\le& c^2 x_j^2 + 4c^2|x_j||x_k| \le 4c^2|x_j|\left(\sum_{k\in[n]} |x_k|\right) \le 4c^2\sqrt{n}|x_j|
\end{aligned}$$

We denote $4c^2\sqrt{n}|x_j|$ by c_{ij}. Clearly, for any U and U', $\left|\, \|Ux\|_2^2 - \|U'x\|_2^2 \,\right| \le \sum_{i,j\in[d]} c_{ij}\mathbb{1}[u_{ij} \neq u'_{ij}]$. Also, notice that $\textsc{Inf}(U) = I_n \otimes \textsc{Inf}(X)$, and therefore $\|\textsc{Inf}(U)\|_2 = \|\textsc{Inf}(X)\|_2$. [7] We apply Lemma 9

[7] $\otimes$ denotes the Kronecker product of the two matrices.

to $\|Ux\|_2^2$ and derive the following inequality:

$$\Pr\left[\left|\,\|Ux\|_2^2 - v^2\right| > t\right] \le 2\exp\left(-\frac{(1-\|\textsc{Inf}(X)\|_2)\,t^2}{\sum_{i\in[m],j\in[n]} c_{ij}^2}\right) = 2\exp\left(-\frac{(1-\|\textsc{Inf}(X)\|_2)\,t^2}{16c^4nm}\right).$$

□

Next, we only consider when the good event happens, that is, for all points x in the ε-net, $\|Ux\|_2 \in \left[\frac{v}{2}, \sqrt{2}v\right]$. Combining Lemma 8 (setting $t = 3/4v^2$) and the union bound, we know that the good event happens with probability at least $1-2\exp\left(-\frac{(1-\|\textsc{Inf}(X)\|_2)9v^4}{256c^4nm} + \ln(3/\varepsilon)\cdot n\right)$. According to Lemma 6, $\sigma_{max}(U) \le \frac{\sqrt{2}v}{1-\varepsilon}$ and $\sigma_{min}(U) \ge \frac{v}{2} - \frac{\varepsilon}{1-\varepsilon}\cdot\sqrt{2}v$. If we choose $\varepsilon = 1/7$, then $\sigma_{max}(U) \le 2v$ and $\sigma_{min}(U) \ge \frac{v}{4}$. □

REFERENCES

[1] Saeed Alaei. 2011. Bayesian Combinatorial Auctions: Expanding Single Buyer Mechanisms to Many Buyers. In *the 52nd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*.

[2] Saeed Alaei, Hu Fu, Nima Haghpanah, Jason Hartline, and Azarakhsh Malekian. 2012. Bayesian Optimal Auctions via Multi- to Single-agent Reduction. In *the 13th ACM Conference on Electronic Commerce (EC)*.

[3] Moshe Babaioff, Yannai A Gonczarowski, and Noam Nisan. 2021. The menu-size complexity of revenue approximation. *Games and Economic Behavior* (2021).

[4] Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S. Matthew Weinberg. 2014. A Simple and Approximately Optimal Mechanism for an Additive Buyer. In *the 55th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*.

[5] Dirk Bergemann and Karl Schlag. 2011. Robust monopoly pricing. *Journal of Economic Theory* 146, 6 (2011), 2527–2543.

[6] Johannes Brustle, Yang Cai, and Constantinos Daskalakis. 2019. Multi-Item Mechanisms without Item-Independence: Learnability via Robustness. *CoRR* abs/1911.02146 (2019). arXiv:1911.02146 http://arxiv.org/abs/1911.02146

[7] Johannes Brustle, Yang Cai, and Constantinos Daskalakis. 2020. Multi-item mechanisms without item-independence: Learnability via robustness. In *EC*.

[8] Yang Cai and Constantinos Daskalakis. 2017. Learning Multi-item Auctions with (or without) Samples. In *FOCS*.

[9] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. 2012. An Algorithmic Characterization of Multi-Dimensional Mechanisms. In *the 44th Annual ACM Symposium on Theory of Computing (STOC)*.

[10] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. 2012. Optimal Multi-Dimensional Mechanism Design: Reducing Revenue to Welfare Maximization. In *the 53rd Annual IEEE Symposium on Foundations of Computer Science (FOCS)*.

[11] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. 2013. Reducing Revenue to Welfare Maximization : Approximation Algorithms and other Generalizations. In *the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*.

[12] Yang Cai, Constantinos Daskalakis, and S. Matthew Weinberg. 2013. Understanding Incentives: Mechanism Design becomes Algorithm Design. In *the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*.

[13] Yang Cai, Nikhil R. Devanur, and S. Matthew Weinberg. 2016. A Duality Based Unified Approach to Bayesian Mechanism Design. In *STOC*.

[14] Yang Cai and Zhiyi Huang. 2013. Simple and Nearly Optimal Multi-Item Auctions. In *the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*.

[15] Yang Cai and Mingfei Zhao. 2017. Simple Mechanisms for Subadditive Buyers via Duality. In *STOC, 2017*.

[16] Sourav Chatterjee. 2005. *Concentration inequalities with exchangeable pairs*. Ph. D. Dissertation. Citeseer.

[17] Shuchi Chawla, Jason D. Hartline, and Robert D. Kleinberg. 2007. Algorithmic Pricing via Virtual Valuations. In *the 8th ACM Conference on Electronic Commerce (EC)*.

[18] Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. 2010. Multi-Parameter Mechanism Design and Sequential Posted Pricing. In *the 42nd ACM Symposium on Theory of Computing (STOC)*.

[19] Shuchi Chawla and J. Benjamin Miller. 2016. Mechanism Design for Subadditive Agents via an Ex-Ante Relaxation. In *Proceedings of the ACM Conference on Economics and Computation (EC)*.

[20] Xi Chen, Ilias Diakonikolas, Anthi Orfanou, Dimitris Paparas, Xiaorui Sun, and Mihalis Yannakakis. 2015. On the complexity of optimal lottery pricing and randomized mechanisms. In *Proceedings of the 56th Annual Symposium on Foundations of Computer Science (FOCS)*.

[21] Constantinos Daskalakis. 2015. Multi-item auctions defying intuition? *ACM SIGecom Exchanges* 14, 1 (2015), 41–75.

[22] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. 2013. Mechanism design via optimal transport. In *ACM Conference on Electronic Commerce, EC '13, Philadelphia, PA, USA, June 16-20, 2013*, Michael J. Kearns, R. Preston McAfee, and Éva Tardos (Eds.). ACM, 269–286. https://doi.org/10.1145/2482540.2482593
[23] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. 2014. The Complexity of Optimal Mechanism Design. In *the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*.
[24] Constantinos Daskalakis, Alan Deckelbaum, and Christos Tzamos. 2017. Strong Duality for a Multiple-Good Monopolist. *Econometrica* 85, 3 (2017), 735–767.
[25] Constantinos Daskalakis, Maxwell Fishelson, Brendan Lucier, Vasilis Syrgkanis, and Santhoshini Velusamy. 2020. Simple, Credible, and Approximately-Optimal Auctions. In *EC*.
[26] David Donoho and Victoria Stodden. 2003. When Does Non-Negative Matrix Factorization Give a Correct Decomposition into Parts?. In *Proceedings of the 16th International Conference on Neural Information Processing Systems* (Whistler, British Columbia, Canada) *(NIPS'03)*. MIT Press, Cambridge, MA, USA, 1141–1148.
[27] Shaddin Dughmi, Li Han, and Noam Nisan. 2014. Sampling and representation complexity of revenue maximization. In *WINE*.
[28] Paul Dütting, Zhe Feng, Harikrishna Narasimhan, David Parkes, and Sai Srivatsa Ravindranath. 2019. Optimal auctions through deep learning. In *International Conference on Machine Learning*. PMLR, 1706–1715.
[29] Zhe Feng, Harikrishna Narasimhan, and David C Parkes. 2018. Deep learning for revenue-optimal auctions with budgets. In *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent Systems*.
[30] Yiannis Giannakopoulos and Elias Koutsoupias. 2014. Duality and optimality of auctions for uniform distributions. In *Proceedings of the 15th ACM conference on Economics and Computation (EC)*.
[31] Kira Goldner and Anna R Karlin. 2016. A prior-independent revenue-maximizing auction for multiple additive bidders. In *International Conference on Web and Internet Economics*. Springer, 160–173.
[32] Yannai A Gonczarowski and S Matthew Weinberg. 2018. The Sample Complexity of Up-to-ε Multi-Dimensional Revenue Maximization. In *FOCS*.
[33] Sergiu Hart and Noam Nisan. 2012. Approximate Revenue Maximization with Multiple Items. In *EC*.
[34] Sergiu Hart, Noam Nisan, et al. 2013. *The menu-size complexity of auctions*. Center for the Study of Rationality.
[35] Ian A Kash and Rafael M Frongillo. 2016. Optimal auctions with restricted allocations. In *Proceedings of the 2016 ACM Conference on Economics and Computation*. 215–232.
[36] Robert Kleinberg and S. Matthew Weinberg. 2012. Matroid Prophet Inequalities. In *the 44th Annual ACM Symposium on Theory of Computing (STOC)*.
[37] Jamie Morgenstern and Tim Roughgarden. 2016. Learning simple auctions. In *Conference on Learning Theory*. PMLR, 1298–1318.
[38] Roger B. Myerson. 1981. Optimal Auction Design. *Mathematics of Operations Research* 6, 1 (1981), 58–73.
[39] Noam Nisan, Tim Roughgarden, E. Tardos, and V. V. Vazirani (Eds.). 2007. *Algorithmic Game Theory*. Cambridge University Press.
[40] Aviad Rubinstein and S. Matthew Weinberg. 2015. Simple Mechanisms for a Subadditive Buyer and Applications to Revenue Monotonicity. In *EC*. https://doi.org/10.1145/2764468.2764510
[41] Mark Rudelson. 2014. Recent developments in non-asymptotic theory of random matrices. *Modern aspects of random matrix theory* 72 (2014), 83.
[42] Weiran Shen, Pingzhong Tang, and Song Zuo. 2019. Automated mechanism design via neural networks. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems*.
[43] Volker Strassen. 1965. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics* 36, 2 (1965), 423–439.
[44] Vasilis Syrgkanis. 2017. A Sample Complexity Measure with Applications to Learning Optimal Auctions. In *Advances in Neural Information Processing Systems (NeurIPS)*.
[45] James M Varah. 1975. A lower bound for the smallest singular value of a matrix. *Linear Algebra and its applications* 11, 1 (1975), 3–5.
[46] Andrew Chi-Chih Yao. 2015. An n-to-1 Bidder Reduction for Multi-item Auctions and its Applications, In SODA. *CoRR* (2015). http://arxiv.org/abs/1406.3278

A MISSING PROOF OF LEMMA 2

Proof of Lemma 2: The proof essentially follows from the same analysis as Theorem 3 in [7]. We only provide a sketch here. Since we are working with the matrix factorization model and can directly exploit the low dimensionality of the latent representation, we manage to replace the dependence on N with $\|A\|_\infty$ in both the revenue loss and violation of the truthfulness constraints. Our proof

relies on the idea of *"simultaneously coupling"* by Brustle et al. [7]. More specifically, it couples $\widehat{F}_{z,i}$ with every distribution $F_{z,i}$ in the ε-Prokhorov-ball around $\widehat{F}_{z,i}$. If we round both $\widehat{F}_{z,i}$ and any $F_{z,i}$ to a random grid G with size δ, we can argue that the *expected total variation distance* (over the randomness of the grid) between the two rounded distributions is $O(\varepsilon + \frac{\varepsilon}{\delta})$ (using Theorem 2 in [7]). Now consider the following mechanism: choose a random grid G, round the bids to the random grid, apply the mechanism M_G that we designed for the rounded distribution of $\bigtimes_i \widehat{F}_{z,i}$. More specifically, M_G is the following mechanism: for each bid b, use $\mathcal{S}_i(b_i, \delta)$ to sample a bid b'_i and run $\widehat{M}$ on the bid profile $(b'_1, \ldots, b'_m)$. Since the expected total variation distance (over the randomness of the grid) between the two rounded distributions is $O(\varepsilon + \frac{\varepsilon}{\delta})$, we only need to argue that when the given distribution and the true distribution are close in total variation distance, we can robustify the mechanism designed for one distribution for the other distribution. This is a much easier task, and we again use a similar argument in [7] to prove it. Combining everything, we can show that the randomized mechanism we constructed is approximately-truthful and only loses a negligible revenue compared to $\widehat{M}$ under any distribution that is within the ε-Prokhorov-ball around the given distribution. □