Learning Simple Auctions

Jamie Morgenstern
Tim Roughgarden

April 11, 2016

Abstract

We present a general framework for proving polynomial sample complexity bounds for the problem of learning from samples the best auction in a class of "simple" auctions. Our framework captures all of the most prominent examples of "simple" auctions, including anonymous and non-anonymous item and bundle pricings, with either a single or multiple buyers. The technique we propose is to break the analysis of auctions into two natural pieces. First, one shows that the set of allocation rules have large amounts of structure; second, fixing an allocation on a sample, one shows that the set of auctions agreeing with this allocation on that sample have revenue functions with low dimensionality. Our results effectively imply that whenever it's possible to compute a near-optimal simple auction with a known prior, it is also possible to compute such an auction with an unknown prior (given a polynomial number of samples).

1 Introduction

The standard economic approach for designing revenue-maximizing auctions assumes all information unknown to the designer is drawn from some prior distribution, about which the designer has perfect information. With this "perfect" prior in hand, the designer fine-tunes an auction to optimize for her *expected* revenue over draws of the unknown information from the prior. While this model allows for quite strong results relating her chosen auction's revenue to the optimal revenue, three related difficulties arise from using this design pattern in practice. First, for any particular setting, it is unlikely that the designer could actually formulate a perfect prior over the market's hidden information. Second, if the market designer has an imperfect prior, it is possible that her optimal auction has overfit to this prior and will have very poor revenue when run on the (similar) true prior. Finally, the optimal auction for a particular prior can be quite complicated and unintuitive.

These obstacles can be addressed in a rigorous manner by designing auctions as a function of several *samples* of the unknown data (usually, buyers' valuations) drawn from an unknown distribution, with the knowledge that our goal is to earn high revenue on a fresh draw from the same distribution. It is reasonable to expect that experienced sellers have previous records of the bids made by previous participants in their market. Moreover, if an auction is guaranteed perform well on future draws from a distribution to which it only had sample access, it will have strong generalization properties if its sample size was sufficiently large.

How many samples are necessary to achieve such a guarantee? The answer depends upon the complexity of the set of auctions the seller might select. The more complex the class of auctions, the lower the class's "representation error" (the higher the revenue the seller might be able to extract); on the other hand, a more complex class of auctions will have higher "generalization error" (loss in revenue from optimizing over the sample rather than the true prior) for a fixed sample size.

The mechanism design community has placed "simplicity" as a design goal for its own sake for single-parameter (Hartline and Roughgarden, 2009) and multi-parameter (Chawla et al., 2007, 2010; Babaioff et al., 2014; Rubinstein and Weinberg, 2015) auctions.[1] Recent work (Morgenstern and Roughgarden, 2015) proposed the use of a class's pseudo-

[1] A canonical "single-parameter" problem is a single-item auction — each bidder either "wins" or "loses". A canonical "multi-parameter" problem is a multi-item auction — with k items, each bidder faces 2^k different possibilities. Multi-parameter problems are well known to be much more ill-behaved than single-parameter problems. For example, there is no general multi-parameter analog of Myerson's (single-parameter) theory of revenue-maximizing auctions Myerson (1981).

dimension as a formal notion of simplicity for single-parameter auctions, and proves for general single-parameter settings there exist classes of auctions with small representation error (which contain a nearly-optimal auction) *and* have small pseudo-dimension or generalization error (a polynomial-sized sample suffices to learn a nearly-optimal auction from that class).

In this work, we give a general framework for bounding the pseudo-dimension of classes of multi-parameter auctions. Our results imply polynomial sample complexity bounds for revenue maximization over all of the aforementioned "simple" auctions. In effect, our results imply that whenever it's possible to compute a near-optimal simple auction with a known prior, it is also possible to compute such an auction with an unknown prior (given a polynomial number of samples).

One concrete example of a class of well-studied "simple" multi-parameter auctions comes from Babaioff et al. (2014). Consider a single bidder whose valuation is *additive* over k items: there is a vector $v \in \mathbb{R}^k$ such that the bidder's valuation for a bundle $B \subseteq [k]$ is $\sum_{j \in B} v_j$. An *item pricing* is defined by a vector $\mathbf{p} \in \mathbb{R}^k$, and offers the agent any bundle B for price $\sum_{j \in B} \mathbf{p}_j$. A *grand bundle pricing* is defined by a single real number $q \in \mathbb{R}$ and offers the bundle $B = [k]$ for the price q. When a single additive buyer's valuation $v \sim \mathcal{D}_1 \times \cdots \times \mathcal{D}_k$ is drawn from a product distribution, either the best item pricing or the best grand-bundle pricing will earn $1/6$ of optimal revenue. Babaioff et al. (2014) assume that the D_j's are known a priori and choose item and bundle prices as a function of the distributions. Can we instead learn from samples the best auction from the class consisting of all item and bundle prices? The main result in Babaioff et al. (2014) provides a bound on the representation error of this class; our work provides the first sample complexity bound (for this and many other classes).

Our Main Results We present a general framework for bounding sample complexity for "simple" combinatorial auctions, when considering auctions as functions from valuations to revenue. Formally, we study the following question, and provide a technique for answering it in many interesting cases:

> Given a class of multiparameter auctions $C : \mathcal{V}^n \to \mathbb{R}$ (each auction maps any n-tuple of combinatorial valuations to the revenue achieved by the auction on those valuations), how large must m be such that the empirical revenue maximizer in C over m samples drawn from $\mathcal{D}$ earns $OPT(C) - \epsilon$ expected revenue on fresh sample drawn from $\mathcal{D}$?

Our main technical contributions are first to show a general way to measure the sample complexity of single-buyer mechanisms, which are interesting in their own right, and second to show a reduction in bounding the sample complexity for the multi-buyer auctions to bounding the sample complexity of single-buyer auctions. This reduction and our general framework apply to all known multi-buyer simple auctions from the literature. We then instantiate this framework and show that it is flexible enough to bound the sample complexity of a large class of simple auction classes from the literature. In particular, we bound the pseudo-dimension of item pricings, grand bundle pricings, and second-price item or bundle auctions with reserves, covering the set of known simple auctions which approximately optimize revenue. The following table summarizes our results, as well as the known approximation guarantees these auctions provide. We note that Theorem D.1 is proven using a direct argument rather than this framework.

Summary of Simple Auction Properties

Class	Valuations	PD anon, nonanon	Rev APX anonymous	Rev APX nonanonymous
Grand bundle pricing	General	$O(1)$, $O(n)$ Corollary 4.3		
Item Pricing	General	$\tilde{O}(k^2)$, $\tilde{O}(k^2 n)$, Corollary 4.4	3 (1 unit-demand bidder) (Chawla et al., 2007)	10.7 (n unit-demand bidders) (Chawla et al., 2010)
Item and Grand Bundle Pricing	General	$\tilde{O}(k^2)$, $\tilde{O}(k^2 n)$	6 (1 additive bidder) (Babaioff et al., 2014) 312 (1 subadditive bidder) (Rubinstein and Weinberg, 2015)	
	Additive	$\tilde{O}(k)$, $\tilde{O}(kn)$, Theorem D.1		
Second-price item auctions with item reserves	Additive	$\tilde{O}(k^2)$, $\tilde{O}(k^2 n)$, Corollary 4.5 $\tilde{O}(k)$, $\tilde{O}(kn)$, Theorem D.1		48 (n additive buyers), (Yao, 2015)

These results imply that a polynomial-sized sample suffices to learn a nearly-optimal auction from these classes of simple auctions. Thus, when combined with results from the literature, it is possible to learn auctions which earn a constant-factor of optimal revenue for a single additive (Babaioff et al., 2014) or subadditive bidder (Rubinstein and Weinberg, 2015), n unit-demand bidders (Chawla et al., 2010), and n additive bidders (Yao, 2015).[2]

For many classes of auctions, our sample complexity bounds do not rely on any structural assumptions about buyers' valuations (only that their utilities are quasilinear in money and that they will behave by maximizing their own utility). We point to this flexibility as a key feature of our techniques: for bidders with general valuation functions, it can be quite complicated to reason about bidder's behavior directly. We also formally describe the allocations of these auction classes as coming from *sequential* allocation procedures all drawn from the same class, and show any class which has allocations which can be described this way also has a provably simple class of allocation functions. This reduction may be of independent interest for proving other classes of auctions have small sample complexity.

1.1 Related Work

There has been a recent surge in the design of single-parameter revenue-maximizing auctions[3] from samples (Elkind, 2007; Balcan et al., 2007, 2008a; Cole and Roughgarden, 2014; Huang et al., 2015; Medina and Mohri, 2014; Roughgarden and Schrijvers, 2015; Morgenstern and Roughgarden, 2015; Devanur et al., 2015); we focus here on the problem of designing auctions from samples for multi-parameter settings. Optimal auctions for combinatorial settings are substantially more complex than for single-parameter settings, even before introducing questions of sample complexity. Item pricings in particular have been the subject of much study with respect to their constant approximations when buyers' values for items are independent,[4] for welfare (Kelso Jr and Crawford, 1982; Feldman et al., 2015) and revenue in the worst-case for for a single (Chawla et al., 2007) and n unit-demand bidders (Chawla et al., 2010), a single (Babaioff et al., 2014) additive buyers, and a single subadditive buyer (Rubinstein and Weinberg, 2015).[5] For the additive and subadditive buyer results, the theorems state that the better of the best item pricing and best grand bundle pricing achieve a constant factor of optimal revenue. Combining the result of Babaioff et al. (2014) and a recent result of Yao (2015), it is possible to earn a constant factor of optimal revenue for n additive buyers using the better of the best grand bundle pricing and an auction format related to item pricing (the second-price item auction with item-specific reserve prices). All n-buyer results for revenue rely on the use of *nonanonymous* item and grand bundle pricings. These results can be thought of as bounding the representation error of using these classes of auctions for revenue maximization; our work can be thought of as complementing these results by bounding the classes' generalization error.

Item pricings are also sufficiently simple that the sample complexity of choosing welfare-optimal (Feldman et al., 2015; Hsu et al., 2016) and revenue-optimal (Balcan et al., 2008a) item pricings has been explored. Balcan et al. (2008a) study the sample complexity of anonymous item pricing for combinatorial auctions with unlimited supply and employ one technique which bears some resemblance to our framework. Fixing a sample of size m, they bound the number of distinct allocation labelings L of that sample by anonymous item pricing using a geometric interpretation of anonymous pricings. Such an argument seems difficult to extend to other classes of auctions (for example, nonanonymous item pricings). We ultimately suggest the use of linear separability as a tool to bound L, an argument which applies to many distinct classes of auctions with finite supply, and doesn't rely on the particular geometry of anonymous item pricings.

We use the concept of linear separability (Daniely and Shalev-Shwartz, 2014) to prove bounds on the pseudo-dimension of many classes of auctions; this tool was also used by Balcan et al. (2014) in studying the sample complexity of learning the valuation function of a single buyer when goods are divisible and the valuation functions are either additive, Leontiff, or Separable Piecewise-Linear Concave; our results apply to multiple bidders, when items are

[2] Our framework also applies to learning simple auctions with good welfare guarantees, as in Feldman et al. (2015); all that changes is the real-valued function associated with an auction. Welfare guarantees are simpler than revenue guarantees (since the objective function value depends on the allocation only) so we concentrate on the latter.

[3] A generalization of single-item auctions, where each buyer can be described by a single real number representing her value for being selected as a winner.

[4] Dughmi et al. (2014) show that when items' values are allowed to be correlated, for a single unit-demand bidder, the sample complexity required to compute a constant-factor approximation to the optimal auction is necessarily exponential (in m).

[5] For more general valuation profiles and without item-wise independence, it is known that item pricings can also achieve super-constant revenue approximations, see Balcan et al. (2008b); Chakraborty et al. (2013).

indivisible, and most of them to arbitrary valuation classes (including superadditive valuations). Hsu et al. (2016) also studied the linear separability and used it to bound the pseudo-dimension of welfare maximization for item pricings as well as the concentration of demand for any particular good.

2 Preliminaries

Bayesian Mechanism Design Preliminaries In this section, we provide the definitions and main results regarding simple multi-parameter mechanism design necessary for proving our main results. We consider the problem of selling k heterogeneous items to n bidders. Each bidder $i \in [n]$ can be described by a *combinatorial valuation function* $v_i \in \mathcal{V} \subseteq (2^k \to \mathbb{R})$, and is assumed to be *quasilinear in money*, meaning that her utility for a bundle B with price $p(B)$ is exactly $u_i(B, p) = v_i(B) - p(B)$. We will assume all valuation functions are *monotone*, $v(B) \leq v(B')$ for all $B \subseteq B'$. An auction $\mathcal{A}$ is comprised of an allocation rule $\mathcal{A}_1 : \mathcal{V}^n \to [n]^k$ and a payment rule $\mathcal{A}_2 : \mathcal{V}^n \to \mathbb{R}^n$. We will only consider direct revelation mechanisms, for which it is the best-response for any buyer to reveal v_i to any mechanism $\mathcal{A}$. The valuation function v_i is assumed to be known to agent i but not to the designer of the auction, who must choose an auction $\mathcal{A}$ before observing $v_1, \ldots, v_n$.

We will assume that bidder i's valuation is drawn independently from some distribution $\mathcal{D}_i$ over valuation profiles. We assume the support of the distribution $\mathcal{D} = \mathcal{D}_1 \times \ldots \times \mathcal{D}_n$ is in $[0, H]^n$. We will refer to the *revenue* of an auction $\mathcal{A}$ on a particular instance $v = (v_1, \ldots, v_n)$ as $\sum_i \mathcal{A}_2(v)_i$, and the (expected) revenue of $\mathcal{A}$ as $\textsc{Rev}(\mathcal{A}, \mathcal{D}) = \mathbb{E}_{v\sim\mathcal{D}}[\sum_i \mathcal{A}_2(v)_i]$. When a bidder's valuation v_i can be represented as $v_{i1}, \ldots, v_{ik}$ such that $v_i(S) = \sum_{j\in S} v_{ij}$, we say that i is *additive*; when v_i can be represented as k numbers $v_{i1}, \ldots, v_{ik}$ such that $v_i(S) = \max_{j\in S} v_{ij}$, we say that i is *unit-demand*. If, for all $S, T \subseteq [k]$, $v_i(S) + v_i(T) \geq v_i(S \cup T)$, we say v_i is *subadditive*.

Several particular kinds of auctions are of particular use when (approximately) optimizing for revenue in multi-parameter settings. An auction is an (anonymous) *item pricing* if it sets price p_j for each item $j \in [k]$, and offers buyers in some fixed order any bundle B of remaining items for price $\sum_{j\in B} p_j$; each buyer then chooses the bundle maximizing her utility. An auction is a *non-anonymous* item pricing if it sets price p_{ij} for each $j \in [k], i \in [n]$, and offers buyers in some fixed order any bundle B of remaining items; buyer i will be offered B for a price of $\sum_{j\in B} p_{ij}$. An anonymous (or nonanonymous) grand bundle pricing sets a single price p (p_i) for the "grand" bundle $[k]$ of all items, and offers the grand bundle to buyers in some fixed order until the grand bundle is sold. Throughout the paper, we will assume this fixed order is fixed (namely, not a parameter of the design space), and that it places bidder 1 first in the ordering, 2 second, and so on. When buyers are additive, we will also consider the *second-price* item auction, which sells each item to the highest bidder for that item at the second-highest bid for that item, and the second-price grand bundle auction which sells the grand bundle to the highest bidder for it at the second-highest bid. Finally, we will consider the second-price item (grand-bundle) auction with both anonymous and non-anonymous item reserves, which sell to the highest bidder for that item at the maximum of the second-highest bid and the item's reserve (for that bidder), or to no one if the highest bidder's bid is below her reserve. These auction classes achieve constant-factor approximations for revenue in many special cases: for one (Chawla et al., 2007) and n (Chawla et al., 2010) unit-demand bidders, for one (Babaioff et al., 2014) and n (Yao, 2015) bidders, and for one subadditive bidder (Rubinstein and Weinberg, 2015) (see Section C, where we have included the formal theorem statements for completeness).

Learning Theory Preliminaries In this section, we provide definitions and useful tools for bounding the sample complexity of learning a class of real-valued functions $\mathcal{F}$. We omit discussion of binary-labeled learning and the definitions of uniform versus PAC learning for reasons of space (see Section B for further details).

Real-Valued Labels Both PAC learnability and uniform learnability of binary-valued functions can be well-characterized in terms of the class's VC dimension. When learning real-valued functions (for example, to guarantee convergence of the revenue of various auctions), we use a real-valued analog to VC dimension (which will give a sufficient but not necessary condition for uniform convergence). We will work with the *pseudo-dimension* (Pollard, 1984), one standard generalization. Formally, let $c : \mathcal{V} \to [0, H]$ be a real-valued function over $\mathcal{V}$, and $\mathcal{F}$ be the class we are learning over. Let S be a sample drawn from $\mathcal{D}$, $|N| = m$, labeled according to c. Both the empirical and true error of a hypothesis $\hat{c}$ are defined as before, though $|\hat{c}(v) - c(v)|$ can now take on values in $[0, H]$ rather than in $\{0, 1\}$. Let

$(r_1, \ldots, r_m) \in [0, H]^m$ be a set of *targets* for N. We say $(r_1, \ldots, r_m)$ *witnesses* the shattering of N by $\mathcal{F}$ if, for each $T \subseteq N$, there exists some $c_T \in \mathcal{F}$ such that $c_T(v_q) \geq r_q$ for all $v_q \in T$ and $c_T(v_q) < r_q$ for all $v_q \notin T$. If there exists some $\vec{r}$ witnessing the shattering of N, we say N is *shatterable* by $\mathcal{F}$. The *pseudo-dimension* of $\mathcal{F}$, denoted $\mathcal{PD}(\mathcal{F})$, is the size of the largest set S which is shatterable by $\mathcal{F}$. We will derive sample complexity upper bounds from the following theorem, which connects the sample complexity of uniform learning over a class of real-valued functions to the pseudo-dimension of the class.

Theorem 2.1 (E.g. Anthony and Bartlett (1999)) *Suppose $\mathcal{F}$ is a class of real-valued functions with range in $[0, H]$ and pseudo-dimension $\mathcal{PD}(\mathcal{F})$. For every $\epsilon > 0, \delta \in [0, 1]$, the sample complexity of (ϵ, δ)-uniformly learning the class $\mathcal{F}$ is*

$$n = O\left(\left(\frac{H}{\epsilon}\right)^2 \left(\mathcal{PD}(\mathcal{F}) \ln \frac{H}{\epsilon} + \ln \frac{1}{\delta}\right)\right).$$

Moreover, a conceptually simple algorithm achieves the guarantee in Theorem 2.1: simply output the function $c \in \mathcal{F}$ with the smallest empirical error on the sample. These algorithms are called *empirical risk minimizers*.

Multi-Labeled Learning The main goal of our work is to bound the sample complexity of revenue maximization for multi-parameter classes of auctions (via bounding these classes' pseudo-dimension); our proofs first bound the number of labelings of *purchased bundles* which these auctions can induce on a sample of size m. Then, we argue about the behavior of the revenue of all auctions which agree on the purchased bundles for every sample to bound the pseudo-dimension. Since bundles are neither binary nor real-valued, we now briefly mention several tools which we use for learning in the so-called *multi-label* setting.

The first of these tools is that of *compression schemes* for a class of functions.

Definition 2.2 A *compression scheme* for $\mathcal{F} : \mathcal{V} \to \mathcal{Y}$, of size d consists of

- a *compression* function
$$\textbf{compress} : (\mathcal{V} \times \mathcal{Y})^m \to (\mathcal{V} \times \mathcal{Y})^d,$$
where $\textbf{compress}(N) \subseteq N$ and $d \leq m$; and
- a *decompression* function
$$\textbf{decompress} : (\mathcal{V} \times \mathcal{Y})^d \to \mathcal{F}.$$

For any $f \in \mathcal{F}$ and any sample $(v_1, f(v_1)), \ldots, (v_m, f(v_m))$, the functions satisfy

$$\textbf{decompress} \circ \textbf{compress}((v_1, f(v_1)), \ldots, (v_m, f(v_m))) = f' \in \mathcal{F}$$

where $f'(v_q) = f(v_q)$ for each $q \in [m]$.

Intuitively, a compression function selects a subset of d "most relevant" points from a sample, and based on these points, the decompression scheme selects a hypothesis. When such a scheme exists, the learning algorithm **decompress** $\circ$ **compress** is an empirical risk minimizer. Furthermore, this compression-based learning algorithm has sample complexity bounded by a function of d, which plays a role analogous to VC dimension in the sample complexity guarantees.

Theorem 2.3 (Littlestone and Warmuth (1986)) *Suppose $\mathcal{F}$ has a compression scheme of size d. Then, the* PAC *complexity of $\mathcal{F}$ is at most $m = O\left(\frac{d \ln \frac{1}{\epsilon} + \ln \frac{1}{\delta}}{\epsilon}\right)$.*

While compression schemes imply useful sample complexity bounds, it can be hard to show that a particular hypothesis class admits a compression scheme. One general technique is to show that the class is linearly separable in a higher-dimensional space.

Definition 2.4 A class $\mathcal{F}$ is *d-dimensionally linearly separable* if there exists a function $\psi : \mathcal{V} \times \mathcal{Y} \to \mathbb{R}^d$ and for any $f \in \mathcal{F}$, there exists some $w^f \in \mathbb{R}^d$ with $f(v) \in \text{argmax}_y \langle w^f, \psi(v, y) \rangle$ and $|\text{argmax}_y \langle w^f, \psi(v, y) \rangle| = 1$.

It is known that a d-dimensional linearly separable class admits a compression scheme of size d.

Theorem 2.5 (Theorem 5 of Daniely and Shalev-Shwartz (2014)) *Suppose $\mathcal{F}$ is d-dimensionally linearly separable. Then, there exists a compression scheme for $\mathcal{F}$ of size d.*

If a class is linearly separable, this greatly restricts the number of labelings it can induce on a sample of size m, a trick used in Hsu et al. (2016) and also in the next section of this paper.

We also briefly mention that if a class $\mathcal{F}$ is linearly separable, post-processing the class with a fixed function also yields a linearly separable class over the resulting label space.

Observation 1
Suppose $\mathcal{F}$ is d-dimensionally linearly separable over Q. Fix some $q : Q \to Q'$. Then, the set $q \circ \mathcal{F} = \{q \circ f | f \in \mathcal{F}\}$ is d-dimensionally linearly separable over Q'.

With these tools in hand, our roadmap is as follows: for a class of auctions, we first prove that the class (which labels valuations by utility-maximizing bundles purchased) is linearly separable, which then implies an upper-bound on how many distinct bundle labelings one can have for a fixed sample. Then, we argue about the pseudo-dimension of the class (which labels a valuation by the revenue achieved when that agent buys her utility-maximizing bundle) by considering only those auctions which all have the same bundle labeling of m samples and arguing about the behavior of revenue of those auctions.

3 A Framework for Bounding Pseudo-dimension Via Intermediate Discrete Labels

We now propose a new framework for bounding the pseudo-dimension of many well-structured classes of real-valued functions. Suppose $\mathcal{F}$ is some set of real-valued functions whose pseudo-dimension we wish to bound. Suppose that, for each $f \in \mathcal{F}$, f can be "factored" into a pair of functions (f_1, f_2) such that $f_2(f_1(x), x) = f(x)$ for any x. There are always "trivial" factorings, where the function $f_2 = f$ or $f_1(x) = x$, but the interesting case arises when both $f_1(x)$ and f_2 (fixing $f_1(x)$) depend in a very limited way upon x. In particular, if the set of functions $\{f_1\}$ are very structured, and fixing $f_1(x)$ the set of functions $\{f_2\}$ only depend upon x in some very mild way, this will imply that $\mathcal{F}$ itself has small pseudo-dimension. Intuitively, this will allow us to "bucket" functions by their values according to f_1 on some sample, and bound the pseudo-dimension of each of those buckets separately.

Our particular technique for showing such a property is first to show that the set of functions $\{f_1\}$ are *linearly separable* in a dimensions, then to fix some sample S of size m and some f_1, and to upper-bound by b the pseudo-dimension of the set of functions f_2 whose associated f_1' agrees with the labeling of f_1 on S. The following definition captures precisely what we mean when we say that the function class $\mathcal{F}$ *factors* into these two other classes of functions. If $f_1(x)$ reveals too much about x, it will be difficult to prove linear separability; similarly, if f_2 depends too heavily on x, it will be difficult to prove a bucket has small pseudo-dimension.

Definition 3.1 ((a, b)-factorable class) Consider some $\mathcal{F} = \{f : \mathcal{X} \to \mathbb{R}\}$. Suppose, for each $f \in \mathcal{F}$, there exists (f_1, f_2), $f_1 : \mathcal{X} \to \mathcal{Y}$, $f_2 : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}$ such that $f_2(f_1(x), x) = f(x)$ for every $x \in \mathcal{X}$. Let

$$\mathcal{F}_1 = \{f_1 : (f_1, f_2) \text{ is a decomposition of some } f \in \mathcal{F}\}$$

and

$$\mathcal{F}_2 = \{f_2 : (f_1, f_2) \text{ is a decomposition of some } f \in \mathcal{F}\}.$$

The set $\mathcal{F}$ *(a, b)-factors over Q* if:

(1) $\mathcal{F}_1$ is a-dimensionally linearly separable over $Q \subseteq \mathcal{Y}$.

(2) For every $f_1 \in \mathcal{F}_1$ and sample $S \subset \mathcal{X}$ of size m, the set

$$\mathcal{F}_{2|f_1(S)} = \{f_2' : \mathcal{X} \to \mathbb{R}, f_2'(x) = f_2(f_1'(x), x) | f_1(S) = f_1'(S) \text{ and } (f_1', f_2) \text{ is a decomposition of some } f \in \mathcal{F}\}$$

has pseudo-dimension at most b.

We now give an example of a simple class which satisfies this definition. One could easily bound the pseudo-dimension of this example class using a direct shattering argument, but it will be instructive to work through our definition of (a, b)-separability.

Example 3.2 Fix some set $G = \{g_1, \ldots, g_k\} \subset \mathbb{R}^k$. Suppose $\mathcal{F} = \{f : f(x) = \max_{g \in G_f \subseteq G} g \cdot x\}$ is the set of all functions which take the maximum of at most k common linear functions in a fixed set G. We will show that $\mathcal{F}$ $(kd, \tilde{O}(kd))$-factors over $[k]$, where each $j \in [k]$ will represent *which* of the k linear functions is maximizing for a particular input. That is, for some $f, G_f \subseteq G$, let $f_1(x) = \text{argmax}_{t: g_t \in G_f} g_t \cdot x$ and $f_2(t, x) = g_t \cdot x$. Thus, we have a valid factoring:

$$f_2(f_1(x), x) = f_2(\text{argmax}_{t: g_t \in G_f} g_t \cdot x, x) = g_{\text{argmax}_{t: g_t \in G_f} g_t \cdot x} \cdot x = \max_{g_t \in G_f} g_t \cdot x = f(x).$$

It remains to show that $\mathcal{F}_1$ is d-dimensionally linearly separable and to bound the pseudo-dimension of $\mathcal{F}_{2|f_1}$. We start with the former. Let $\Psi(x, t)_{t'j} = \mathbb{I}[t' = t] \cdot x_j$ for $t' \in [k], j \in [d]$. Then, let $w^f_{tj} = \mathbb{I}[g_t \in G_f] \cdot g_{tj}$. The dot product will then be

$$\Psi(x, t) \cdot w^f = \sum_{t'} \mathbb{I}[t' = t] \cdot \mathbb{I}[g_{t'} \in G_f] g_{t'} \cdot x$$

which will be maximized when $t = \text{argmax}_{t': g_{t'} \in G_f} g_{t'} \cdot x$, or when $t = f_1(x)$. So, $\mathcal{F}_1$ is linearly separable in kd dimensions over $[k]$.

Now, fix $f_1 \in \mathcal{F}_1$; we will show the pseudo-dimension of $\mathcal{F}_{2|f_1}$ is at most $\tilde{O}(kd)$. For any fixed sample $S = (x^1, \ldots, x^m)$, $f_1(x^t)$ is fixed for all $t \in [m]$, implying that the input to all $f_2 \in \mathcal{F}_{2|f_1}$, $(f_1(x^t), x^t)$, is fixed. Finally, by definition of f_2',

$$f_2'(x^t) = f_2(f_1(x^t), x^t) = g_{f_1(x^t)} \cdot x^t.$$

Thus, for each $j \in [k]$, the subset $S_j \subseteq S$ for which $f_1(x^t) = j$ for all $x^t \in S_j$, f_2' is just a linear function in d dimensions of x^t with coefficients g_j, Thus, since linear functions in d dimensions have pseudo-dimension at most $d+1$, there are at most m^{d+1} labelings which can be induced on S_j, and at most $m^{k(d+1)}$ labelings of all of S. This implies $\mathcal{PD}(\mathcal{F}_{2|f_1})$ is at most $\tilde{O}(kd)$.

We now present the main theorem about the pseudo-dimension of classes which are (a, b)-factorable. The proof of this theorem first exploits the fact that linearly separable classes have a "small" number of possible outputs for a sample of size m. Then, fixing the output of the linearly separable function, the second set of functions' pseudo-dimension is small. The proof of the theorem is relegated to the appendix due to space considerations.

Theorem 3.3 *Suppose $\mathcal{F}$ is (a, b)-factorable over Q. Then,*

$$\mathcal{PD}(\mathcal{F}) = O\left(\max\left((a + b)\ln(a + b), a \ln |Q|\right)\right).$$

Intuitively, when $\mathcal{F}_1$ is linearly separable in a dimensions, it can induce at most $m^a |Q|^a$ many labelings of m samples, and fixing such a sample and its labeling, because $\mathcal{F}_2$ has pseudo-dimension at most b, it can induce at most m^b labelings of m samples with respect to their thresholds.

While the range of $\mathcal{F}_1$ might be all of Q, it will regularly be helpful to only need to prove linear separability of $\mathcal{F}_1$ only over "realizable" labels for particular inputs. If $\mathcal{F}_1$ has the property that for every input x, every $f_1 \in \mathcal{F}_1$ labels x with one of a smaller set of labels $Q_x \subsetneq Q$, then it suffices to prove linear separability *for* x over Q_x. The following remark makes this claim formal; its proof can be found in Section E. So, we will be able to focus on proving linear separability of $\mathcal{F}_1$ over a label space which depends upon the inputs x. This will be particularly useful when describing auctions in the next section, whose allocations are in certain cases highly restricted by their inputs.

Remark 3.4 Suppose for each $x \in \mathcal{X}$, there exists some $Q_x \subseteq Q$ such that $f_1(x) \in Q_x \subseteq Q$ for all $f_1 \in \mathcal{F}_1$, and that for for each x, $\mathcal{F}_1$ is linearly separable in a dimensions for that x over Q_x. Assume there is a subset of dimension $T^+ \subseteq [a]$ for which $w^f_{t \in T^+} \geq 0$ and $\sum_{t \in T^+} w^f_t > 0$ for all f. Suppose that for all $x \in X, f \in \mathcal{F}_1$, $\max_{y \in Q_x} \Psi(x, y) \cdot w^f \geq 0$. Then, $\mathcal{F}_1$ is linearly separable over Q in a dimensions as well.

4 Consequences for Learning Simple Auctions

We now present applications of the framework provided by Theorem 3.3 to prove bounds on the pseudo-dimension for many classes of "simple" mechanisms. The implication is that these classes, which have been shown in many special cases to have small *representation error* also have have small *generalization error* when auctions are chosen after observing a polynomially sized sample. We now describe how one can translate a class of auctions into a class of functions which has an obvious and useful factorization. An auction $\mathcal{A} : \mathcal{V}^n \to [n]^k \times [0, H]^n$ has two components, its *allocation function* $\mathcal{A}_1 : \mathcal{V}^n \to [n]^k$ and its *revenue function* $\mathcal{A}_2 : \mathcal{V}^n \to [0, H]^n$. We will abuse notation and refer to $\mathcal{A}_2(\mathbf{v}) = \sum_i A(\mathbf{v})_{2i}$ as the revenue function for an auction. Our goal is to bound the sample complexity of picking some high-revenue function from a class. All omitted proofs are found in Section E. For the remainder of this section we use $\mathcal{F}$ to represent a class of auctions, $f \in \mathcal{F}$ to represent a particular auction, and $\mathcal{F}_1, \mathcal{F}_2$ to be the corresponding allocation functions and revenue functions which result from this decomposition. When $\mathcal{F}_1$ is linearly separable, this implies there can only be so many distinct allocations possible for a fixed set of valuation profiles S, and when $\mathcal{F}_2$ (fixing some allocation for S) has small pseudo-dimension, the class of auctions itself has small pseudo-dimension.

This "trivial" decomposition of an auction's revenue function describes its revenue function as a function of both the allocation chosen by $f_1 \in \mathcal{F}_1$ for $\mathbf{v}$ and the valuation profile $\mathbf{v}$. Since A_2 is a function only of $\mathbf{v}$, there is clearly enough information in $(f_1(\mathbf{v}), \mathbf{v})$ to compute $\mathcal{A}_2(\mathbf{v})$ (one can simply ignore $f_1(\mathbf{v})$ and output $f_2(f_1(\mathbf{v}), \mathbf{v}) = \mathcal{A}_2(\mathbf{v})$). The reason we consider this decomposition is that fixing an allocation, revenue functions of simple auctions are generally very simple to describe as a function of the input valuation profile $\mathbf{v}$. If one fixes the allocation choice for a sample S of m valuations, many auctions' classes of revenue functions are either constant functions on S which do not depend upon $\mathbf{v}$ at all (for example, a posted price auction for a single item offered to a single bidder earns its posted price if the item sells and 0 when the item doesn't sell, both of which are constants when the allocation is fixed) or depends only in a very mild way (for example, a second-price single-item auction with a reserve earns the maximum of its reserve and the second-highest bid when the item sells and 0 when it doesn't).

Most of the "simple" auctions with multiple buyers and items that have been considered are *sequential auctions* which interact with buyers "one at a time": first, bidder 1 is offered a menu of several possible allocations at different prices, she chooses some bundle, then bidder 2 is offered one of several allocations of the remaining items, and so on. We assume for the remainder of the paper that there are no ties, (that is, there are no menus or bidders for which $|\text{argmax}_B u(B)| > 1$.[6] These auctions are simple enough that they can actually be run in practice, and yet expressive enough that in many cases can earn constant fractions of the optimal revenue.

We next work toward a general reduction, from bounding the sample complexity of sequential auctions (with multiple buyers) to that of single-buyer problems. The following definition captures two particularly common forms of these auctions. The first definition captures the setting where the function selecting the menu to bidder i may depend upon i's identity; the second refers to when the menu is *anonymous*: what may be offered to bidder i can be different than what is offered to bidder i', but only due to the differences in bids $v_i, v_{i'}$ and the remaining available items $X_i(v), X_{i'}(v)$.

For example, consider a single item for sale. Suppose n buyers are approached in some fixed order and bidder i is offered the item at price p_i if no earlier buyer has purchased the item. If $p_i = p_{i'}$ for all $i, i' \in [n]$, then the auction applied to each buyer is the same, and we say this auction applies an n-wise repeated allocation associated with a single posted price. If $p_i \neq p_{i'}$ for some $i, i' \in [n]$, then the allocation function applied to each bidder is an allocation rule associated with some single posted price, though the particular posted price and therefore the allocation function varies from bidder to bidder; this auction's allocation is therefore an n-wise sequential allocation drawn from the class of posted prices.

[6] We elide further discussion on this technical point, though we note it is possible to encode a tie-breaking rule over utility-maximizing bundles in a way which is linearly separable (see Hsu et al. (2016) for more details).

For a slightly more complex example, consider a set of k heterogeneous items $[k]$ for sale to n bidders. Consider an auction which sets a price p_{ij} for each item $j \in [k]$ and each bidder $i \in [n]$, and serves bidders in some fixed order. Bidder i is offered any bundle B for which no item has been selected by some previous bidder at price $p_i(B) = \sum_{j \in B} p_{ij}$. This allocation is reached by applying a posted item pricing allocation to each buyer in turn, so these allocations are n-wise sequential allocations drawn from posted item pricing allocations. If $p_{ij} = p_{i'j}$ for all $j \in [k]$ and all $i, i' \in [n]$, then the same allocation rule is being applied to all bidders, and the overall allocation is therefore an n-wise anonymous sequential allocation rule.

Definition 4.1 (**n-fold anonymous and nonanonymous sequential allocations**) Let $\mathcal{H}$ be some class with $h : \mathcal{V} \times \{0,1\}^k \to Q$ for all $h \in \mathcal{H}$ and some $Q \subseteq \{0,1\}^k$. For some n functions $h_1, \ldots, h_n \in \mathcal{H}$ and every $\mathbf{v} \in \mathcal{V}^n$, inductively define $X_1(\mathbf{v}) = [k], X_i(\mathbf{v}) = X_{i-1}(\mathbf{v}) \setminus h_{i-1}(\mathbf{v}_{i-1}, X_{i-1}(\mathbf{v}))$. Then, define the n-wise product function $\prod_{(h_1,\ldots,h_n)}$ to be

$$\prod_{(h_1,\ldots,h_n)} (\mathbf{v}) = (h_1(\mathbf{v}_1, X_1(\mathbf{v})), h_2(\mathbf{v}_2, X_2(\mathbf{v})), \ldots h_n(\mathbf{v}_n, X_n(\mathbf{v}))).$$

Then, we call any such function an n-wise *nonanonymous sequential allocation drawn from $\mathcal{H}$*, or n-wise sequential allocation drawn from $\mathcal{H}$ for short. If $h_1 = h_2 = \ldots = h_n$, we call $\prod_{h_1,\ldots,h_n}$ an n-wise anonymous sequential allocation drawn from $\mathcal{H}$.

The sets $X_1, \ldots, X_n$ correspond to the sets of remaining available items for each bidder after the previous bidders have purchased their bundles according to their allocation functions: what is remaining for bidder i is whatever bidder $i-1$ had available less whatever was allocated to bidder $i-1$. The two previous examples fit into this scenario perfectly. The per-bidder allocation functions are fixed up-front: the allocation rules brought about by (anonymous) a price for a single item or (anonymous) prices for each item. In some fixed order, the bidders are allocated according to their allocation rule run on their valuation and the remaining items, and whatever items they didn't purchase are available for the next bidder and her allocation rule. When the prices don't depend on the index i, the allocation function for each bidder is the same, so those cases correspond to n-wise anonymous sequential allocation rules.

In the event that some class of auctions' allocation functions $\mathcal{F}_1$ are made up of n-wise sequential allocations from a class $\mathcal{H}$ which is linearly separable, the linear separability is imparted upon $\mathcal{F}_1$. This intuition is made formal by the following theorem.

Theorem 4.2 *Suppose $\mathcal{F}$ is a class of auctions, and let $\mathcal{F}_1 : \mathcal{V}^n \to Q \subseteq [n]^k$ be their (feasible) allocation function. Suppose $\mathcal{F}_1$ is a set of n-wise sequential allocations from some $\mathcal{H}$ which is a-dimensionally linearly separable, whose dot products are upper-bounded by H. Then $\mathcal{F}_1$ is an-dimensionally linearly separable. Similarly, if $\mathcal{F}_1$ is a set of n-wise anonymous sequential allocations drawn from $\mathcal{H}$ which is a-dimensionally linearly separable, then $\mathcal{F}_1$ is also a-dimensionally linearly separable.*

Roughly speaking, this proof takes the maps guaranteed by linear separability of $\mathcal{H}$ and concatenates them n times, "blowing up" the relative importance of the earlier bidders with large coefficients.

We now present the three main corollaries of Theorems 3.3 and 4.2 which bound the pseudo-dimension of several auction classes of interest to the mechanism design community. In particular, we focus on "grand bundle" pricings (Corollary 4.3), where each bidder in turn is offered the entire set of items $[k]$ at some price, "item pricings" (Corollary 4.4), where each bidder in turn is offered all remaining items and each item j has some price for purchasing it, and "second-price item auctions with reserves" (Corollary 4.5), where each bidder submits a bid for each item j, and item j is sold to the highest bidder for j at the larger of the item's reserve price and the second-highest bid for that item. Each of these auctions have two versions: the anonymous version, where the relevant design parameters are the same for all bidders, and the non-anonymous version, where those parameters can be bidder-specific. As one would suspect, anonymous pricings have fewer degrees of freedom, and have lower pseudo-dimension. More formally, the allocations which come from anonymous pricings can be formulated as n-wise repeated allocations, while we formulate non-anonymous pricings' allocations as n-wise sequential allocations (which, by Theorem 4.2 loses a factor of n in the upper bound on these classes' pseudo-dimensions). In each case, $\mathcal{F}_1$ will represent allocation functions: $f_1 \in \mathcal{F}_1$ corresponds to the allocation function which the auction will implement for quasilinear bidders. For every class $\mathcal{F}$, we define for every auction $f \in \mathcal{F}$ the function f_2 to be the *revenue* function, which as a function of an allocation and the valuation profile outputs the revenue for that auction with that allocation for that valuation profile. The decomposition of $\mathcal{F}$ into $\mathcal{F}_1, \mathcal{F}_2$ is trivial; the work comes in showing that $\mathcal{F}_1$ is linearly separable and $\mathcal{F}_{2|f_1}$ has small pseudo-dimensions.

Our first two results use the framework to that grand bundle pricings and item pricings have small pseudo-dimension. The second case requires a more delicate treatment of the valuation profiles (buyers are now choosing arbitrary subsets of items, and will choose utility-maximizing bundles based on the per-item prices). It also requires us to consider a larger set of intermediate labels (the set of all possible allocations grows to $[n]^k$ from $[n]$).

Corollary 4.3 *Let $\mathcal{F}$ be the class of anonymous grand bundle pricings. Then,*

$$\mathcal{PD}(\mathcal{F}) = O(1).$$

If $\mathcal{F}$ is the class of non-anonymous grand bundle pricings, then

$$\mathcal{PD}(\mathcal{F}) = O(n \log n).$$

Corollary 4.4 *Let $\mathcal{F}$ be the class of anonymous item pricings. Then,*

$$\mathcal{PD}(\mathcal{F}) = O\left(k^2\right).$$

If $\mathcal{F}$ is the class of nonanonymous item pricings, then

$$\mathcal{PD}(\mathcal{F}) = O\left(nk^2 \ln(n)\right).$$

Finally, we present our final application of this technique, and bound the pseudo-dimension of the class of second-price item auctions with (non-anonymous) item reserves. A result of Yao (2015) implies this class has small representation error for additive buyers; Corollary 4.5 shows it also has small generalization error. We briefly note that we have a slightly tighter bound on this class's pseudo-dimension, using a stylized argument found in Appendix D.

Corollary 4.5 *Suppose $\mathcal{V}$ is some set of additive valuations. Let $\mathcal{F}$ be the class of second-price item auctions with anonymous reserves. Then,*

$$\mathcal{PD}(\mathcal{F}) = O\left(k^2\right).$$

If $\mathcal{F}$ is the class of second-price item auctions with nonanonymous reserves, then

$$\mathcal{PD}(\mathcal{F}) = O\left(nk^2 \ln(n)\right).$$

Proof: We will show that for both classes, $\mathcal{F}_1$ is linearly separable (in k and nk dimensions, respectively). We do this by showing that the anonymous class's' allocations can be described as n-wise anonymous sequential allocations and the nonanonymous class's allocations as n-wise nonanonymous sequential allocations from some class of single-buyer allocation rules $\mathcal{H}$ which is k-dimensionally linearly separable. The natural candidate for $\mathcal{H}$ is the set of allocations defined by item pricings, which we showed in the proof of Corollary 4.4 is linearly separable in k dimensions.

The fact that $\mathcal{F}_1$'s feasible allocations can be described as n-wise anonymous and nonanonymous sequential allocations over (single buyer) item pricings is not immediately obvious: a player's price for an item j is not just the item's reserve price, but the maximum of that reserve and the second-highest bid for j. Thus, a buyer maximizing her quasilinear utility with respect to her reserves would not necessarily purchase the same bundle as when facing item prices which are the larger of her reserve and the second-highest bid for the item, even if she is the highest bidder for each item. However, since the valuations are *additive*, i^*_j the highest bidder for j *will* maximize her utility by purchasing item j if $v_{i^*_j}(j) \geq p_j$, since $v_{i^*_j}(j) \geq \max_{i' \neq i^*_j} v_{i'}(j)$ implies $v_{i^*_j}(j) \geq \max\left(p_j, \max_{i' \neq i^*_j} v_{i'}(j)\right)$, and an *additive* bidder will buy any item j for which her value for that item is (weakly) higher than the price for that item. Thus, the utility-maximizing bundle for some i with respect to her item prices *over the set of bundles for which she only wins items for which she is the highest bidder* will also be utility-maximizing for an additive bidder needing to pay $\max(p_j, \max_{i' \neq i^*_j} v_{i'}(j)$ for each j. Remark 3.4 implies that we need only show linear separability over $Q_\mathbf{v}$ for each $\mathbf{v} \in \mathcal{V}^n$, where $Q_\mathbf{v} = \{f_1(\mathbf{v}) | f_i \in \mathcal{F}_1, \mathbf{v} \in \mathcal{V}^n\}$ is the range that the allocation rules might have for a particular $\mathbf{v}$.

Thus, since this class only sells j to the highest bidder for j, we need only show linear separability over allocations for good j is either unallocated or sold to the highest bidder for j. Thus, the allocation for each bidder i will be correctly predicted by the item pricing linear separator (over the label space which only has highest bidders winning items).

Thus, by Theorem 4.2, the second-price item auctions with anonymous item reserves is linearly separable in $k+1$ dimensions, and with nonanonymous reserves in $n(k+1)$ dimensions. We will now show that for both classes, for any $f \in \mathcal{F}$ and corresponding $f_1 \in \mathcal{F}_1$, the class $\mathcal{F}_{2|f_1}$ has pseudo-dimension $\tilde{O}(k)$ and $\tilde{O}(nk)$, respectively.

First, fix $\mathcal{F}$ to be the set of the second-price item auctions with anonymous item reserves and pick some $f_1 \in \mathcal{F}_1$. Then, for each $f_2' \in \mathcal{F}_{2|f_1}$ and each $\mathbf{v}^t \in S$, we have

$$f_2'(\mathbf{v}^t) = f_2(f_1(\mathbf{v}^t), \mathbf{v}^t) = \sum_j \max(p_j^f, \max_{i' \neq i_j^*} \mathbf{v}_{i'}^t(\{j\})) \cdot f_1(\mathbf{v}^t)_j.$$

For each item j, suppose the relative ordering of p_j^f and $\max_{i' \neq i_j^*} \mathbf{v}_{i'}^t(\{j\})$ were fixed. Then, $f_2'(\mathbf{v}^t)$ is just a *linear* function in k dimensions of $\mathbf{v}^t$ and p^f, which have pseudo-dimension at most $k+1$, and therefore can induce at most m^{k+1} labelings with respect to $(r^1, \ldots, r^m)$. There are $m+1$ possible relative orderings of these parameters, or $(m+1)^k$ for all items simultaneously. Thus, in total, there can be at most $m^{k+1} \cdot (m+1)^k$ labelings of S with respect to $(r^1, \ldots, r^m)$ by $\mathcal{F}_{2|f_1}$, so $\mathcal{PD}(\mathcal{F}_{2|f_1}) = O(k \ln(k))$.

The proof that non-anonymous item reserves has pseudo-dimension $O(nk \ln(nk))$ is analogous, with a few small exceptions. First, we consider those $p^f \in \mathbb{R}^{nk}$ with a fixed ordering of (for all $j \in [k]$) the parameters $\{\mathbf{p}_{ij}^f | i \in [n]\}$ and the set $\{\mathbf{v}_{i_j'}^t(\{j\}) | t \in [m]\}$; there are therefore $\binom{mn}{n}$ of these relative orderings for a fixed item, or $O((mn)^{nk})$ over all items. Fixing this ordering, the revenue on each sample is again a linear function (in nk dimensions) of $\mathbf{v}^t, \mathbf{p}^f$ for each $\mathbf{v}^t$. Thus, $\mathcal{F}_{2|f_1}$ can induce at most $m^{nk+1} \cdot (mn)^{nk}$ many labelings of S w.r.t. $(r^1, \ldots, r^m)$, implying $\mathcal{PD}(\mathcal{F}_{2|f_1}) = O(nk \ln(nk))$.

Then, applying Theorem 3.3 to the two classes (which are $(k, k \log k)$-factorable over $Q \subseteq \{0,1\}^k, |Q| \leq 2^k$ and $(nk, nk \ln(nk))$-separable over $Q \in [n]^k$) implies the pseudo-dimensions are at most $O(k^2)$ and $O(nk^2 \ln(n))$, respectively. ■

References

Martin Anthony and Peter L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, NY, NY, USA, 1999.

Moshe Babaioff, Nicole Immorlica, Brendan Lucier, and S. Matthew Weinberg. A simple and approximately optimal mechanism for an additive buyer. In *Symposium on Foundations of Computer Science (FOCS 2014)*. IEEE – Institute of Electrical and Electronics Engineers, October 2014.

Maria-Florina Balcan, Avrim Blum, and Yishay Mansour. Single price mechanisms for revenue maximization in unlimited supply combinatorial auctions. Technical report, Carnegie Mellon University, 2007.

Maria-Florina Balcan, Avrim Blum, Jason D Hartline, and Yishay Mansour. Reducing mechanism design to algorithm design via machine learning. *Jour. of Comp. and System Sciences*, 74(8):1245–1270, 2008a.

Maria-Florina Balcan, Avrim Blum, and Yishay Mansour. Item pricing for revenue maximization. In *Proceedings of the 9th ACM conference on Electronic commerce*, pages 50–59. ACM, 2008b.

Maria-Florina Balcan, Amit Daniely, Ruta Mehta, Ruth Urner, and Vijay V Vazirani. Learning economic parameters from revealed preferences. In *Web and Internet Economics*, pages 338–353. Springer, 2014.

Tanmoy Chakraborty, Zhiyi Huang, and Sanjeev Khanna. Dynamic and nonuniform pricing strategies for revenue maximization. *SIAM Journal on Computing*, 42(6):2424–2451, 2013.

Shuchi Chawla, Jason Hartline, and Robert Kleinberg. Algorithmic pricing via virtual valuations. In *Proceedings of the 8th ACM Conf. on Electronic Commerce*, pages 243–251, NY, NY, USA, 2007. ACM.

Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the Forty-second ACM Symposium on Theory of Computing*, pages 311–320, NY, NY, USA, 2010. ACM.

Richard Cole and Tim Roughgarden. The sample complexity of revenue maximization. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 243–252, NY, NY, USA, 2014. SIAM.

Amit Daniely and Shai Shalev-Shwartz. Optimal learners for multiclass problems. In *COLT 2014*, pages 287–316, 2014. URL http://arxiv.org/abs/1405.2420.

Nikhil R. Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. The sample complexity of auctions with side information. *CoRR*, abs/1511.02296, 2015. URL http://arxiv.org/abs/1511.02296.

Shaddin Dughmi, Li Han, and Noam Nisan. Sampling and representation complexity of revenue maximization. In *Web and Internet Economics*, volume 8877 of *Lecture Notes in Computer Science*, pages 277–291. Springer Intl. Publishing, Beijing, China, 2014.

Andrzej Ehrenfeucht, David Haussler, Michael Kearns, and Leslie Valiant. A general lower bound on the number of examples needed for learning. *Information and Computation*, 82(3):247–261, 1989. URL https://www.cis.upenn.edu/~mkearns/papers/lower.pdf.

Edith Elkind. Designing and learning optimal finite support auctions. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 736–745. SIAM, 2007.

Michal Feldman, Nick Gravin, and Brendan Lucier. Combinatorial auctions via posted prices. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 123–135. SIAM, 2015.

Steve Hanneke. The optimal sample complexity of PAC learning. *CoRR*, abs/1507.00473, 2015. URL http://arxiv.org/abs/1507.00473.

Jason D. Hartline and Tim Roughgarden. Simple versus optimal mechanisms. In *ACM Conf. on Electronic Commerce*, Stanford, CA, USA., 2009. ACM.

Justin Hsu, Jamie Morgenstern, Ryan Rogers, Aaron Roth, and Rakesh Vohra. Do prices coordinate markets? In *STOC*, page Forthcoming, 1 2016.

Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. Making the most of your samples. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, page forthcoming, New York, NY, USA, 2015. ACM.

Alexander S Kelso Jr and Vincent P Crawford. Job matching, coalition formation, and gross substitutes. *Econometrica: Journal of the Econometric Society*, pages 1483–1504, 1982.

Nick Littlestone and Manfred Warmuth. Relating data compression and learnability. Technical report, University of California, Santa Cruz, 1986. URL https://users.soe.ucsc.edu/~manfred/pubs/lrnk-olivier.pdf.

Andres Munoz Medina and Mehryar Mohri. Learning theory and algorithms for revenue optimization in second price auctions with reserve. In *Proceedings of The 31st Intl. Conf. on Machine Learning*, pages 262–270, 2014.

Jamie H Morgenstern and Tim Roughgarden. On the pseudo-dimension of nearly optimal auctions. In *Advances in Neural Information Processing Systems*, pages 136–144, 2015.

Roger B Myerson. Optimal auction design. *Mathematics of operations research*, 6(1):58–73, 1981.

David Pollard. *Convergence of stochastic processes*. David Pollard, New Haven, Connecticut, 1984.

T. Roughgarden and O. Schrijvers. Ironing in the dark. Submitted, 2015.

Aviad Rubinstein and S. Matthew Weinberg. Simple mechanisms for a subadditive buyer and applications to revenue monotonicity. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, pages 377–394, New York, NY, USA, 2015. ACM.

Vladimir N Vapnik and A Ya Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability & Its Applications*, 16(2):264–280, 1971.

Vladimir Naumovich Vapnik and Samuel Kotz. *Estimation of dependences based on empirical data*. Springer, 1982.

Andrew Chi-Chih Yao. An n-to-1 bidder reduction for multi-item auctions and its applications. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 92–109. SIAM, 2015.

A Open Problems

We propose the following open problems resulting from our work.

1. Is it possible to construct "compression-style" arguments which bound the pseudo-dimension of the revenue of the class of item pricings for additive bidders which are tight (giving a bound of k and nk, as in Theorem D.1, rather than k^2 and nk^2)?

2. For general or even subadditive valuations, do item pricings have pseudo-dimension $O(nk)$ or strictly larger?

3. Is it possible to frame the allocations which result from item pricings with item-specific reserves as n-fold sequential allocation rules from some simple class, for general valuation functions? We were able to show it for additive valuations, which allowed us to use the "trick" where the highest bidder for an item is willing to pay anything less than her bid for that item (independent of other prices); thus, if she's willing to pay the reserve, by virtue of being the highest bidder for the item she's willing to pay the second-highest bid as well. For more general valuations, she may or may not optimize her utility by paying some combination of item prices and second-highest bids for a bundle which was utility-optimal if she were only paying item prices.

4. Relatedly, what is the pseudo-dimension of second-price item auctions with item-specific reserves when bidders have valuations which are more general than additive or unit-demand? One can use a proof similar to the proof of Theorem D.1 to achieve a bound for unit-demand bidders, but what about for submodular or subadditive bidders? It isn't clear that the relative ordering of a small number of "relevant" parameters (such as per-item price and per-bidder single-item values) of the auction and sample are sufficient to fix the most-preferred bundle for each agent from a sample.

B Binary Labeled Learning

Suppose there is some domain $\mathcal{V}$, and let c be some unknown target function $c : \mathcal{V} \to \{0, 1\}$, and some unknown distribution $\mathcal{D}$ over $\mathcal{V}$. We wish to understand how many labeled samples $(v, c(v))$, with $v \sim \mathcal{D}$, are necessary and sufficient to be able to compute a $\hat{c}$ which agrees with c almost everywhere with respect to $\mathcal{D}$. The distribution-independent sample complexity of learning c depends fundamentally on the "complexity" of the set of binary functions $\mathcal{F}$ from which we are choosing $\hat{c}$. We review two standard complexity measures next.

Let N be a set of m samples from $\mathcal{V}$. The set N is said to be *shattered* by $\mathcal{F}$ if, for every subset $T \subseteq N$, there is some $c_T \in \mathcal{F}$ such that $c_T(v) = 1$ if $v \in T$ and $c_T(v') = 0$ if $v' \notin T$. That is, ranging over all $c \in \mathcal{F}$ induces all $2^{|N|}$ possible projections onto N. The *VC dimension* of $\mathcal{F}$, denoted $\mathcal{VC}(\mathcal{F})$, is the size of the largest set S that can be shattered by $\mathcal{F}$.

Let $\text{err}_N(\hat{c}) = (\sum_{v \in N} |c(v) - \hat{c}(v)|)/|N|$ denote the empirical error of $\hat{c}$ on N, and let $\text{err}(\hat{c}) = \mathbb{E}_{v \sim \mathcal{D}}[|c(v) - \hat{c}(v)|]$ denote the *true* expected error of $\hat{c}$ with respect to $\mathcal{D}$. We say $\mathcal{F}$ is (ϵ, δ)-PAC *learnable with sample complexity* m if there exists an algorithm $\mathcal{A}$ such that, for all distributions $\mathcal{D}$ and all target functions $c \in \mathcal{F}$, when $\mathcal{A}$ is given a sample S of size m, it produces some $\hat{c} \in \mathcal{F}$ such that $\text{err}(\hat{c}) < \epsilon$, with probability $1 - \delta$ over the choice of the sample. The PAC sample complexity of a class $\mathcal{F}$ can be bounded as a polynomial function of $\mathcal{VC}(\mathcal{F})$, ϵ, and $\ln \frac{1}{\delta}$ (Vapnik and Chervonenkis, 1971); furthermore, any algorithm which (ϵ, δ)-PAC learns $\mathcal{F}$ over all distributions $\mathcal{D}$ *must* use nearly as many samples to do so. The following theorem states this well-known result formally.[7]

Theorem B.1 (Upper bound (Hanneke, 2015), Lower bound, Corollary 5 of (Ehrenfeucht et al., 1989)) *Suppose $\mathcal{F}$ is a class of binary functions. Then, $\mathcal{F}$ can be (ϵ, δ)-PAC learned with a sample of size*

$$m = O\left(\frac{\mathcal{VC}(\mathcal{F}) + \ln \frac{1}{\delta}}{\epsilon}\right).$$

[7]The upper bound stated here is a quite recent result which removes a $\ln \frac{1}{\epsilon}$ factor from the upper bound; a slightly weaker but long-standing upper bound can be attributed to Vapnik and Kotz (1982).

Furthermore, any (ϵ, δ)-PAC *learning algorithm for* $\mathcal{F}$ *must have sample complexity*

$$m = \Omega\left(\frac{\mathcal{VC}(\mathcal{F}) + \ln\frac{1}{\delta}}{\epsilon}\right).$$

There is a stronger sense in which a class $\mathcal{F}$ can be learned, called *uniform learnability*. This property implies that, with a sufficiently large sample, the error of *every* $c \in \mathcal{F}$ on the sample is close to the true error of c. We say $\mathcal{F}$ is (ϵ, δ)-*uniformly learnable with sample complexity* m if, for every distributions $\mathcal{D}$, given a sample N of size m, with probability $1 - \delta$, $|\text{err}_N(c) - \text{err}(c)| < \epsilon$ for every $c \in \mathcal{F}$. Notice that, if $\mathcal{F}$ is (ϵ, δ)-uniformly learnable with m samples, then it is also (ϵ, δ)-PAC learnable with m samples. We now state a well-known upper bound on the uniform sample complexity of a class as a function of its VC dimension.

Theorem B.2 (See, e.g. Vapnik and Chervonenkis (1971)) *Suppose* $\mathcal{F}$ *is a class of binary functions. Then,* $\mathcal{F}$ *can be* (ϵ, δ)*-uniformly learned with a sample of size*

$$m = O\left(\frac{\mathcal{VC}(\mathcal{F})\ln\frac{1}{\epsilon} + \ln\frac{1}{\delta}}{\epsilon^2}\right).$$

C Formal Statements of Known Revenue Guarantees for Simple Mechanisms

In various special cases, it has been shown that the aforementioned auctions earn a constant fraction of the optimal revenue. All of these results rely on buyers' valuations displaying some kind of independence across items: for additive and unit-demand buyers, this just means that for all i, $v_i = (v_{i1}, \ldots, v_{ik}) \sim \mathcal{D} = \mathcal{D}_1 \times \ldots \mathcal{D}_k$ is drawn from a product distribution. Under this assumption, Chawla et al. (2010) showed that individualized item pricings are sufficient to earn a constant fraction of optimal revenue.

Theorem C.1 *[Chawla et al. (2010)] Suppose each* $i \in [n]$ *has a unit-demand valuation* $v_i \sim \mathcal{D}_i = (\mathcal{D}_{i1} \times \ldots \times \mathcal{D}_{ik})$*. Then, there exists some nonanonymous item pricing* $p \in \mathbb{R}^{kn}$ *such that*

$$\textsc{Rev}(p, \mathcal{D}) \geq \frac{1}{10.67}\textsc{Rev}(\text{OPT}).$$

For a single item-independent additive buyer, the better of the best item pricing and grand bundle pricing also earns a constant fraction of optimal revenue for that setting (Babaioff et al., 2014).

Theorem C.2 *[Babaioff et al. (2014)] Suppose there is a single buyer which has an additive valuation* $v_i \sim \mathcal{D}_i = (\mathcal{D}_{i1} \times \ldots \times \mathcal{D}_{ik})$*. Then, for an item pricing* $p \in \mathbb{R}^k$ *and* q *a grand bundle price,* $q \in \mathbb{R}$

$$\max(\max_{p\in\mathbb{R}^k}\textsc{Rev}(p, \mathcal{D}_i), \max_{q\in\mathbb{R}}\textsc{Rev}(q, \mathcal{D}_i)) \geq \frac{1}{6}\textsc{Rev}(\text{OPT}, \mathcal{D}_i).$$

A recent result of Yao (2015) showed that one can reduce the problem of designing approximately optimal mechanisms for n additive buyers to the problem of designing approximately optimal mechanisms for each single additive buyers, subject to selling each item to the highest bidder for that item (while losing a constant factor in terms of revenue). When combined with the aforementioned result for a single additive bidder, this implies that the best of second price item auctions with the best individualized item reserves and second price grand bundle auctions with the best individualized bundle reserve, is also approximately revenue-optimal for n (non-identically distributed) buyers with valuations which are independent across items.

Theorem C.3 *[Applying Yao (2015) to Babaioff et al. (2014)] Suppose each buyer* $i \in [n]$ *has an additive valuation* $v_i \sim \mathcal{D}_i = (\mathcal{D}_{i1} \times \ldots \times \mathcal{D}_{ik})$*. Let* s_p *for* $p \in \mathbb{R}^{kn}$ *represent the second-price item auction with reserve* p_{ij} *for buyer* i

and item j (and similarly, let s_q for $q \in \mathbb{R}^n$ represent the second-price grand bundle auction with reserve q_i for buyer i). Then,

$$\max(\max_{p \in \mathbb{R}^{kn}} \text{REV}(s_p, \mathcal{D}_i), \max_{q \in \mathbb{R}^n} \text{REV}(s_q, \mathcal{D}_i)) \geq \frac{1}{8} \text{REV}(\text{OPT}, \mathcal{D}_i).$$

The final well-known result for approximately optimal Rubinstein and Weinberg (2015), "simple" revenue-maximizing mechanisms states that, for an appropriately generalized definition of valuations distributed "independently across items", one can approximately maximize revenue selling to a single subadditive buyer with item or grand bundle pricings. We now present the formal definition of independence they use for these more complicated valuation functions, and present their main result.

Definition C.4 (Rubinstein and Weinberg (2015)) A distribution $\mathcal{D}$ over valuation functions $v : 2^k \to \mathbb{R}$ is subadditive over independent items if:

1. All v in the support of $\mathcal{D}$ are monotone; $v(K \cup K') \geq v(K)$ for all K, K'.
2. All v in the support of $\mathcal{D}$ are subadditive: $v(K \cup K') \leq v(K) + v(K')$ for all K, K'.
3. All v in the support of $\mathcal{D}$ exhibit no externalities: there exists some $\mathcal{D}^{\vec{x}}$ over $\mathbb{R}^k$ and a function V such that $\mathcal{D}$ is a distribution that samples $\vec{x} \sim \mathcal{D}^{\vec{x}}$ and outputs v such that $v(K) = V(\{x_\kappa\}_{\kappa \in K}, K)$ for all K.
4. $\mathcal{D}^{\vec{x}}$ is product across its k dimensions.

Theorem C.5 *[Rubinstein and Weinberg (2015)] Suppose $\mathcal{D}_i$ is subadditive over independent items. Then, there exists a universal constant $c \geq 1$ such that*

$$\max(\max_{p \in \mathbb{R}^k} \text{REV}(p, \mathcal{D}_i), \max_{q \in \mathbb{R}} \text{REV}(q, \mathcal{D}_i)) \geq \frac{1}{c} \text{REV}(\text{OPT}, \mathcal{D}_i).$$

D A tighter bound on the pseudo-dimension of second-price item auctions with reserves for additive bidders

We now present a tighter analysis of second-price item auctions with reserves which exploits the total independence of buyers' behavior on items j, j'.

Theorem D.1 *The pseudo-dimension of item auctions and second-price item auctions with anonymous item reserves is $O(k \log k)$ and with nonanonymous item prices/reserves is $O(nk \log(nk))$ when bidders are additive.*

Proof: We present the proof for the class of second-price item auctions with item reserves; the item price result follows easily since the winner for j always pays her item price (rather than the maximum of that and the second-highest bid for j).

Rather than proving the allocation rules are linearly separable, we upper-bound the number of intermediate labelings these classes can induce for m samples, where the intermediate label space we consider is the allocation combined with, for each item, whether the winner for that item is paying the item's reserve or second price for that item. Fix some sample $S = (\mathbf{v}^1, \ldots, \mathbf{v}^m)$ where $v^t \in \mathcal{V}^n$ and $(r^1, \ldots, r^m) \in \mathbb{R}^m$.

This can be encoded in $\{0, 1\}^{2k}$ for anonymous item reserves (a bit for whether or not an item is sold at its reserve and another for whether it is sold for its second-price), and $\{0, 1\}^{2nk}$ for nonanonymous reserves (where each item is labeled as being allocated to some bidder, along with whether it is sold for that bidder's item-specific reserve or the second-highest bid). In the latter case, there is a post-processing rule which can reduce the label space to have size $O(n^{2k})$, since all allocations are feasible allocations. In both cases, we will use y^t to denote the intermediate label for sample $\mathbf{v}^t$.

We begin with anonymous item reserves. Since buyers are additive, we can consider each item separately. We consider item $j \in [k]$. There are $2m + 1$ relevant quantities which affect the revenue any reserve achieves for item j:

p_j, the reserve for j, and for each $t \in [m]$, $v^t_{i^*_j}(\{j\})$ and $v^t_{i'_j}(\{j\})$, where i^*_j, i'_j are the first and second highest bidders for j from sample t, respectively. When $p_j \leq v^t_{i'_j}(\{j\})$, let $y^t_j = 1$ and $y^t_{k+j} = 0$, when $v^t_{i^*_j}(\{j\}) \geq p_j > v^t_{i'_j}(\{j\})$, let $y^t_j = 0$ and $y^t_{k+j} = 1$, and when $p_j > v^t_{i^*_j}(\{j\})$, let $y^t_j = y^t_{n+j} = 0$. Thus, when the relative ordering of these $2m+1$ parameters is fixed, the jth and $n+j$th coordinates for all m samples are fixed. Varying p_j can induce at most $2m+2$ distinct labelings of all of S. Thus, for all k items, there are at most $(2m+2)^k$ distinct vectors $(y^1, \ldots, y^t)$.

Now, fix some intermediate labeling $(y^1, \ldots, y^m)$ of S. Then, the revenue for a particular reserve vector $(p_1, \ldots, p_k)$ which induces this labeling on the sample is easy to describe as a linear of this labeling. Namely,

$$rev(\mathbf{v}^t, \mathbf{p}, y^t) = \sum_{j: y^t_j = 1} v^t_{i'_j}(\{j\}) + \sum_{j: y^t_{n+j} = 1} \mathbf{p}_j$$

which is a linear function in $2k$ dimensions of $\mathbf{p}$ and $\mathbf{v}^t, y^t$ (which are constants). Thus, since linear functions in $2k$ dimensions have VC-dimension $2k+1$, the item reserves which agree with $(y^1, \ldots, y^m)$ can induce at most m^{2k+1} labelings of S with respect to $(r^1, \ldots, r^m)$.

Thus, the set of all item reserves can induce at most $m^{2k+1} \cdot (2m+2)^k$ labelings with respect to $(r^1, \ldots, r^m)$, so if S is shatterable it must be that $2^m \leq m^{2k+1} \cdot (2m+2)^k$, or that $m = O(k \log k)$.

With nonanonymous reserves, each sample will instead be given an intermediate label in $\{0,1\}^{nk+k}$, where there is a bit for each item/bidder pair (corresponding to whether or not that bidder wins the item and pays her individualized reserve for the item), and an additional bit for each item (corresponding to whether or not that item is sold for its second-highest bid). There are at most $[n+1]^k$ valid labelings of a single sample (each item is sold to at most one bidder, and is either sold to her at her reserve or at the second-highest price). For m samples, for a particular item j, there are now $2m+n$ parameters whose ordering matters (the highest and second-highest bids and the bidder-specific reserves for that item); the bidder-specific item reserves for that item can induce at most $(2m+n)^n$ distinct orderings of these parameters; fixing this ordering, the intermediate label is also fixed for all samples. Furthermore, once one has fixed the intermediate label for all samples, the revenues of all individualized item reserve auctions which agree with that intermediate labeling are again expressible as a linear function in $2nk$ dimensions. Thus, if the sample is shatterable, $2^m \leq (2m+n)^{nk} \cdot m^{2nk}$, implying $m = O(nk \ln(nk))$. ■

E Omitted Proofs

Proof of Remark 3.4: For each $x \in X, y \in Q_x$, there exists $\Psi(x,y)$ and for $f \in \mathcal{F}_1$, some $w^f \in \mathbb{R}^d$ such that

$$\text{argmax}_{y \in Q_x} \Psi(x,y) \cdot w^f = f(x).$$

We simply must extend the definition of $\Psi(x,y)$ to be defined over all $y \in Q$

$$\Psi(x,y') \cdot w^f < \max_{y \in Q_x} \Psi(x,y) \cdot w^f$$

for $y' \in Q \setminus Q_x$. Define $\Psi(x,y')_t = 0$ for any $t \notin T^+$, and $\Psi(x,y')_t = -1$ for all $t \in T^+$. Then, for any $y' \in Q \setminus Q_x$, the dot product $\Psi(x,y') \cdot w^f < 0 \geq \max_{y \in Q_x} \Psi(x,y) \cdot w^f$, so the maximizing label y will still be in Q_x. ■

Proof of Theorem 3.3: Consider a sample $S = (x^1, \ldots, x^m) \in \mathcal{X}^m$ of size m with targets $r = (r^1, \ldots, r^m) \in \mathbb{R}^m$. We first claim that, since $\mathcal{F}_1$ is a-dimensionally linearly separable, $\mathcal{F}_1$ can label S in at most $\binom{m}{a} \cdot |Q|^a$ distinct ways. Theorem 2.5 implies that $\mathcal{F}_1$ must admit a compression scheme **compress**, **decompress** of size at most a. Let $f_1(S)$ denote the labeling of all of S by some fixed $f_1 \in \mathcal{F}_1$. Then, $\mathcal{F}_1$ can label S in at most $|\text{range}_{f_1 \in \mathcal{F}_1}(\textbf{decompress} \circ \textbf{compress})(S, f_1(S))|$ ways since this is a compression scheme for $\mathcal{F}_1$. The decompression function takes as input a *labeled* examples which are a subset of S, so it will have one of $\binom{m}{a} \cdot |Q|^a$ inputs for a fixed set S (some subset of S labeled in some arbitrary way), and therefore at most that many outputs, which upper-bounds the total number of possible labelings of S by the same quantity.

Then, fixing the labeling of S to be consistent with some $f_1 \in \mathcal{F}_1$, we know that the pseudo-dimension of $\mathcal{F}_{2|f_1}$ is at most b, so it can induce at most m^b many labelings of S according to r. Thus, there are at most $m^a|Q|^a m^b$ binary labelings of S with respect to r over all of $\mathcal{F}_2$ (and, therefore over all of $\mathcal{F}$). If S is shatterable, it must be that

$$2^m \leq m^a |Q|^a m^b$$

implying $m \leq (a+b)\ln(m) + a \ln |Q|$, as desired. ■

Proof of Theorem 4.2: In either case, Q is a set of feasible allocations, so we only must show linear separability over the set of feasible allocations (that is, we need only show separability over labels $\mathbf{B} : \mathbf{B}_i \cap \mathbf{B}_j = \emptyset$).

We start with the first case of sequential allocations. We will show that $\mathcal{F}_1$ is an-dimensionally linearly separable. By definition, $\mathcal{F}_1$ is a set of n-wise sequential allocations from some $\mathcal{H}$ which is a-dimensionally linearly separable over $\{0,1\}^k$. This means there exists some $\Psi : (\mathcal{V} \times \{0,1\}^k) \times \{0,1\}^k \to \mathbb{R}^a, w^h \in \mathbb{R}^d$ such that $\text{argmax}_B \Psi((v, X), B) \cdot w^h = h(v, X)$ for all $h \in \mathcal{H}, (v, X) \in \mathcal{V} \times \{0,1\}^k$.

We simply need to construct some new $\hat{\Psi} : \mathcal{V}^n \times Q \to \mathbb{R}^{an}, \hat{w}^{h_1,\ldots,h_n} \in \mathbb{R}^{an}$ such that

$$\text{argmax}_{\mathbf{B}=(B_1 \ldots B_n)} \hat{\Psi}(\mathbf{v}, \mathbf{B}) \cdot \hat{w}^{h_1,\ldots,h_n} = (h_1(\mathbf{v}_1, X_1(v)), h_2(\mathbf{v}_2, X_2(\mathbf{v})), \ldots, h_n(\mathbf{v}_n, X_n(\mathbf{v}))).$$

Define $\alpha_i = 2^i H$, and define

$$\Psi((\mathbf{v}, \mathbf{B})_{ij} = \alpha_i \cdot \Psi((\mathbf{v}_i, [k] \setminus \cup_{i'<i} \mathbf{B}_{i'}) \mathbf{B}_i)_j$$

Then, for some $\prod_{(h_1,\ldots,h_n)} \in \mathcal{F}_1$, define

$$\hat{w}_{ij}^{h_1,\ldots,h_n} = w_j^{h_i}$$

Now, inspecting the dot product for some $\mathbf{v}, \mathbf{B}$ we see

$$\hat{\Psi}(\mathbf{v}, \mathbf{B}) \cdot \hat{w}^{h_1,\ldots,h_n} = \sum_i \alpha_i \Psi((\mathbf{v}_i, [k] \setminus \cup_{i'<i} \mathbf{B}_{i'}), \mathbf{B}_i) \cdot w^{h_i}$$

which, by the definition of α_i and the assumption that $\Psi((\mathbf{v}_i, X), B) \cdot w^h \leq H$ for all $\mathbf{v}_i, X, B, h \in \mathcal{H}$ implies that the maximizing label $\mathbf{B}$ will first pick $\mathbf{B}_1 \subseteq [k] = X_1(\mathbf{v})$ to maximize $\Psi((\mathbf{v}_1, X_1(\mathbf{v})), \mathbf{B}_1) \cdot w^{h_1}$, then will pick $\mathbf{B}_2 \subseteq [k] \setminus \mathbf{B}_1 = X_2(\mathbf{v})$ to maximize $\Psi((\mathbf{v}_2, X_2(\mathbf{v})), \mathbf{B}_2) \cdot w^{h_2}$, and so on. Thus, $\mathcal{F}_1$ is an-dimensionally linearly separable.

Now, suppose $\mathcal{F}_1$ is a set of n-wise repeated allocations. Since $\mathcal{H}$ is a-dimensionally linearly separable, we know that for all v, X, B_i, there exists $\Psi((v_i, X), B_i)$, and for all $h \in \mathcal{H}$ there is some w^h such that

$$\text{argmax}_{B_i} \Psi((v_i, X), B_i) \cdot w^h = h(v_i, X).$$

We simply need to define some $\hat{\Psi} : \mathcal{V}^n \times Q \to \mathbb{R}^a$, $\hat{w}^h \in \mathbb{R}^a$ such that

$$\text{argmax}_{\mathbf{B}} \hat{\Psi}(\mathbf{v}, \mathbf{B}) \cdot \hat{w}^h = (h(\mathbf{v}_1, X_1(\mathbf{v})), h(\mathbf{v}_2, X_2(\mathbf{v})), \ldots, h(\mathbf{v}_n, X_n(\mathbf{v}))).$$

Then, define

$$\hat{\Psi}(\mathbf{v}, \mathbf{B})_x = \sum_i \alpha_i \cdot \Psi((\mathbf{v}_i, [k] \setminus \cup_{i'<i} \mathbf{B}_{i'}), \mathbf{B}_i)_x$$

Then, for some $\prod_{h,\ldots,h)} \in \mathcal{F}_1$, define

$$\hat{w}_x^h = w_x^h$$

Then, the dot product

$$\Psi(\mathbf{v}, \mathbf{B}) \cdot \hat{w}^h = \sum_x \sum_i \alpha_i \cdot \Psi((\mathbf{v}_i, [k] \setminus \cup_{i'<i} \mathbf{B}_{i'}), \mathbf{B}_i)_x \cdot w_x^h = \sum_i \alpha_i \cdot \Psi((\mathbf{v}_i, [k] \setminus \cup_{i'<i} \mathbf{B}_{i'}), \mathbf{B}_i) \cdot w^h,$$

which by the definition of α_i and the guaranteed upper bound on the dot product $\Psi \cdot w^h \leq H$, we know will be maximized by first picking some $\mathbf{B}_1 \subseteq [k] = X_1(\mathbf{v})$ which maximizes $\Psi((\mathbf{v}_1, X_1(\mathbf{v})), \mathbf{B}_1) \cdot w^h$, then picking $B_2 \subseteq X_1(\mathbf{v}) \setminus h(\mathbf{v}_1, X_1(\mathbf{v})) = X_2(\mathbf{v})$ which maximizes $\Psi((\mathbf{v}_2, X_2(\mathbf{v})),\ B_2) \cdot w^h$, and so on. Thus, $\mathcal{F}_1$ is a-dimensionally linearly separable. ■

Proof of Corollary 4.3: We first prove first that for a single buyer, the grand-bundle mechanism is 2-dimensionally linearly separable over $\{0, 1\}$. Let $\mathcal{H}$ denote the class of single-buyer grand bundle pricings. For some $h \in \mathcal{H}$, we define $h_1 : \mathcal{V} \to \{0, 1\}$ as $h_1(v) = \mathbb{I}[v \geq p^h]$, where $p^h \in \mathbb{R}$ represents the price of the grand bundle under h. We will show $\mathcal{H}$ is 2-dimensionally linearly separable over $\{0, 1\}$. Define $\Psi(v, b) = \mathbb{I}[b = 1](v([k]), 1)$ for each $b \in \{0, 1\}$ and $w^h = (1, -p^h)$. Then,

$$\text{argmax}_b \Psi(v, b) \cdot w^h = \text{argmax}_b \mathbb{I}[b = 1](v([k] - p^f) = \mathbb{I}[v \geq p^h] = f_1(v)$$

since the penultimate expression is maximized by $b = 1$ only if $v([k]) \geq p^f$. Thus, $\mathcal{H}$ is 2-dimensionally separable.

Notice that when $\mathcal{F}$ is the set of anonymous grand bundle pricings, its allocation rules $\mathcal{F}_1$ are n-fold repeated allocations from $\mathcal{H}$. Thus, by Theorem 4.2, anonymous grand bundle pricings' allocations are linearly separable in 2 dimensions. The obvious intermediate label space $Q = \{\vec{0}\} \cup \{e_i | i \in [n]\}$, the set of standard basis vectors, contains more information than is needed to compute the revenue of these auctions. Define $q(x) = \mathbb{I}[||x|| > 0]$; Observation 1 implies that $\mathcal{F}_1' = q \circ \mathcal{F}_1$ is 2-dimensionally linearly separable over $Q' = \{0, 1\}$. Now we prove for each $f_1 \in \mathcal{F}_1'$ that $\mathcal{F}_{2|f_1}$ has pseudo-dimension $O(1)$. Fix some $f_1 \in \mathcal{F}_1'$. Then, we have that

$$f_2'(\mathbf{v}) = f_2(f_1(\mathbf{v}), \mathbf{v}) = p^f \cdot f_1(\mathbf{v})$$

so, the class $\mathcal{F}_{2|f_1}$ is a class of linear functions in 1 dimensions, which have pseudo-dimension at most 2. Thus, $\mathcal{F}$ is $(2, 2)$-factorable over $\{0, 1\}$, and Theorem 3.3 implies that the pseudo-dimension of anonymous grand bundle pricings is $O(1)$.

Similarly, when $\mathcal{F}$ is the the set of non-anonymous grand bundle pricings, $\mathcal{F}_1$ are n-fold sequential allocations from $\mathcal{H}$. Thus, these allocation rules are $2n$-dimensionally linearly separable, respectively. In this case, we leave the intermediate label space as $Q = \{\vec{0}\} \cup \{e_i | i \in [n]\}$. For any $f \in \mathcal{F}$, let $p^f \in \mathbb{R}^n$ denote the price vector for the grand bundle, that is p_i^f is i's price for purchasing the grand bundle. Fix some $f_1 \in \mathcal{F}_1$; we claim that $\mathcal{F}_{2|f_1}$ has pseudo-dimension $O(n)$. For any $f_2' \in \mathcal{F}_2$ and any f which is decomposed into (f_1, f_2) , we have that $f_2'(\mathbf{v}) = f_2(f_1(\mathbf{v}), \mathbf{v}) = p^f \cdot f_1(v)$, which again is a linear function when f_1 is fixed, in this case in n dimensions. Thus, $\mathcal{F}$ is $(2n, n)$-factorable over Q, so Theorem 3.3 implies that the pseudo-dimension of nonanonymous grand bundle pricings is $O(n \log n)$. ■

Proof of Corollary 4.4: As in the previous proof, we claim that $\mathcal{F}_1$, the allocation rules of these auctions are n-wise repeated allocations and n-wise sequential allocations from the single-buyer item pricings allocation set $\mathcal{H}$. We begin by showing $\mathcal{H}$ is $k + 1$-dimensionally linearly separable. For some $h \in \mathcal{H}$, let $p^h \in \mathbb{R}^n$ denote the item pricing faced by the single buyer. Then, define for $v \in \mathcal{V}$, $B \in \{0, 1\}^k$,

$$\Psi(v, B)_j = \begin{cases} \mathbb{I}[j \in B] & \text{if } j \in [k] \\ v(B) & \text{if } j = k + 1 \end{cases}$$

and for $h \in \mathcal{H}$, define

$$w_j^h = \begin{cases} -p_j^h & \text{if } j \in [k] \\ 1 & \text{if } j = k + 1 \, . \end{cases}$$

Then, we have that $\Psi(v, B) \cdot w^h = v(B) - \sum_{j \in B} p_j^h$, which will be maximized by B which maximizes v's utility. Thus, $h(v) = \text{argmax}_B v(B) - \sum_{j \in B} p_j^h = \text{argmax}_B \Psi(v, B) \cdot w^h$, so $\mathcal{H}$ is $k + 1$-dimensionally linearly separable.

Consider $\mathcal{F}$ the class of anonymous item prices. Theorem 4.2 implies that this class is $k+1$-dimensionally linearly separable over $Q = [n]^k$. Again, the intermediate label space suggested by this reduction to the single-buyer case, $Q = [n]^k$, is larger than necessary to compute revenue. We define $q(\mathbf{B})_j = \mathbb{I}[j \in \cup_i \mathbf{B}_i]$, and by Observation 1, $\mathcal{F}_1' = q \circ \mathcal{F}_1$ is $k + 1$-dimensionally linearly separable over $Q' = \{0, 1\}^k$. We now show that, for a fixed $f_1 \in \mathcal{F}_1$, the class $\mathcal{F}_{2|f_1}$ has pseudo-dimension $O(k)$. Notice that for any $f_2' \in \mathcal{F}_{2|f_1}$, we have that

$$f_2'(\mathbf{v}) = f_2(f_1(\mathbf{v}), \mathbf{v}) = p^f \cdot f_1(v)$$

which, again is a k-dimensional linear function for some fixed f_1, and therefore has pseudo-dimension at most $k+1$. Thus, the class $\mathcal{F}$ can be $(k+1, k+1)$-factored over $\{0,1\}^k$, and Thereof 3.3 implies the pseudo-dimension is thus at most $O(k^2)$.

The proof for the nonanonymous case is identical, with two changes. First, $\mathcal{F}_1$ is a set of n-wise *nonanonymous* sequential allocations, so it is linearly separable in $n(k+1)$ dimensions. Second, we cannot compress the intermediate label space $Q \subset \{0,1\}^{nk}, |Q| \leq [n]^k$, since $f_2'(\mathbf{v}) = f_2(f_1(\mathbf{v}), \mathbf{v}) = p^f \cdot f_1(v)$ only expresses the revenue of the auction if $f_1(v)$ expresses which buyers purchase which items; thus, the set $\mathcal{F}_{2|f_1}$ has pseudo-dimension at most $O(nk)$. Thus, the class $\mathcal{F}$ can be $(O(nk), O(nk))$-factored over Q with $|Q| \leq [n]^k$, and Thereof 3.3 implies the pseudo-dimension is thus at most $O(nk^2 \ln(n))$. ■