Mechanism Design via
Consensus Estimates, Cross Checking, and Profit Extraction

Bach Q. Ha
Northwestern University
bach@u.northwestern.edu

Jason D. Hartline
Northwestern University
hartline@eecs.northwestern.edu

1 Introduction

There is only one technique for *prior-free optimal mechanism design* that generalizes beyond the structurally benevolent setting of digital goods. This technique uses random sampling to estimate the distribution of agent values and then employs the *Bayesian optimal mechanism* for this estimated distribution on the remaining players. Though quite general, even for digital goods, this random sampling auction has a complicated analysis and is known to be suboptimal. To overcome these issues we generalize the consensus technique from [5] to structurally rich environments that include, e.g., single-minded combinatorial auctions.

The classical economic theory of mechanism design is Bayesian: it is assumed that the preferences of the agents are drawn at random from a known probability distribution and the designer aims to optimize their objective in expectation over this randomization. This leads to mechanisms that are tailored to the distributional setting. In contrast prior-free mechanism design looks at mechanisms that perform well without knowledge or assumptions on agent preferences. While these mechanisms do not perform as well as ones tailored to the distribution, in many setting they provide good approximations, and are more robust.

The simplest environment in which to explore mechanism design is that of selling a *digital good*, i.e., where the seller has no constraint over the subsets of agents that can be served simultaneously. Recent contributions to the literature on prior-free mechanism design have focused on extending results for digital good environments to ones that are more structurally rich. From least-general to most-general, these include *multi-unit environments*, where there are a given k number of units available for sale (i.e., any subset of the agents of size at most k can be served); *matching environments*, where feasible sets correspond to one side of a bipartite matching; and *downward-closed environments*, where the only constraint on feasible sets is that any subset of a feasible set is feasible.

The only prior-free mechanisms known to give good approximations for general downward-closed environments are variants of the random sampling auction. This auction first gathers distributional information from a random sample of the agents and then simulates the Bayesian optimal auction for the empirical distribution on the remaining agents. The agents in the sample are always rejected. Tight analysis of the random sampling auction is difficult, upper-bounds on its approximation factor are 4.68, 25, 50, and 2560 for digital good [1], multi-unit [3], matching [6], and downward closed environments [6], respectively. Other mechanism design techniques give 3.25 and 6.5-approximations for digital-good [8] and multi-unit environments [6], respectively, and notably the 3.25 approximation for digital goods surpasses the lower-bound of 4 which is known for

1

the random sampling auction. These limitations suggest the need to consider other techniques for obtaining good approximations for general downward-closed environments.

To obtain good approximation mechanisms for general downward-closed environments, we generalize the digital good auction technique of consensus estimates from [5]. The two main ingredients of this approach are a *profit extraction* mechanism and a *consensus* function. Given a target profit, the profit extraction mechanism should approximate the target if the target is less than the optimal revenue possible. If we had a good estimate of the revenue, we could then obtain a good revenue with the profit extractor. The consensus function is used to get an estimate of the revenue from the reports of the agents in a way that is non-manipulable. In particular, for each agent we can calculate the optimal profit from the other agents, plug this profit into the consensus function, and with high probability the estimated profit produced will be the same for all agents. We can then simulate the profit extraction mechanism for each agent with their consensus estimate. If the estimates agree, the result of this simulation is the agreed-upon profit, otherwise, it is at least zero.

There are two main challenges to extending this approach for general downward-closed environments. The first challenge is in designing a profit-extraction mechanism for these environments. Our profit extraction mechanism will be parameterized by a *revenue curve*, the revenue as a function of number of winners (without taking into account any feasibility constraints). Given a target revenue curve that is below the actual revenue curve, our mechanism obtains revenue comparable to that which would be obtained by the optimal mechanism on the input that corresponds to the target revenue curve. The second challenge is in ensuring infeasible outcomes are not produced in the case that the mechanism does not have a consensus. Note that for digital good environments, there is no feasibility constraint that could be violated when the estimates do not reach consensus. The same is not so for general downward-closed environments. A parameterized mechanism (such as a profit extractor) is of course required to always produced feasible outcomes. However, if we determine the outcome for each agent by simulating the parameterized mechanism with different parameters, the combined outcome may not be feasible. To address this potential inconsistency we give a *cross checking* approach for identifying a subset of agents which which consensus is achieved.

Our mechanism is a 30.4-approximation in general downward-closed environments.

Related Work. This paper derives its framework for prior-free mechanism design and analysis from Hartline and Yan [6]. It extends the profit-extraction and consensus techniques from Goldberg and Hartline [5, 7]. Other than these, the closest related work to ours is that of Dhangwatnotai, Roughgarden, and Yan [4]. They consider abstract service provision in downward-closed environments with the added assumption that the values of the agents are distributed according to a unknown distribution that satisfies a standard *monotone hazard rate* assumption. Under this assumption, they give an (essentially) 4-approximation mechanism. In contrast, our mechanism gives a worse bound, but does so without the distributional assumption.

Organization. In Section 2 we will formally describe our auction environment, design and analysis framework, and review the consensus technique. In Section 3 we describe our cross-checking approach as it applies to obtaining a consistent consensus estimate. In Section 4 we describe a mechanism for extracting the profit suggested by a given target revenue curve. In Section 5 we describe an approach for obtaining a consensus estimate on revenue curves. Finally, in Section 6 we combine the three parts to give a good mechanism and analyze its performance.

2 Preliminaries

Here we describe the abstract setting in which we consider mechanism design and the structural tools that we will be using to design and analyze mechanisms.

Model. Let N be a set of $n \geq 2$ bidders. Each bidder $i \in N$ has a private *valuation* v_i for receiving some abstract service. A bidder i, upon reporting his valuation, will be served with a probability x_i and charged a payment of p_i. We denote the *valuation profile*, *allocation vector*, and *payment vector* by $\mathbf{v} = (v_1, \ldots, v_n)$, $\mathbf{x} = \{x_1, \ldots, x_n\} \in [0,1]^n$, and $\mathbf{p} = (p_1, \ldots, p_n)$ respectively. Without loss of generality we index the agents in decreasing order of value, i.e., $v_i \geq v_{i+1}$.

There is a *feasibility* constraint which describes the subsets of the agents that can be served simultaneously. We assume that this feasibility constraint is downward closed, i.e., any subset of a feasible set is feasible. As described in the introduction, many common environments for mechanism design are downward-closed.

We allow the feasibility constraint to be probabilistic, i.e., given by a convex combination of downward closed set systems. We assume the following semantics for a randomized feasibility constraint: (i) agents bid, (ii) the randomization over the feasibility constraint is realized, and then (iii) the mechanism runs on the reported valuations and the realized set system. For the purpose of calculating revenue the allocation $\mathbf{x}$ and payments $\mathbf{p}$ are taken in expectation over the randomization in the mechanism and the set system. That said, the mechanisms we consider will be incentive compatible even when the order of steps (i) and (ii) are reversed.

A fundamental and potentially restrictive assumption that we will make is one of symmetry. Given an asymmetric set system we can always make it symmetric by randomly permuting the identities of the agents. This assumption is akin to standard assumptions for the secretary problem and in settings where one might consider the agents to be a priori identical (e.g., if there values were drawn from an i.i.d. distribution) it is without loss. The resulting feasibility constraint we refer to as a *downward-closed permutation environment*.

Our mechanisms will be based on simple algorithms. Given weights $\mathbf{w} = (w_1, \ldots, w_n)$ indexed in non-increasing order, we assume we have an algorithm for selecting a feasible set to optimize the sum of the selected weights (for the realized set system) with ties broken randomly. As suggested above, x_i will denote the probability (over randomization in the set system and random tie-breaking) that the ith largest weight is selected. Clearly $\mathbf{x}$ maximizes $\sum_i w_i x_i$ subject to feasibility and so we will refer to $\mathbf{x}$ as the *maximizer* for weights $\mathbf{w}$.

We assume the standard *risk-neutral quasi-linear* utility model, i.e., an agent i wishes to maximize their expected utility which is given by $u_i = v_i x_i - p_i$. We will focus solely on incentive compatible (IC) mechanisms; which means for any agent, reporting their true valuation would be a dominant strategy; and we assume that agents follow this dominant strategy. We view a mechanism as a function from reports to allocation and payments and denote these functions by $\mathbf{x}(\cdot)$ and $\mathbf{p}(\cdot)$. A mechanism is incentive compatible if and only if [9]:

1. $x_i(\mathbf{v})$ is monotone non-decreasing in v_i, and

2. $p_i(\mathbf{v})$ satisfies the payment identity:

$$p_i(\mathbf{v}) = v_i x_i(\mathbf{v}) - \int_0^{v_i} x_i(z, \mathbf{v}_{-i}) \, dz, \tag{1}$$

where $(z, \mathbf{v}_{-i})$ is the valuation profile with v_i replaced with z. The second equation is referred to as the *payment identity*. When we give a mechanism we will describe only the *allocation rule* and infer the *payment rule* from this identity.[1] We will denote the expected payment for agent i with allocation rule $\mathbf{x}(\cdot)$ as $\mathrm{IC}_i^{\mathbf{x}}(\mathbf{v})$ and the total revenue as $\mathrm{IC}^{\mathbf{x}}(\mathbf{v})$.

Revenue Analysis. We adopt the framework from Hartline and Yan [6] wherein the revenue of the designed mechanism is compared to the *envy-free revenue* benchmark. We will denote the maximum envy-free revenue by $\mathrm{EFO}(\mathbf{v})$. For technical reasons we define our benchmark to be $\mathrm{EFO}^{(2)}(\mathbf{v}) = \mathrm{EFO}(v_2, v_2, v_3, \ldots, v_n)$. The goal of such a design and analysis framework is then to give a mechanism that obtains a revenue that is a good approximation to the benchmark $\mathrm{EFO}^{(2)}(\mathbf{v})$ in worst case over valuation profiles $\mathbf{v}$.

Envy-free revenue is defined only for allocation vectors $\mathbf{x}$ that are monotone, i.e., $x_i \geq x_{i+1}$. The envy-free payment for agent i with monotone allocation $\mathbf{x}$ on $\mathbf{v}$ is given by [6]:

$$\mathrm{EF}_i^{\mathbf{x}}(\mathbf{v}) = \sum_{j \geq i}^{n} v_j \cdot (x_j - x_{j+1}) \tag{2}$$

The envy-free revenue, denoted $\mathrm{EF}^{\mathbf{x}}(\mathbf{v})$, is the sum of the envy-free payments.

The envy-free revenue can be understood structurally in terms of the *revenue curve* and *virtual values*. The revenue curve $\mathbf{R}$ for $\mathbf{v}$ describes the optimal revenue as a function of the number of agents served (when feasibility constraints are ignored). The ith coordinate of the revenue curve, R_i, can be calculated by evaluating at i the smallest concave non-decreasing function that contains the point set $\{(i, iv_i) : i \in \{1, \ldots, n\}\}$. The virtual value at i is the left-slope of this function, i.e., $\phi_i = R_i - R_{i-1}$. See Figure 1.

Lemma 1 (Hartline and Yan [6]). *The envy-free revenue of monotone allocation $\mathbf{x}$ satisfies*

$$\mathrm{EF}^{\mathbf{x}}(\mathbf{v}) = \sum_{i=1}^{n} \phi_i \cdot x_i = \sum_{i=1}^{n} R_i \cdot (x_i - x_{i+1}) \tag{3}$$

as long as $x_i = x_{i+1}$ whenever $\phi_i = \phi_{i+1}$.

The optimal envy-free revenue, $\mathrm{EFO}(\mathbf{v})$, can be found from Lemma 1, in particular, by optimizing *virtual surplus*, i.e., $\sum_i \phi_i x_i$, with random tie-breaking. Random tie-breaking results in an allocation $\mathbf{x}$ that satisfies $x_i = x_{i+1}$ whenever $\phi_i = \phi_{i+1}$.

Notice that for the same allocation rule $\mathbf{x}$, the envy-free payments (2) and incentive compatible payments (1) are distinct, i.e., $\mathrm{IC}_i^{\mathbf{x}}(\mathbf{v}) \neq \mathrm{EF}_i^{\mathbf{x}}(\mathbf{v})$.

Consensus estimates. A central ingredient in our approach is the technique of *consensus estimates* that was introduced by Goldberg and Hartline [5]. A *consensus function* maps shared randomness and a statistic to an estimate of the statistic. The objective of such a consensus function is that, when applied individually to each of a set of statistics that are within some bounded range, with high probability (in the shared randomness) the estimates will coincide, i.e., there will be a *consensus* among the *estimates*.

Definition 1. *For implicit parameter $c > 1$ and shared randomness $\sigma \sim U[0,1]$, the* consensus function *on statistic s is,*

$$\mathrm{Consens}(\sigma, s) = \lfloor s \rfloor_{\{c^{\sigma+d} : d \in \mathbb{Z}\}},$$

where $\lfloor s \rfloor_S$ denotes s rounded down to the nearest element of S.

---

[1] Of course such a payment can be easily calculated, e.g., with techniques from Archer et al. [2].
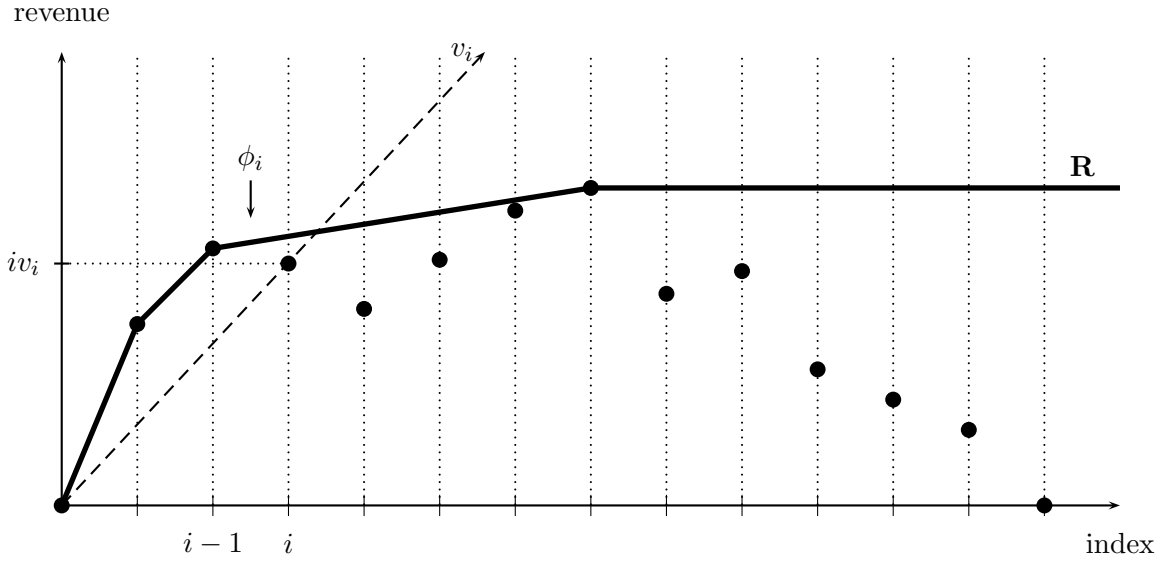
Figure 1: Agents are indexed in decreasing order by value. The point set $\{(i, iv_i) : i \in \{1, \ldots, n\}\}$ is depicted. The revenue curve is the smallest non-decreasing concave function that upper-bounds this point set. The value of agent i can be represented by a line from the origin with slope v_i; the virtual value of agent i is the left-slope of the revenue curve at i.

Lemma 2. *[5] For $c \geq \beta$, The probability (over randomization of σ) that the consensus function is constant on interval $[s/\beta, s]$ is $1 - \log_c \beta$.*

3   Cross-checking

Given some statistic on valuation profiles, we will be using the consensus function (Definition 1) to get an estimate of this statistic, e.g., by calculating $\mathrm{Consens}(\sigma, s(\mathbf{v}_{-i}))$ for each i where $s(\cdot)$ calculates the statistic for a valuation profile. Notice that if we had some mechanism $\mathcal{M}_s$ that was parameterized by statistic s then, if there is consensus, simulations of $\mathcal{M}_{\mathrm{Consens}(\sigma, s(\mathbf{v}_{-i}))}$ to determine the allocation x_i and payment p_i for agent i are internally consistent. I.e., the outcome produced by $\mathcal{M}_s$ is feasible for any s; therefore, so is the combined outcome. Unfortunately, when consensus is not achieved then these simulations may not be consistent.

In this section we give a method of cross-checking to ensure that consistent estimates of the statistic for some subset of the agents. For environments with downward closed feasibility constraints, such a method can be used in mechanism design as agents outside this consistent subset can be rejected.

Definition 2. *For shared randomness σ, statistic s, and consensus function* Consens, *and valuation profile* $\mathbf{v}$ *calculate the following:*

1. For all pairs $i \neq j \in \{1, \ldots, n\}$, calculate the consensus estimate $\tilde{s}_{i,j} = \mathrm{Consens}(\sigma, s(\mathbf{v}_{-i,j}))$.

2. I is the set of agents i that have consensus on $\tilde{s}_{i,j}$ for all j; or $\emptyset$ if no such i exists.

3. $\tilde{s}$ is the consensus $\tilde{s}_{i,j}$ of any $i \in I$ and any j (they are all the same).

The cross-checked consensus function *is defined as*

$$\text{CrossConsens}(\sigma, s, \mathbf{v}) = (\tilde{s}, I).$$

Cross-checked consensus estimates are non-manipulable in a strong sense. Whether or not an agent i is in I is not a function of that agents value. Furthermore, the final consensus is not a function of the report of any agent $i \in I$. This implies that mechanisms which take the following form are incentive compatible.

Definition 3 (cross-checked consensus estimate composition). *Given an incentive compatible mechanism $\mathcal{M}_s$ that is parameterized by some statistic s and a consensus function* Consens *for the statistic, compose them as follows:*

1. *Calculate cross-checked consensus estimate $(\tilde{s}, I) = \text{CrossConsens}(\sigma, s, \mathbf{v})$.*

2. *Simulate incentive compatible mechanism $\mathcal{M}_{\tilde{s}}$ on $\mathbf{v}$.*

3. *For agents $i \in I$ output result of simulation, reject all others.*

Theorem 1. *Mechanisms produced by the cross-checked consensus estimate construction are incentive compatible.*

4 Profit Extraction Mechanism

Using the techniques in the previous section we will construct a consensus estimate for the revenue curve. In this section we will show how to design a mechanism with good revenue that is parameterized by an approximation of the revenue curve. Such a mechanism is termed a *profit extractor*. Given a target revenue curve that is upper-bounded by our actual revenue curve, this mechanism will obtain at least the optimal envy-free revenue for the target revenue curve. The target revenue curve will be provided to the mechanism in the form of the valuation profile $\tilde{\mathbf{v}}$ that generates it. We will denote by $\tilde{\mathbf{R}}$ and $\tilde{\phi}$ the revenue curve and virtual values for $\tilde{\mathbf{v}}$.

Definition 4 (Profit Extractor, $\text{PE}_{\tilde{\mathbf{v}}}$). *Parameterized by non-increasing valuation vector $\tilde{\mathbf{v}}$:*

1. *Sort the bids in a non-increasing order, break ties arbitrary. If $\tilde{v}_i > v_i$ for some i, reject everyone and charge nothing.*

2. *Assign weights $\tilde{\phi}$ to agents in the same order as their values.*

3. *Serve the set of agents to maximize the sum of their assigned weights.*

We will show that the IC revenue obtained from the profit extractor for $\tilde{\mathbf{v}}$ on $\mathbf{v}$ is higher than the optimal envy-free revenue for $\tilde{\mathbf{v}}$. Furthermore, for appropriately chosen $\tilde{\mathbf{v}}$, this revenue approximates the optimal envy-free revenue for $\mathbf{v}$.

Lemma 3. *For any $\tilde{\mathbf{v}} \le \mathbf{v}$, the revenue of the profit extractor on $\mathbf{v}$ is at least the envy-free optimal revenue for $\tilde{\mathbf{v}}$. Moreover, the inequality holds on each agent's payment, i.e., $\text{IC}_i^{\text{PE}_{\tilde{\mathbf{v}}}}(\mathbf{v}) \ge \text{EFO}_i(\tilde{\mathbf{v}})$.*

Proof. We show the second condition of the lemma for any agent i (the first condition follows). Let $\tilde{\mathbf{x}}$ be the allocation for $\text{EFO}(\tilde{\mathbf{v}})$. This is the same allocation as used by $\text{PE}_{\tilde{\mathbf{v}}}$ unless $\mathbf{v} \geq \tilde{\mathbf{v}}$ fails to hold.

First, notice that the IC payments, from equation (1), and EF payments, from equation (2), correspond to the area in the region bounded by $x \leq \tilde{x}_i$ (above), $v \geq 0$ (left), and the "allocation rule" (bottom right). For IC payments, this allocation rule is the probability that the agent is served for any possible misreport z. For EF payments, this "allocation rule" is the smallest monotone function that upper-bounds the point set $\{(\tilde{v}_i, \tilde{x}_i) : i \in \{1, \dots, n\}\}$. To prove the lemma we need only show that the IC allocation rule gives a weaker bound than the EF "allocation rule."

For any $j > i$, the EF allocation rule drops from $\tilde{x}_j$ to $\tilde{x}_{j+1}$ at $\tilde{v}_j$. We claim that the IC allocation rule makes the same drop but at a value that is at least $\tilde{v}_j$. To see this claim, consider the minimum bid z that agent i can make to secure allocation probability at least $\tilde{x}_j$. First, she must out bid agent $j+1$, i.e., $z \geq v_{j+1}$. Assuming she out bids $j+1$, she must also $z \geq \tilde{v}_j$ otherwise, the condition $(z, \mathbf{v}_{-i}) \geq \tilde{\mathbf{v}}$ is not met and all agents are rejected. Therefore, if i bids $z < \tilde{v}_j$ then she is allocated with probability at most $\tilde{x}_{j+1}$. Hence the IC allocation rule drops from $\tilde{x}_j$ to (at most) $\tilde{x}_{j+1}$ at value at least $\tilde{v}_j$. $\square$

Lemma 4. *For any $\tilde{\mathbf{v}}$ and $\mathbf{v}$ with $\tilde{\mathbf{R}} \geq \frac{1}{\beta}\mathbf{R}$, the envy-free optimal revenue for $\tilde{\mathbf{v}}$ is a β-approximation to that from $\mathbf{v}$, i.e., $\text{EFO}(\tilde{\mathbf{v}}) \geq \frac{1}{\beta} \cdot \text{EFO}(\mathbf{v})$.*

Proof. Let $\mathbf{x}$ and $\tilde{\mathbf{x}}$ be the allocations $\text{EFO}(\mathbf{v})$ and $\text{EFO}(\tilde{\mathbf{v}})$, respectively.

$$\text{EFO}(\tilde{\mathbf{v}}) = \sum_i \tilde{R}_i \cdot (\tilde{x}_i - \tilde{x}_{i+1}) \geq \sum_i \tilde{R}_i \cdot (x_i - x_{i+1}) \geq \frac{1}{\beta} \sum_i R_i \cdot (x_i - x_{i+1}) = \frac{1}{\beta}\text{EFO}(\mathbf{v}).$$

The first inequality follows from the optimality of $\tilde{\mathbf{x}}$ for $\tilde{\mathbf{v}}$ and Lemma 1. The second inequality follows from monotonicity of $\mathbf{x}$ and the assumption that $\forall i, \tilde{R}_i \geq \frac{1}{\beta} \cdot R_i$. $\square$

Combining these lemmas, we see that with the right $\tilde{\mathbf{v}}$, $\text{PE}_{\tilde{\mathbf{v}}}(\mathbf{v})$ can approximate the optimal envy free revenue on $\mathbf{v}$.

Theorem 2. *For any $\tilde{\mathbf{v}} \leq \mathbf{v}$ with $\tilde{\mathbf{R}} \geq \frac{1}{\beta}\mathbf{R}$, the profit extractor for $\tilde{\mathbf{v}}$ on $\mathbf{v}$ is a β-approximation to the optimal envy-free revenue for $\mathbf{v}$, i.e., $\text{IC}^{\text{PE}_{\tilde{\mathbf{v}}}}(\mathbf{v}) \geq \frac{1}{\beta} \cdot \text{EFO}(\mathbf{v})$.*

5   Consensus Estimates of Revenue Curves

Our objective now is to get a consensus estimate of the revenue curve. We will express the estimate revenue curve, $\tilde{\mathbf{R}}$, in terms of an estimate valuation profile, $\tilde{\mathbf{v}}$, that generates it.

In addition to the implicit parameter $c > 1$ in the definition of consensus (Definition 1) we will also use implicit parameter $\alpha > 1$ and a minimum required support $m \in \mathbb{Z}_+$. A statistic we will be interested in getting consensus on is the number of agents with values at least α^j for any given j. We will use $n_j(\mathbf{v})$ to denote this statistic. As per our notation in the previous section, we will denote $\tilde{n}_j(\sigma, \mathbf{v}) = \lceil \text{Consens}(\sigma, n_j(\mathbf{v})) \rceil$ where we round the estimate up to the nearest integer because it is an integer statistic. Estimates that do not have the minimum required support of $\tilde{n}_j(\sigma, \mathbf{v}) \geq m$ will be discarded. We will use the remaining estimates to construct an estimate of the valuation profile $\tilde{\mathbf{v}}(\sigma, \mathbf{v})$ and revenue curve $\tilde{\mathbf{R}}(\sigma, \mathbf{v})$ as follows.

Definition 5 (estimated revenue curve and valuation profile)**.** *For any j for which the estimates $\tilde{n}_j$ of the number of agents with values α^j is at least the minimum required support m, define point $Q_j = (\tilde{n}_j, \alpha^j \tilde{n}_j)$. The estimated revenue curve, $\tilde{\mathbf{R}}$, is the minimum non-decreasing concave function that upper-bounds the point set $\{Q_j\}_{j \in \mathbb{Z}}$ and the origin. Let j_k denote the kth largest index such that point Q_{j_k} is on $\tilde{\mathbf{R}}$. The estimated valuation profile, $\tilde{\mathbf{v}}$, has $\tilde{n}_{j_k} - \tilde{n}_{j_{k-1}}$ values equal to α^{j_k}. Pad the remainder of $\tilde{\mathbf{v}}$ with zeros to get an n-vector.*

Figure 2: The dotted vertical lines are $\lceil c^{d+\sigma} \rceil$ for $d \in \mathbb{Z}$; while the diagonal lines from the origin represent value α^j (as their slopes) for $j \in \mathbb{Z}$. Each value v_i is represented with "•" at point (i, iv_i). The line upper-bounding the black dots represents the original revenue curve, $\mathbf{R}$. For each diagonal line representing α^j, $n_j(\mathbf{v})$ is the number of values that lie above it; the dotted vertical line immediately to the left of the right-most among these values represents $\tilde{n}_j(\sigma, \mathbf{v})$; finally the intersection between this line and α^j is the point Q_j, represented with "×". $\tilde{\mathbf{R}}$ and $\tilde{\mathbf{v}}$ are constructed from estimates $\tilde{n}_j(\sigma, \mathbf{v}) \geq m$ and their corresponding Q_js. The thick dotted function upper-bounding these Q_js is $\tilde{\mathbf{R}}$, and each $\tilde{v}_i$ is represented by the point $(i, i\tilde{v}_i)$ with a "□".

In the above construction $\tilde{\mathbf{v}}$ is the smallest (point-wise) valuation profile that has revenue curve $\tilde{\mathbf{R}}$, and furthermore, $\tilde{\mathbf{v}} \leq \mathbf{v}$. For statistical estimates $\tilde{n}_j(\sigma, \mathbf{v})$ the estimated revenue curve and valuation profile will be denoted $\tilde{\mathbf{v}}(\sigma, \mathbf{v})$ and $\tilde{\mathbf{R}}(\sigma, \mathbf{v})$. Our goal now is to show that with high probability the estimated revenue curve (and valuation profile) has consensus when a few agents, S, are omitted, i.e., $\tilde{\mathbf{R}}(\sigma, \mathbf{v}) = \tilde{\mathbf{R}}(\sigma, \mathbf{v}_{-S})$. To do this we define a notion of relevance for statistics j and show that with high probability there is simultaneous consensus for all relevant statistics.

Definition 6 (t-consensus on $\mathbf{v}$)**.** *Given a fixed valuation vector $\mathbf{v}$, a positive integer constant t and a fixed choice of σ, the jth statistic has a t-consensus on $\mathbf{v}$ if for every set $S \subset \{1, 2, \ldots, n\}$ of no more than t elements,*

$$\tilde{n}_j(\sigma, \mathbf{v}) = \tilde{n}_j(\sigma, \mathbf{v}_{-S}).$$

Definition 7. *For a given valuation vector* $\mathbf{v}$ *and a positive integer* t*, the* j*th statistic* n_j *is relevant if there exists* $\sigma \in [0,1]$*, and a set* $S \subseteq \{1, 2, \ldots, n\}$ *of no more than* t *elements such that the point* $Q_j(\sigma, \mathbf{v}_{-S})$ *is on* $\tilde{\mathbf{R}}(\sigma, \mathbf{v}_{-S})$*.*

Notice that when t-consensus happens for a relevant statistic j, then points $Q_j(\sigma, \mathbf{v})$ and $Q_j(\sigma, \mathbf{v}_{-S})$ in the construction of the estimated revenue curve are identical.

We now argue that the probability that any statistic j has does not have consensus is roughly proportional to $1/n_j(\mathbf{v})$, that for relevant statistics the $n_j(\mathbf{v})$ values are geometrically increasing, and thus the union bound implies that all estimates of relevant statistics, and thus the estimated revenue curve, have consensus with high probability. This approach is adapted from Goldberg and Hartline [7].

Lemma 5. *For any* $\mathbf{v}$ *and* $\sigma \sim U[0,1]$*, the probability that the* j*th statistic has a* t*-consensus on* $\mathbf{v}$ *is at least* $1 + \log_c \left(1 - \frac{t}{n_j(\mathbf{v})} \right)$*.*

Proof. Observe that for every set S of no more than t elements, $n_j(\mathbf{v}) - t \leq n_j(\mathbf{v}_{-S}) \leq n_j(\mathbf{v})$. These inequalities hold since when some bids are removed, the number of bids above any α^j decrease, but only by at most the size of the removed set. Thus the probability that $\mathrm{Consens}(\sigma, n_j(\mathbf{v})) = \mathrm{Consens}(\sigma, n_j(\mathbf{v}_{-S}))$ is at least $1 - \log_c \frac{n_j(\mathbf{v})}{n_j(\mathbf{v}) - t}$ as suggested by Lemma 2. The lemma follows from the power rule for logarithm. $\square$

Lemma 6. *The values above successive relevant statistics are bounded by a geometrically increasing function: for any relevant statistic* j*,* $n_j(\mathbf{v}) \geq m\alpha^{r-j}$ *where* r *is the largest index of any relevant statistic.*

Proof. First note that the largest index of a relevant statistic r is well defined. For any j that is relevant, it must be that $n_j(\mathbf{v}) \geq m$; otherwise, j would be discarded by the estimated revenue curve construction. Thus, the largest index that may not be discarded is $\lfloor v_m \rfloor_{\{\alpha_j : j \in \mathbb{Z}\}}$.

From the definition of $\tilde{n}_j(\sigma, \mathbf{v}_{-S})$, we have $\alpha^j \tilde{n}_j(\sigma, \mathbf{v}_{-S}) \leq \alpha^j n_j(\mathbf{v}_{-S}) \leq \alpha^j n_j(\mathbf{v})$. Since j is relevant, the corresponding point $Q_j(\sigma, \mathbf{v}_{-S})$ must be higher than $Q_r(\sigma, \mathbf{v}_{-S})$; therefore, $\alpha^j \tilde{n}_j(\sigma, \mathbf{v}_{-S}) \geq \alpha^r \tilde{n}_r(\sigma, \mathbf{v}_{-S}) \geq m\alpha^r$. Combining this with the previous inequality, we have the desired claim. $\square$

Lemma 7. *The probability of* t*-consensus at all relevant values is at least*

$$1 + \log_c \left[1 - \frac{t\alpha}{m(\alpha - 1)} \right].$$

Proof. We will first bound the probability of consensus at one relevant value, then use the union bound to find the lower-bound of the probability of consensus at all relevant values. For any relevant statistic j, let $\mathcal{E}_j$ denote the event that n_j has a t-consensus on $\mathbf{v}$.

$$\Pr[\mathcal{E}_j] \geq 1 + \log_c \left(1 - \frac{t}{n_j(\mathbf{v})} \right) \geq 1 + \log_c \left(1 - \frac{t}{m}\alpha^{-(r-j)} \right).$$

The first inequality is from Lemma 5, while the second inequality is from Lemma 6. Let $R = \{j \mid n_j \text{ is relevant}\}$. The probability that all relevant statistics have t-consensus on $\mathbf{v}$, using the

union bound, is

$$\Pr[t\text{-consensus}] \geq 1 - \sum\nolimits_{j \in R} \Pr[\neg \mathcal{E}_j] \geq 1 + \sum\nolimits_{j \in R} \log_c \left(1 - \tfrac{t}{m}\alpha^{-(r-j)}\right)$$

$$\geq 1 + \sum\nolimits_{i \geq 0} \log_c \left(1 - \tfrac{t}{m}\alpha^{-i}\right) = 1 + \log_c \left[\prod\nolimits_{i \geq 0} \left(1 - \tfrac{t}{m}\alpha^{-i}\right)\right]$$

$$\geq 1 + \log_c \left[1 - \sum\nolimits_{i \geq 0} \tfrac{t}{m}\alpha^{-i}\right] = 1 + \log_c \left[1 - \tfrac{t}{m}\tfrac{\alpha}{\alpha-1}\right].$$

$\square$

The last thing that we need for our estimated revenue curves is for them to be good estimates. This follows directly from their definition.

Lemma 8. *For any σ and $\mathbf{v}$, the consensus revenue curve $\tilde{\mathbf{R}}(\sigma, \mathbf{v})$ is a $c\alpha$-approximation of the revenue curve $\mathbf{R}^{(m')}$ for $m' = \lfloor mc \rfloor$ and truncated valuation profile $\mathbf{v}^{(m')} = (v_{m'}, \ldots, v_{m'}, v_{m'+1}, v_{m'+2}, \ldots, v_n)$, i.e., $\tilde{\mathbf{R}}(\sigma, \mathbf{v}) \geq \frac{1}{c\alpha}\mathbf{R}^{(m')}$.*

Proof. It is sufficient to show that $\tilde{R}_i \geq \frac{1}{c\alpha}iv_i$ for $i \geq m'$ as concavity of revenue curves would then imply the lemma. Consider then any index $i \geq m'$ and let j be the index of the statistic that satisfies $\alpha^j \leq v_i < \alpha^{j+1}$. Since $n_j(\mathbf{v}) \geq i$, by the definition of $\tilde{n}_j$ and m', respectively, $\tilde{n}_j \geq \lceil i/c \rceil \geq m$; therefore, statistic j is not discarded in the first step of the construction of $\tilde{\mathbf{R}}$. Furthermore, the point $Q_j = (\tilde{n}_j, \alpha^j \tilde{n}_j)$ is above and to the left of $(i, \frac{1}{c\alpha}iv_i)$ because $\tilde{n}_j \leq i$, $\alpha^j \geq v_i/\alpha$, and $\tilde{n}_j \geq i/c$. Monotonicity of $\tilde{\mathbf{R}}$, then, implies the desired $\tilde{R}_i \geq \frac{1}{c\alpha}iv_i$. $\square$

6 Designed Mechanisms

We now proceed to define a mechanism that is a good approximation to $\text{EFO}^{(2)}(\mathbf{v})$. This mechanism will be a convex combination of a primitive cross-checked consensus-estimate profit-extraction mechanism and an extension of the Vickrey auction to downward-closed permutation environments.

Definition 8. *The* primitive cross-checked consensus-estimate profit-extraction *mechanism,* CCEPE$'$ *is the profit extraction mechanism* PE$_{\tilde{\mathbf{v}}}$ *(Definition 4) composed (Definition 2) with the valuation profile estimate (Definition 5).* CCEPE$'$ *is parameterized implicitly by α, c, and m.*

Definition 9. *The Pseudo-Vickrey auction,* Vic, *serves the highest valued agent (and charge her the second highest agent's value) if doing so is feasible with respect to the set system, otherwise, serve no one.*

Definition 10. *The* cross-checked consensus-estimate profit-extraction *mechanism,* CCEPE, *is a convex combination of the Pseudo-Vickrey auction (with probability p) and* CCEPE$'$ *(with probability $1 - p$).* CCEPE$'$ *is parameterized implicitly by α, c, m, and p.*

The Pseudo-Vickrey mechanism is intended to obtain good revenue from the highest-valued agents where as CCEPE$'$ is intended to obtain good revenue from the lower-valued agents. The convex combination obtains good revenue over all. This analysis is given by the following lemmas and theorem.

Lemma 9. *For any* $\mathbf{v}$ *and* $m' = \lfloor mc \rfloor$*, the expected revenue of* CCEPE$'$ *is a* β' *approximation of approximates* EFO$^{(m')}$ *with*

$$\beta' \leq c\alpha \left[1 + \log_c \left(1 - \frac{2\alpha}{m(\alpha-1)} \right) \right]^{-1}.$$

Proof. Let $(\tilde{\mathbf{v}}, I)$ denote the outcome of the cross-checked consensus estimate of the valuation profile. By Lemma 7 (with $t = 2$) all agents are cross-checked and $I = \{1, \ldots, n\}$ with probability at least $1 + \log_c \left[1 - \frac{2\alpha}{m(\alpha-1)} \right]$. In this case, $\tilde{\mathbf{v}} \leq \mathbf{v}$ (Definition 5) and $\tilde{\mathbf{R}} \geq \frac{1}{c\alpha} \mathbf{R}^{(m')}$ (Lemma 8) therefore, the profit-extraction mechanism for $\tilde{\mathbf{v}}$, PE$_{\tilde{\mathbf{v}}}$, on $\mathbf{v}$ obtains revenue at least $\frac{1}{c\alpha}$ EFO$^{(m')}(\mathbf{v})$ (Theorem 2). $\square$

Lemma 10. *For any* $\mathbf{v}$*, the Pseudo-Vickrey auction revenue is at least the envy-free optimal payment of the highest-valued agent, i.e.,* IC$^{\text{Vic}}(\mathbf{v}) \geq$ EFO$_1^{(2)}(\mathbf{v})$*.*

Proof. Assume that $v_1 = v_2$, this is without loss for this lemma because both Pseudo-Vickrey's revenue and EFO$^{(2)}$'s revenue is the same on $\mathbf{v}$ and $\mathbf{v}^{(2)} = (v_2, v_2, v_3, \ldots, v_n)$. The payment for agent 1 upon winning in Pseudo-Vickrey is $v_2 = v_1$; the payment upon winning in EFO$^{(2)}$ is at most $v_1 = v_2$. The probability that agent 1 wins in Pseudo-Vickrey is the highest of any feasible allocation (because agent 1 wins whenever serving agent 1 is feasible); in particular it is as high as that of EFO$^{(2)}$. Therefore, the revenue from agent 1 in Pseudo-Vickrey is at least that of EFO$^{(2)}$. $\square$

Theorem 3. *For any* $\mathbf{v}$*,* CCEPE *is a* β*-approximation to* EFO$^{(2)}(\mathbf{v})$ *where* β *satisfies*

$$\beta \leq \max \left\{ \frac{\lfloor mc \rfloor}{p}, \frac{c\alpha}{1-p} \left[1 + \log_c \left(1 - \frac{2\alpha}{m(\alpha-1)} \right) \right]^{-1} \right\}.$$

Proof. We will separate our revenue into two parts; the first part is obtained from the top $m' = \lfloor mc \rfloor$ agents, denoted $H = \{1, \ldots, m'\}$; and the second part is obtained from the remaining $n - m'$, denoted $L = \{m' + 1, \ldots, n\}$.

The contribution to the envy-free optimal revenue by the top agents satisfies EFO$_H^{(2)}(\mathbf{v}) \leq m' \, \text{IC}^{\text{Vic}}(\mathbf{v})$. This bound follows from Lemma 10 and the observation that envy-free payments are monotonically non-increasing in agent values.

The contribution to the envy-free optimal revenue by the bottom agents satisfies EFO$_L^{(2)}(\mathbf{v}) \leq$ EFO$^{(m')}(\mathbf{v})$. This bound follows as EFO$^{(m')}$ could try to simulate the outcome of EFO$^{(2)}$ and would then receive the same contribution to revenue from agents L as EFO$^{(2)}$; of course, its revenue from all agents must only be higher.

In conclusion, EFO$^{(2)}(\mathbf{v}) \leq m' \, \text{IC}^{\text{Vic}}(\mathbf{v}) + \text{EFO}^{(m')}(\mathbf{v})$.

The revenue of our mechanism, CCEPE, the sum of a $\beta_1 = m'/p$ approximation to $m' \, \text{IC}^{\text{Vic}}(\mathbf{v})$ and a $\beta_2 = \beta'/(1-p)$ approximation to EFO$^{(m')}(\mathbf{v})$, with β' as defined in Lemma 9. Therefore, it is a $\beta = \max(\beta_1, \beta_2)$ approximation to EFO$^{(2)}(\mathbf{v})$. $\square$

We can optimize the parameters of CCEPE to obtain the following corollary.

Corollary 1. *For any* $\mathbf{v}$*,* CCEPE *with* $p = 0.627$*,* $c = 1.666$*,* $\alpha = 2.734$ *and* $m = 12$ *is a* 30.4*-approximation to* EFO$^{(2)}(\mathbf{v})$*.*

References

[1] Saeed Alaei, Azarakhsh Malekian, and Aravind Srinivasan. On random sampling auctions for digital goods. In *Proceedings of the 10th ACM conference on Electronic commerce*, EC '09, pages 187–196, New York, NY, USA, 2009. ACM.

[2] Aaron Archer, Christos Papadimitriou, Kunal Talwar, and Éva Tardos. An approximate truthful mechanism for combinatorial auctions with single parameter agents. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '03, pages 205–214, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.

[3] Nikhil R. Devanur and Jason D. Hartline. Limited and online supply and the bayesian foundations of prior-free mechanism design. In *Proceedings of the 10th ACM conference on Electronic commerce*, EC '09, pages 41–50, New York, NY, USA, 2009. ACM.

[4] Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. Revenue maximization with a single sample. In *Proceedings of the 11th ACM conference on Electronic commerce*, EC '10, pages 129–138, New York, NY, USA, 2010. ACM.

[5] Andrew V. Goldberg and Jason D. Hartline. Competitiveness via consensus. In *Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms*, SODA '03, pages 215–222, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.

[6] Jason Hartline and Qiqi Yan. Envy, truth, and profit. In *Proceedings of the 12th ACM conference on Electronic commerce*, EC '11, pages 243–252, New York, NY, USA, 2011. ACM.

[7] Jason D. Hartline and Andrew V. Goldberg. Collusion-resistant mechanisms for single-parameter agents. In *Proceedings of the 16th Annual ACM-SIAM Sympsium on Discrete Algorithms*, SODA '05, pages 620–629, Vancouver, BC, Canada, January 2005. Society for Industrial and Applied Mathematics.

[8] Jason D. Hartline and Robert McGrew. From optimal limited to unlimited supply auctions. In *Proceedings of the 6th ACM conference on Electronic commerce*, EC '05, pages 175–182, New York, NY, USA, 2005. ACM.

[9] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, February 1981.