2016 IEEE 57th Annual Symposium on Foundations of Computer Science

Learning in Auctions: Regret is Hard, Envy is Easy

Constantinos Daskalakis
CSAIL and EECS, MIT
costis@csail.mit.edu

Vasilis Syrgkanis
Microsoft Research, New England
vasy@microsoft.com

***Abstract*—An extensive body of recent work studies the welfare guarantees of simple and prevalent combinatorial auction formats, such as selling m items via simultaneous second price auctions (SiSPAs) [1], [2], [3]. These guarantees hold even when the auctions are repeatedly executed and the players use no-regret learning algorithms to choose their actions. Unfortunately, off-the-shelf no-regret learning algorithms for these auctions are computationally inefficient as the number of actions available to the players becomes exponential. We show that this obstacle is inevitable: there are no polynomial-time no-regret learning algorithms for SiSPAs, unless $\mathtt{RP} \supseteq \mathtt{NP}$, even when the bidders are unit-demand. Our lower bound raises the question of how good outcomes polynomially-bounded bidders may discover in such auctions.**

To answer this question, we propose a novel concept of learning in auctions, termed "no-envy learning." This notion is founded upon Walrasian equilibrium, and we show that it is both efficiently implementable and results in approximately optimal welfare, even when the bidders have valuations from the broad class of fractionally subadditive (XOS) valuations (assuming demand oracle access to the valuations) or coverage valuations (even without demand oracles). No-envy learning outcomes are a relaxation of no-regret learning outcomes, which maintain their approximate welfare optimality while endowing them with computational tractability.

Our positive and negative results extend to several auction formats that have been studied in the literature via the smoothness paradigm. Our positive results for XOS valuations are enabled by a novel Follow-The-Perturbed-Leader algorithm for settings where the number of experts and states of nature are both infinite, and the payoff function of the learner is non-linear. We show that this algorithm has applications outside of auction settings, establishing significant gains in a recent application of no-regret learning in security games. Our efficient learning result for coverage valuations is based on a novel use of convex rounding schemes and a reduction to online convex optimization.

***Keywords*-online learning, auctions, mechanism design, price of anarchy, computational complexity**

I. Introduction

A central theme in Algorithmic Mechanism Design is understanding the effectiveness and limitations of mechanisms to induce economically efficient outcomes in a computationally efficient manner. A practically relevant and actively studied setting for performing this type of investigation are *combinatorial auctions*.

Supported by a Microsoft Research Faculty Fellowship, and NSF Award CCF-0953960 (CAREER), CCF-1551875 and CCF-1617730. Work done in part while the authors were visiting the Simons Institute for the Theory of Computing.

The setting involves a seller with a set $[m]$ of indivisible items that he wishes to sell to a set $[n]$ of buyers. Each buyer $i \in [n]$ is characterized by a monotone valuation function $v_i : 2^{[m]} \to \mathbb{R}_+$, mapping each bundle S_i of items to the buyer's value $v_i(S_i)$ for this bundle. This function is known to the buyer, but is unknown to the seller and the other buyers. The seller's goal is to find a partition $S_1 \sqcup S_2 \sqcup \ldots \sqcup S_n = [m]$ of the items together with prices $p_1, \ldots, p_n$ so as to maximize the total welfare resulting from allocating bundle S_i to each buyer i and charging him p_i. The total buyer utility from such an allocation would be $\sum_i (v_i(S_i) - p_i)$ and the seller's revenue would be $\sum_i p_i$, so the total welfare from such an allocation would simply be $\sum_i v_i(S_i)$.

Given the seller's uncertainty about the buyer's valuations, she needs to interact with them to select a good allocation. However, the buyers are strategic, aiming to optimize their own utility, $v_i(S_i) - p_i$. Hence, the seller needs to design her allocation and price computation rules carefully so that a good allocation is found despite the agents' being strategic in response to these rules. How much of the optimal welfare can the seller guarantee?

A remarkable result in Economics is that optimal welfare can be attained, as long as we have unbounded computational and communication resources, via the celebrated VCG mechanism [4], [5], [6]. This mechanism asks bidders to report their valuations, uses their reports to select an optimal partition of the items, and computes payments in a way that it is in the best interest of all bidders to truthfully report their valuations; in particular, it is a *dominant strategy truthful* mechanism, and because of its truthfulness it guarantees that an optimal allocation is truly selected.

Despite its optimality and truthfulness, the VCG mechanism is overly demanding in terms of both computation and communication. Reporting the whole valuation function is too expensive for the bidders to do for most interesting types of valuations. Moreover, optimizing welfare exactly with respect to the reported valuations is also difficult in many cases. Unfortunately, if we are only able to do it approximately, the truthfulness of the VCG mechanism disappears, and no welfare guarantees can be made. Even with computational concerns set aside, it is widely acknowledged that the VCG mechanism is rarely used in practice [7]. Indeed, many practical scenarios involve the allocation of items through simple mechanisms which are often not centrally designed and non-truthful. Take eBay, for example, where several

0272-5428/16 $31.00 © 2016 IEEE
DOI 10.1109/FOCS.2016.31

different items are sold simultaneously and sequentially via ascending price and other types of auctions. Or consider sponsored search where several keywords are auctioned simultaneously and sequentially using generalized second price auctions. For most interesting families of valuations such environments induce non truthful behavior, and are thus difficult to study analytically.

The prevalence of such simple decentralized auction environments provides motivation for a quantitative analysis of the quality of outcomes in simple non-truthful mechanisms. A growing volume of research has taken up this challenge, developing tools for studying the welfare guarantees of non-truthful mechanisms; see e.g. [8], [1], [2], [9], [10], [11], [3]. Using the approximation perspective, this literature bounds the Price-of-Anarchy (PoA) of simple non-truthful mechanisms, and has provided remarkable insights into their economic efficiency.

To illustrate these results, let us consider Simultaneous Second Price Auctions, which we will abbreviate to "SiSPAs" in the remainder of this paper. While we focus our attention on these auctions, our results extend to the most common other forms of auctions studied in the PoA literature, as we display in the full version. As implied by its name, a SiSPA asks every bidder to bid on each of the items separately and allocates each item using a second price auction based on the bids submitted solely for this item.

Facing a SiSPA, a bidder whose valuation is non-additive is not able to express his complex preferences over bundles of items. It is thus a priori not clear how he will bid, and what the resulting welfare will be. One situation where a prediction can be made is when the bidders have some information about each other, either knowing each other's valuations, or knowing a distribution from which each other's valuations are drawn. In this case, we can study the SiSPA's Nash or Bayesian Nash equilibrium behavior, computing the welfare in equilibrium. Remarkably, the work on the PoA of mechanisms has shown that the equilibrium welfare of SiSPAs (and of other types of simple auctions) is guaranteed to be within a constant factor of optimum, even when the bidders' valuations are subadditive [3].[1] When bidders have no information about each other, the problem becomes ill-posed, as it is impossible for the bidders to form beliefs about each others bids and choose their bid optimally.

A way out of the conundrum comes from the realization that simple mechanisms often occur repeatedly, involving the same set of bidders; think sponsored search. In such a setting it is natural to assume that bidders engage in learning to compute their new bids as a function of their experience so far. One of the most standard types of learning behavior is *no-regret learning*. A bidder's bids over T executions of a SiSPA satisfy the no-regret learning guarantee if the bidder's cumulative utility over the T executions is within an additive $o(T)$ of the cumulative utility that the bidder would have achieved from the best in hindsight vector of bids $b_1, \ldots, b_m$, if he were to place the same bid b_j on item j in all T executions of the SiSPA. Assuming that bidders use no-regret learning to update their bids in repeated executions of a SiSPA (or other types of simple auctions) the afore-referenced work has shown that the average bidder welfare across the T executions is within a constant factor of the optimal welfare, even when the bidders' valuations are subadditive [3].

These guarantees are astounding, especially given known intractability results for dominant strategy truthful mechanisms, which hold even when the bidders have submodular valuations [12], [13], [14]—a family of valuations that is smaller than subadditive.[2] However, moving to simple non-truthful auctions does not come without a cost. Cai and Papadimitriou [15] have recently established intractability results for computing Bayesian-Nash equilibria in SiSPAs, even for quite simple types of valuations, namely mixtures of additive and unit-demand [15].[3] At the same time, implementing no-regret learning in combinatorial auctions is quite tricky as the action space of the bidders explodes. For example, in SiSPAs there is a continuum of possible bid vectors that a bidder may submit and, even if we tried to discretize this set, their number would generally be exponential in the number of items in order to maintain a good approximation from the discretization. Unfortunately, every step of a no-regret algorithm would typically require computation that is linear in the number of available actions, thus exponential in the number of items in our setting.

An important open question in the literature has thus been whether this obstacle can be overcome via specialized no-regret algorithms that only need polynomial computation. Our first result shows that this obstacle is insurmountable. We show that in one of the most basic settings where no-regret learning is non-trivial, it cannot be implemented in polynomial-time unless $\texttt{RP} \supseteq \texttt{NP}$.

Theorem 1. *Suppose that a unit-demand bidder whose value for each item $i \in [m]$ is v bids in T executions of a SiSPA. Unless $\texttt{RP} \supseteq \texttt{NP}$, there is no learning algorithm running in time polynomial in m, v, and T and whose regret is any polynomial in m, v, and $\frac{1}{T}$. The computational hardness holds even when the learner faces i.i.d. samples from a fixed distribution of competing bids, and whether or not no-overbidding is required of the bids produced by the learner.*

Note that our theorem proves an intractability result even if pseudo-polynomial dependence on the description of v is permitted in the regret bound and the running time. The *no-*

[1] A *subadditive* valuation v is one satisfying $v(S \cup T) \leq v(S) + v(T)$, for all $S, T \subseteq [m]$.

[2] A *submodular* valuation v is one satisfying $v(S \cup T) + v(S \cap T) \leq v(S) + v(T)$, for all $S, T \subseteq [m]$.

[3] A *unit-demand* valuation v is one satisfying $v(S) = \max_{i \in S} v(\{i\})$, for all $S \subseteq [m]$.

overbidding assumption mentioned in the statement of our theorem represents a collection of conditions under which no-regret learning in second-price auctions gives good welfare guarantees [1], [3]. An example of such no-overbidding condition is this: For each subset $S \subseteq [m]$, the sum of bids across items in S does not exceed the bidder's value for bundle S. Sometimes this condition is only required to hold on average. It will be clear that our hardness easily applies whether or not no-overbidding is imposed on the learner, so we do not dwell on this issue more in this paper.

How can we show the in-existence of computationally efficient no-regret learning algorithms? A crucial (and general) connection that we establish in this paper is that it suffices to prove an inapproximability result for a corresponding offline combinatorial optimization problem. More precisely, we prove Theorem 1 by establishing an inapproximability result for an offline optimization problem related to SiSPAs, together with a "transfer theorem" that transfers inapproximability from the offline problem to intractability for the online problem. The transfer theorem is a generic statement applicable to any online learning setting. In particular, we show the following—see Section III for details:

1) In SiSPAs, finding the best response payoff against a polynomial-size supported distribution of opponent bids is strongly NP-hard to additively approximate for a unit-demand bidder. Another way to say this is that one step of a specific learning algorithm, namely Follow-The-Leader (FTL), is inapproximable.

2) If it is intractable to additively approximate the optimum of the average function of a given, efficiently samplable distribution over functions in some set $\mathcal{F}$, then there exists no efficient no-regret online algorithm against sequences of functions from $\mathcal{F}$, unless $\mathtt{RP} \supseteq \mathtt{NP}$. This result is generic: whenever one step of FTL is inapproximable, there is no efficient no-regret learning algorithm. Independent work of [16], provides a weaker reduction, which requires hardness for explicit distributions, rather than efficiently samplable.

The intractability result of Theorem 1 shows that computationally bounded learners cannot reach no-regret outcomes in SiSPAs. Our intractability results can be easily adapted to Simultaneous First Price Auctions and, we expect, several other commonly studied mechanisms for which PoA bounds on the welfare guarantees of no-regret learning outcomes are known. Complementing our results, recent work of Braverman et al. [17] shows that, in a large class of auctions where no-regret learning *is* efficiently implementable, no better than the logarithmic approximation of [18] to the optimal welfare is attainable.

The afore-described impossibility results raise the question of what welfare we should expect of SiSPAs, or other types of combinatorial auctions, when bidders are computationally bounded. We propose a way to overcome these barriers by introducing a new type of learning dynamics, which we call *no-envy*, and which is founded upon the concept of Walrasian equilibrium. In all our results, no-envy learning outcomes are a super-set of no-regret learning outcomes. We show that this super-set simultaneously achieves two important properties: i) while being a broader set, it still maintains the welfare guarantees of the set of no-regret learning outcomes established via PoA analysis; ii) there exist computationally efficient no-envy learning algorithms; when these algorithms are used by the bidders, their joint behavior converges (in a decentralized manner) to the set of no-envy learning outcomes for a large class of valuations (which includes submodular). Thus no-envy learning provides a way to overcome the computational intractability of no-regret learning in auctions with implicitly given exponential action spaces. We describe our results in the following section. We will focus our attention on SiSPAs but the definition of no-envy learning naturally extends to any mechanism and all our positive results extend to a large class of smooth mechanisms (see full version).

A. No-Envy Dynamics: Computation and Welfare

No-envy dynamics is a twist to no-regret dynamics. Recall that in no-regret dynamics the requirement is that the cumulative utility of the bidder after T rounds be within an additive $o(T)$ error of the optimum utility he would have achieved had he played the best fixed bid in hindsight. In no-envy dynamics, we require that the bidder's cumulative utility be within an additive $o(T)$ of the optimum utility that he would have achieved if he was allocated the best in hindsight fixed bundle of items in all rounds and paid the price of this bundle in each round. The guarantee is inspired by Walrasian equilibrium: In auctions, the prices that a bidder faces on each bundle of items is determined by the bids of the other bidders. Viewed as a price-taker, the bidder would want to achieve utility at least as large as the one he would have achieved if he purchased his favorite bundle at its price. No-envy dynamics require that the average utility of the bidder across T rounds is within $o(1)$ of what he would have achieved by purchasing the optimal bundle at its average price in hindsight.

Inspired by Walrasian equilibrium, no-envy learning defines a natural benchmark against which to evaluate an online sequence of bids. It is easy to see that in SiSPAs the no-regret learning requirement is stronger than the no-envy learning requirement. Indeed, the no-envy requirement is implied by the no-regret requirement against a subset of all possible bid vectors, namely those in $\{0, +\infty\}^m$. So no-envy learning is more permissive than no-regret learning, allowing for a broader set of outcomes. This not true necessarily for other auction formats, but it holds for the types of valuation functions and auctions studied in this paper. In particular, in all our no-envy learning upper bounds the set of outcomes reachable via no-envy dynamics is always a superset of the outcomes reachable via no-regret dynamics. This is true even if the no-envy dynamics are constrained to not overbid.

While no-regret learning outcomes are intractable, we show that this broader set of outcomes is tractable. At the same time, we show that this broader set of outcomes maintains approximate welfare optimality. So we have increased the set of possible outcomes, but maintained their economic efficiency and endowed them with computational efficiency.

We proceed to describe our results for the computational and economic efficiency of no-envy learning. First, observe that even though the no-envy learning guarantee is a relaxation of the no-regret learning guarantee, the problem of implementing no-envy learning sequences remains similarly challenging. Take SiSPAs, for example. As we have noted no-envy learning is tantamount to requiring the bidder to not have regret against all bid vectors in $\{0, +\infty\}^m$. This set is exponential in the number of items m, so it is unclear how to run an off-the-shelf no-regret learner efficiently. We are still suffering from the combinatorial explosion in the number of actions, which lead to our lower bound of Theorem 1. Yet the curse of dimensionality is now much more benign. Our upper bounds, discussed next, establish that we can harness big computational savings when we move from competing against any bid vector to competing against bid vectors in $\{0, +\infty\}^m$. Except to do this we still need to develop new general-purpose, no-regret algorithms for online learning settings where the number of experts is exponentially large and the cost/utility functions are arbitrary.

1) Efficient No-Envy Learning: We show that no-envy learning can be efficiently attained for bidders with fractionally subadditive (XOS) valuations. A valuation $v(\cdot)$ belongs to this family if for some collection of vectors $\mathcal{V} = (v^\ell)_\ell$, where each $v^\ell \in \mathbb{R}^m_+$, it satisfies:

$$v(S) = \max_{v^\ell \in \mathcal{V}} \sum_{j \in S} v^\ell_j, \forall S \subseteq [m]. \quad (1)$$

Note that the XOS class is larger than that of submodular valuations. In many applications, the set $\mathcal{V}$ describing an XOS valuation may be large. Thus instead of inputting this set explicitly into our algorithms, we will assume that we are given an oracle, which given S returns the vector $v^\ell \in \mathcal{V}$ such that $v(S) = \sum_{j \in S} v^\ell_j$. Such an oracle is known as an *XOS oracle* [19], [20]. We will also sometimes assume, as it is customary in Walrasian equilibrium, that we are given access to a *demand oracle*, which given a price vector $p \in \mathbb{R}^m_+$ returns the bundle S maximizing $v(S) - \sum_{j \in S} p_j$. We show the following.

Theorem 2. *Consider a bidder with an XOS valuation $v(\cdot)$ participating in a sequence of SiSPAs. Assuming access to a demand and an XOS oracle for $v(\cdot)$,[4] there exists a polynomial-time algorithm for computing the bidder's bid vector b^t at every time step t such that after T iterations the bidder's average expected utility is at least:*

$$\max_S \left(v(S) - \sum_{j \in S} \hat{\theta}^T_j \right) - O\left(\frac{m^2(D+H)}{\sqrt{T}} \right), \quad (2)$$

where $\hat{\theta}^T_j$ is the average cost of item j in the T executions of the SiSPA as defined by the bids of the competing bidders, D is an upper bound on the competing bid for any item and H is an upper bound on $\max_S v(S)$. The learning algorithm with the above guarantee also satisfies the no overbidding condition that the sum of bids for any set of items is never larger than the bidder's value for that set. Moreover, the guarantee holds with no assumption about the behavior of competing bidders.

The proof of Theorem 2 is carried out in three steps, of which the first and last are specific to SiSPAs, while the second provides a general-purpose Follow-The-Perturbed-Leader (FTPL) algorithm in online learning settings where the number of experts is exponentially large and the cost/utility functions are arbitrary:

1) The first ingredient is simple, using the XOS oracle to reduce no-envy learning in SiSPAs to no-regret learning in a related "online buyer's problem," where the learner's actions are not bid vectors but instead what bundle to purchase before seeing the prices; see Definition 6. Theorem 6 provides the reduction from no-envy learning to this problem.

2) The second step proposes a FTPL algorithm for general online learning problems where the learner chooses some action $a^t \in A$ and the environment chooses some state $\theta^t \in \Theta$, *from possibly infinite, unstructured sets A and* Θ, and where the learner's reward is tied to these choices through some function $u(a^t, \theta^t)$ *that need not be linear.* Since A need not have finite-dimensional representation and u need not be linear, we cannot efficiently perturb (either explicitly or implicitly) the cumulative rewards of the elements in A as required in each step of FTPL [22]; see [23] and its references for an overview of such approaches. Instead of perturbing the cumulative rewards of actions in A directly, our proposal is to do this indirectly by augmenting the history $\theta^1, \ldots, \theta^{t-1}$ that the learner has experienced so far with some randomly chosen fake history, and run Follow-The-Leader (FTL) after this augmentation. While it is not a priori clear whether our perturbation approach is a useful one, it is clear that our proposed algorithm only needs an offline optimization oracle to be implemented, as each step is an FTL step after the fake history is added. When applying this algorithm to the online buyer's problem from Step 1, the required offline optimization oracle will conveniently end up being simply a demand oracle.

Our proposed general purpose learner is presented in Section V. The way our learner accesses function u is via an optimization oracle, which given a finite multiset of elements from Θ outputs an action in A that is optimal against the uniform distribution over the multiset. See Definition 7.

[4]For submodular valuations this is equivalent to assuming access only to demand oracles, as XOS oracles can be simulated in polynomial time assuming demand oracles [21]

In Theorem 8, we bound the regret experienced by our algorithm in terms of u's stability. Roughly speaking, the goal of our randomized augmentations of the history in each step of our learning algorithm is to smear the output of the optimization oracle applied to the augmented sequence over A, allowing us to couple the choices of Be-The-Perturbed-Leader and Follow-The-Perturbed-Leader for that sequence. **3)** To apply our general purpose algorithm from Theorem 8 to the online buyer's problem for SiSPAs from Step 1, we need to bound the stability of the bidder's utility function subject to a good choice of a history augmentation sampler. It turns out there is a simple sampler for us to use here, where only one price vector is added to the history, whose prices are independently distributed according to an exponential distribution with mean $O(\sqrt{T})$ and variance $O(T)$. **4)** While our motivation comes from mechanism design, our FTPL algorithm from Step 2 is general purpose, and we believe it will find applications in other settings. For instance, in the full version we show that it provides quantitative improvements on the regret bounds of a recent paper of Balcan et al. [24] for security games.

In the absence of demand oracles, we provide positive results for the subclass of XOS called coverage valuations. To explain these valuations, consider a bidder with k needs, $1, \ldots, k$, associated with values $a_1, \ldots, a_k$. There are m available items, each covering a subset of these needs. We can view each item as a set $\beta_i \subseteq [k]$ of the needs it satisfies. The value that the bidder derives from a set $S \subseteq [m]$ of the items is the total value from the needs that are covered:

$$v(S) = \textstyle\sum_{\ell \in \cup_j \beta_j} a_\ell. \tag{3}$$

Theorem 3. *Consider a bidder with an explicitly given coverage valuation $v(\cdot)$ participating in a sequence of SiSPAs. There exists a polynomial-time algorithm for computing the bidder's bid vector b^t at every time step t such that after T iterations the bidder's average expected utility is at least:*

$$\max_S \left(\left(1 - \frac{1}{e}\right) v(S) - \sum_{j \in S} \hat{\theta}_j^T \right) - 3m \frac{H + \sqrt{D}}{\sqrt{T}}, \tag{4}$$

where $\hat{\theta}_j^T$, H and D are as in Theorem 2, and the algorithm satisfies the same no overbidding condition stated in that theorem. There is no assumption about the behavior of the competing bidders.

Note that our no-envy guarantee (4) in Theorem 3 has incurred a loss of a factor of $1-1/e$ in front of $v(S)$, compared to the guarantee in (2). This relaxed guarantee is an even broader relaxation of the no-regret guarantee. Still, as we show in the next section this does not affect our approximate welfare guarantees. We prove Theorem 3 via an interesting connection between the online buyer's problem for coverage valuations and the convex rounding approach for truthful welfare maximization proposed by [25]. In the online buyer's problem, the buyer needs to decide what set to buy at each step, prior to seeing the prices. It is natural to have the buyer include each item to his set independently, thereby defining an expert for all points $x \in [0,1]^m$, where x_i is the probability that item i is included. It turns out that the expected utility of the buyer under such distribution x is not necessarily convex, so this choice of experts turns our online learning problem non-convex. We propose to massage each expert $x \in [0,1]^m$ into a distribution $D(x) \in \Delta(2^{[m]})$. With the right choice of $D(\cdot)$ the expected value function becomes convex in x, so we may run online convex optimization algorithms on the massaged experts.

2) Welfare Maximization: We show that SiSPAs attain a constant factor approximation to the optimal welfare at every no-envy or approximate no-envy learning outcome for the broad class of XOS valuations. Thus the relaxation from no-regret to no-envy learning does not degrade the quality of the welfare guarantees, and has the added benefit that no-envy outcomes can be attained by computationally bounded players in a decentralized manner, using Theorems 2 and 3. In the full version, we show that this property applies to many mechanisms that have been analyzed in the literature via the *smoothness* paradigm [11].

Corollary 4. *When each bidder $i \in \{1, \ldots, n\}$ participating in a sequence of SiSPAs has an XOS valuation (endowed with a demand and XOS oracle) or an explicitly given coverage valuation $v_i(\cdot)$, there exists a polynomial-time computable learning algorithm such that, if each bidder i employs this algorithm to compute his bids b_i^t at each step t, then after T rounds the expected average welfare is guaranteed to be at least:*

$$\frac{1}{2}\left(1 - \frac{1}{e}\right) \textsc{Opt}(v_1, \ldots, v_n) - O\left(m^2 \cdot n \cdot \max_{S,i} v_i(S) \sqrt{\frac{1}{T}}\right).$$

If all bidders have XOS valuations with demand and XOS oracles the factor in front of OPT is $1/2$.

We regard Corollary 4, in particular our result for XOS valuations with demand queries, as going a long way to alleviate our intractability results for no-regret learning. It also offers a new approach to mechanism design, namely *mechanism design for no-envy bidders.* While only super-constant approximation factors to the optimal welfare are known for dominant strategy truthful mechanisms, even for submodular bidders [12], [13], [14], [26], we attain constant factor approximations for the larger class of XOS valuations, albeit with the different solution concept of no-envy learning. Inded, we regard no-envy learning as a fruitful new approach to mechanism design going forward.

II. Preliminaries

We analyze the online learning problem that a bidder faces when participating in a sequence of repeated executions of a simultaneous second price auction (SiSPA) with m items.

While we focus on SiSPAs our results extend to the most commonly studied formats of simple auctions, as discussed in the full version. A sequence of repeated executions of a SiSPA corresponds to a sequence of repeated executions of a game involving n players (bidders). At each execution t, each player i submits a bid b_{ij}^t on each item j. We denote by b_i^t the vector of bidder i's bids at time t and by b^t the profile of bids of all players on all items. Given these bids, each item is given to the bidder who bids for it the most and this bidder pays the second highest bid on the item. Ties are broken according to some arbitrary tie-breaking rule. Each player i has some fixed (across executions) valuation $v_i : 2^{[m]} \to \mathbb{R}_+$ over bundles of items. If at time t he ends up winning a set of items S^t and is asked to pay a price of θ_j^t for each item $j \in S^t$, then his utility is $v_i(S^t) - \sum_{j \in S^t} \theta_j^t$, i.e. his utility is assumed quasi-linear. An important class of valuations that we will consider in this paper is that of XOS valuations, defined in Equation (1), which are a super-set of submodular valuations but a subset of subadditive valuations. We will also consider the class of coverage valuations, defined in Equation 3, which are a subset of XOS. Different results will consider different types of access to an XOS valuation through an XOS oracle, a demand oracle, or a value oracle, as described in the introduction (see also [21]).

Online bidding problem: From the perspective of a single player i, all information that he needs in order to calculate his utility as a function of his bid in a SiSPA is the highest bid submitted by the other bidders on each item j, as well as the probability that he wins each item j if he ties with the highest other bid on that item. For simplicity of notation, we will assume throughout the paper that the player always loses an item when he ties first. All our results, both positive and negative, easily extend to the more general case of arbitrary bid-profile dependent tie-breaking. Since we analyze learning from the perspective of a single player, we drop the index of player i. For a fixed bid profile of the opponents, we refer to the highest other bid on item j as the threshold of item j and denote it with θ_j. We denote with $\theta = (\theta_1, \ldots, \theta_m)$. The player wins an item j if he submits a bid $b_j > \theta_j$ and loses the item otherwise. When he wins item j, he pays θ_j. We are interested in learning algorithms that achieve a no-regret guarantee even when the thresholds of the items are decided as is customary by an adversary. Thus, the online learning problem that a player faces in a SiSPA is defined as follows:

Definition 1 (Online bidding problem)**.** *At each execution/day/time/step t, the player picks a distribution over bid vectors $b^t \in [0, B]^m$ and the adversary picks adaptively (based on the history of the player's past bid vector distributions (including round t) and bid realizations (excluding round t)) a threshold vector $\theta^t \in [0, H]^m$. The player wins set $S(b^t, \theta^t) = \{j \in [m] : b_j^t > \theta_j^t\}$ and gets reward:*

$$u(b^t, \theta^t) = v\left(S\left(b^t, \theta^t\right)\right) - \sum_{j \in S(b^t, \theta^t)} \theta_j^t. \tag{5}$$

At each t, the distribution over b^t chosen by the algorithm may depend on the history of threshold vectors $\theta_{<t}$.

We evaluate learning algorithms based on their *regret* with respect to the best fixed bid vector in hindsight.

Definition 2 (Regret)**.** *The regret of an online learning algorithm against an adaptive adversary generating a sequence $\theta^{1:T} = (\theta^1, \ldots, \theta^T)$ of threshold vectors as per Definition 1 (namely the thresholds depend on the realized bids) is:*

$$\mathtt{Reg}\left(\theta^{1:T}\right) = \sup_{b^*} \left\{ \mathbb{E}_{b^{1:T}} \left[\frac{1}{T} \sum_{t=1}^{T} \left(u(b^*, \theta^t) - u(b^t, \theta^t) \right) \right] \right\},$$

where each b^t is itself random and depends on $\theta^{1:t-1}$, as specified by the algorithm. The regret $r(T)$ of the algorithm against adaptive adversaries is the maximum regret against any adaptive adversary. An algorithm has polynomial regret rate if $r(T) = \mathrm{poly}(T^{-1}, m, \max_S v(S), B, H)$, where B, H are as in Definition 1.

III. Hardness of No-Regret Learning

We show that there does not exist an efficiently computable learning algorithm with polynomial regret rate for the online bidding problem for SiSPAs unless $\mathtt{RP} \supseteq \mathtt{NP}$, proving Theorem 1.

We first examine a related offline optimization problem which we show is NP-hard to approximate to within a small additive error. We then show how this inapproximability result implies the non-existence of polynomial-time no-regret learning algorithms for SiSPAs unless $\mathtt{RP} \supseteq \mathtt{NP}$. Our reduction from offline inapproximability to online intractability is a special case of a generic reduction provided in the full version.

Throughout this section we will consider the following very restricted class of valuations: the player is unit-demand and has a value v for getting any item, i.e. his value for any set of items is given by $v(S) = v \cdot 1\{S \neq \emptyset\}$. Our intractability results hold even if we assume that v is provided in unary representation. Consider the following problem:

Definition 3 (Optimal Bidding Problem)**.** *A distribution D of threshold vectors θ over a set of m items is given explicitly as a list of k vectors, where D is assumed to choose a uniformly random vector from the list. A bidder has a unit-demand valuation with the same value v for each item, given in unary. The problem asks for a bid vector that maximizes the bidder's expected utility against distribution D. In fact, it only asks to compute the expected value from an optimal bid vector, i.e.*

$$\sup_b \left\{ v \cdot \Pr\left[\exists j \in [m] : b_j > \theta_j\right] - \sum_{j \in [m]} \theta_j \cdot \Pr\left[b_j > \theta_j\right] \right\}.$$

We show that the optimal bidding problem is NP-hard via a reduction from r-regular set-cover. In fact we show that

it is hard to approximate, up to an additive approximation that is inverse-polynomially related to the input size. This will be useful when using the hardness of this problem to deduce the in-existence of efficiently computable learning algorithms with polynomial regret rates.

Theorem 5 (Hardness of Approximately Optimal Bidding)**.** *The optimal bidding problem is* NP*-hard to approximate to within an additive ξ even when: the k threshold vectors in the support of (the explicitly given distribution) D take values in $\{1, H\}^m$, $H = k^2 \cdot m^2$, $v = 2 \cdot k \cdot m$ and $\xi = \frac{1}{2k}$.*

Given the hardness of *optimal bidding* in SiSPAs, we are ready to sketch the proof of our main impossibility result (Theorem 1) for *online bidding* in SiSPAs. Our result holds even if the possible threshold vectors that the bidder may see take values in some known discrete finite set. It also holds even if we weaken the regret requirements of the online bidding problem, only requiring that the player achieves no-regret with respect to bids of the form $\{0, v/2m\}^m$, i.e., the bid on each item is either 0 or an $2m$-th faction of the player's value. Notice that any such bid is a non-overbidding bid. We will refer to the afore-described weaker learning task as the *simplified online bidding problem*. We sketch here how to deduce from the inapproximability of optimal bidding the impossibility of polynomial-time no-regret learning, deferring details to the full version.

Proof sketch of Theorem 1. We present the structure of our proof and the challenges that arise. Consider a hard distribution D for the optimal bidding problem from Theorem 5, and let b^* be the bid vector that optimizes the expected utility of the bidder when a threshold vector is drawn from D. Also, let u^* be the corresponding optimal expected utility. (Theorem 5 says that approximating u^* is NP-hard.) Now let us draw T i.i.d. samples $\theta = (\theta^1, \ldots, \theta^T)$ from D. Clearly, if T is large enough, then, with high probability, the expected utility $\tilde{u}_\theta$ of b^* against the uniform distribution over $\theta^1, \ldots, \theta^T$ is approximately equal to u^*.

Now let us present the sequence $\theta^1, \ldots, \theta^T$ to a no-regret learning algorithm. The learning algorithm is potentially randomized so let us call $\hat{u}_\theta$ the expected average utility (over the randomness in the algorithm and keeping sequence θ fixed) that the algorithm achieves when facing the sequence of threshold vectors $\theta^1, \ldots, \theta^T$. If the regret of the algorithm is $r(T)$, this means that $\hat{u}_\theta \geq \sup_b(\frac{1}{T}\sum_{t=1}^{T} u(b, \theta^t)) - r(T) \geq \tilde{u}_\theta - r(T)$. In particular, if $r(T)$ scales polynomially with $1/T$ then, for large enough T, $\hat{u}_\theta$ is lower bounded by $\tilde{u}_\theta$ (minus some small error), and hence by u^* (minus some small error). Hence, $\hat{u}_\theta$ (plus some small error) provides an upper bound to u^*. Moreover, if we run our no-regret learning algorithm a large enough number of times N against the same sequence of threshold vectors and average the average utility achieved by the algorithm in these N executions, we can get a very good estimate of $\hat{u}_\theta$, and hence a very good upper bound for u^*, with high probability.

The challenge that we need to overcome now is that, in principle, the expected average utility $\hat{u}_\theta$ of our no-regret learner against sequence $\theta^1, \ldots, \theta^T$ could be much larger than $\sup_b(\frac{1}{T}\sum_{t=1}^{T} u(b, \theta^t))$ and hence $\tilde{u}_\theta$ and u^*, as the algorithm is allowed to change its bid vector in every step. We need to argue that this cannot happen. In particular, we would like to upper bound $\hat{u}_\theta$ by u^*. We do this via a Martingale argument exploiting the randomness in the choice of the sequence θ. Using Azuma's inequality, we show that for large enough T, the $\hat{u}_\theta$ is upper bounded by u^* plus some small error with high probability. In fact we show something stronger: if T is large enough then, with high probability, u^* plus some small error upper bounds the algorithm's average utility (not just average expected utility), where now both the threshold and the bid vectors are left random. Hence, we can argue that, with high probability, if we run our algorithm N times over a (long enough) sequence of random threshold vectors and we compute the average (across the N executions) of the average (across the T steps) utility of our algorithm, then this double average is upper bounded by u^* plus some small error. Hence, we get a lower bound on u^*. If we choose T, N large enough then we can get an approximation of u^* to within an additive error that depends inverse polynomially with the input size. Since the latter is NP-hard, we get that there cannot exist a polynomial-time no-regret algorithm unless RP $\supseteq NP$. ■

IV. Walrasian Equilibria and No-Envy Learning

The hardness of no-regret learning in simultaneous auctions motivates the investigation of other notions of learning that have rational foundations and at the same time admit efficient implementations. Our inspiration in this paper comes from the study of markets and the well-studied notion of *Walrasian equilibrium*. Recall that an allocation of items to buyers together with a price on each item constitutes a Walrasian equilibrium if no buyer envies some other allocation at the current prices. That is the bundle S allocated to each buyer maximizes the difference $v(S) - p(S)$ of his value $v(S)$ for the bundle minus the cost of the bundle. Implicitly the Walrasian equilibrium postulates some degree of rationality on the buyers: given the prices of the items, each buyer wants a bundle of items such that he has *no-envy* against getting any other bundle at the current prices.

We adapt this *no-envy* requirement to SiSPAs. In a SiSPA a player is facing a set of prices on the items, which are determined by the bids of the other players and are hence unknown to him when he is choosing his bid vector. In a sequence of repeated executions of a SiSPA, the player needs to choose a bid vector at every time-step. The fact that he does not know the realizations of the item prices when making his choice turns the problem into a learning problem. We say that the sequence of actions that he took satisfies the no-envy guarantee, if in the long run he does not regret not buying any fixed set S at its average price.

Definition 4 (No-Envy Learning Algorithm)**.** *An algorithm for the online bidding problem of Definition 1 is* no-envy *if, for any adaptive adversary generating threshold vectors $\theta^{1:T}$, the bid vectors $b^{1:T}$ chosen by the algorithm satisfy:*

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} u(b^t,\theta^t)\right] \geq \max_{S\subseteq[m]}\left(v(S)-\mathbb{E}\left[\sum_{j\in S}\hat{\theta}_j^T\right]\right)-\epsilon(T),$$

where $\hat{\theta}_j^T = \frac{1}{T}\sum_{t=1}^{T}\theta_j^t$, $\epsilon(T)\to 0$. It has polynomial envy rate if $\epsilon(T)=\text{poly}(T^{-1},m,\max_S v(S),B,H)$.

To allow for even larger classes of settings to have efficiently computable no-envy learning outcomes, we will also define a relaxed notion of no-envy. In this notion the player is guaranteed that his utility is at least some α-fraction of his value for any set S, less the average price of that set. The latter is a more reasonable relaxation in the online learning setting given that, unlike in a market setting, the players do not know the realization of the prices when they make their decision.

Definition 5 (Approximate No-Envy Learning Algorithm)**.** *An algorithm for the online bidding problem is an α-approximate no-envy algorithm if, for any adaptively chosen sequence of threshold vectors $\theta^{1:T}$ by an adversary, the bid vectors $b^1,\ldots,b^T$ chosen by the algorithm satisfy:*

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} u(b^t,\theta^t)\right] \geq \max_{S\subseteq[m]}\left(\frac{v(S)}{\alpha}-\mathbb{E}\left[\sum_{j\in S}\hat{\theta}_j^T\right]\right)-\epsilon(T)$$

To gain some intuition about the difference between no-envy and no-regret learning guarantees consider the following. When we compute the utility from a fixed bid vector in hindsight, then in every iteration the set of items that the player would have won is nicely correlated with that round's threshold vector in the sense that the player wins an item in that round only when the item's threshold is low. On the contrary, when evaluating the player's utility had he won a specific set of items in all rounds the player may win and pay for an item even when the price of the item is high. The results of this section imply that for XOS valuations, the no-regret condition is stronger than the no-envy condition. Hence, when we analyze no-envy learning algorithms for XOS bidders we relax the algorithm's benchmark. Correspondingly, if the bidders of a SiSPA are XOS and use no-envy learning algorithms to update their bid vectors, the set of outcomes that they may converge to is broader than the set of no-regret outcomes.

Online Buyer's Problem: We first show that we can reduce the no-envy learning problem to a related online learning problem, which we call the *online buyer's problem*.

Definition 6 (Online buyer's problem)**.** *Imagine a buyer with some valuation $v(\cdot)$ over a set of m items who is asked to request a subset of the items to buy each day before seeing their prices. In particular, at each time-step t an adversary picks a set of thresholds/prices θ_j^t for each item j adaptively based on the past actions of the buyer. Without observing the thresholds at step t, the buyer picks a set S^t of items to buy. His instantaneous reward is:*

$$u(S^t,\theta^t) = v(S^t) - \mathbb{E}\left[\textstyle\sum_{j\in S^t}\theta_j^t\right], \tag{6}$$

i.e., the buyer receives the set S^t and pays the price for each item in the set.

For simplicity, we overload notation and denote by $u(b,\theta)$ the reward in the online bidding problem from a bid vector b and with $u(S,\theta)$ the reward in the online buyer's problem from a set S. We relate the online buyer's problem to the online bidding problem in SiSPAs in a black-box way, by showing that when the valuations are XOS (and assuming access to an XOS oracle), then any no-regret or "approximate" no-regret algorithm for the *online buyer's problem* can be turned in a black-box and efficient manner into a no-envy algorithm for the *online bidding problem*.

Lemma 6 (From buyer to bidder)**.** *Suppose that we have access to an efficient learning algorithm for the online buyer's problem which guarantees for any adaptive adversary that the buyer's expected average reward is at least:*

$$\max_S\left(\tfrac{1}{\alpha}v(S)-\textstyle\sum_{j\in S}\hat{\theta}_j^T\right)-\epsilon(T), \tag{7}$$

where $\hat{\theta}_j^T=\frac{1}{T}\sum_{t=1}^{T}\theta_j^t$. Then we can construct an efficient α-approximate no-envy algorithm for the online bidding problem, assuming access to XOS value oracles. Moreover, this algorithm never submits an overbidding bid.

We conclude by showing that if all players in a SiSPA use an α-approximate no-envy learning algorithm, then the average welfare is a 2α-approximation to the optimal welfare, less an additive error term stemming from the envy of the players. In other words the price of anarchy of α-approximate no-envy dynamics is upper bounded by 2α.

Theorem 7. *If n players participating in repeated executions of a SiSPA use an α-approximate no-envy learning algorithm with envy rate $\epsilon(T)$ and which does not overbid, then in T executions of the SiSPA the average bidder welfare is at least $\frac{1}{2\alpha}\text{OPT}-n\cdot\epsilon(T)$, where* OPT *is the optimal welfare for the input valuation profile $v=(v_1,\ldots,v_n)$.*

V. Online Learning with Oracles

In this section we devise novel follow-the-perturbed leader style algorithms for general online learning problems. We then apply these algorithms and their analysis to get no-envy learning algorithms for the online bidding problem. In the full version, we give implications to security games [24].

Consider an online learning problem where at each time-step an adversary picks a parameter $\theta^t\in\Theta$ and the algorithm picks an action $a^t\in A$. The algorithm receives a reward: $u(a^t,\theta^t)$, which could be positive or negative. We will assume that the rewards are uniformly bounded by some

function of the parameter θ, for any action $a \in A$, i.e.: $\forall a \in A : u(a,\theta) \in [-f_-(\theta), f_+(\theta)]$. We will denote with $\theta^{1:t}$ a sequence of parameters $\{\theta_1, \theta_2, \ldots, \theta_t\}$. Moreover, we denote with: $U(a,\theta^{1:t}) = \sum_{\tau=1}^{t} u(a,\theta^\tau)$, the cumulative utility of a fixed action $a \in A$, under sequence $\theta^{1:t}$.

Definition 7 (Optimization oracle)**.** *We will consider the case where we are given oracle access to the following optimization problem: given a sequence of parameters $\theta^{1:t}$ compute some optimal action for this sequence:*

$$M\left(\theta^{1:t}\right) = \arg\max_{a\in A} U(a,\theta^{1:t}). \tag{8}$$

We define a new type of perturbed leader algorithms where the perturbation is introduced in the form of extra samples of parameters:

Algorithm 1 (Follow the perturbed leader with sample perturbations)**.** *At each time-step t:*

1. *Draw a random sequence of parameters $\{x\}^t = \{x^1, \ldots, \ldots, x^k\}^t$ independently and based on some time-independent distribution over sequences. Both the length of the sequence and the parameter $x^i \in \Theta$ at each iteration of the sequence can be random.*

2. *Denote with $\{x\}^t \cup \theta^{1:t-1}$ the augmented sequence of parameters where we append the extra parameter samples $\{x\}^t$ at the beginning of sequence $\theta^{1:t-1}$*

3. *Invoke oracle and play: $a^t = M\left(\{x\}^t \cup \theta^{1:t-1}\right)$.*

Using a reduction of [27] (see their Lemma 12) we can show that to bound the regret of Algorithm 1 against adaptive adversaries it suffices to bound the regret against oblivious adversaries (who pick the sequence non-adaptively), of the following algorithm, which only draws the samples once.

Algorithm 2 (Follow the perturbed leader with fixed sample perturbations)**.** *Draw a random sequence of parameters $\{x\} = \{x^1, \ldots, \ldots, x^k\}$ based on some distribution over sequences and at the beginning of time. At each time-step t, invoke oracle M and play action: $a^t = M\left(\{x\} \cup \theta^{1:t-1}\right)$.*

We give a general theorem on the regret of a perturbed leader algorithm with sample perturbations.

Theorem 8. *Suppose that the distribution over sample sequences $\{x\}$, satisfies the* stability property *that for any sequence of parameters $\theta^{1:T}$ and for any $t \in [1:T]$:*

$$\mathbb{E}_{\{x\}}\left[u(a^{t+1},\theta^t) - u(a^t,\theta^t)\right] \le g(t) \tag{9}$$

where a^t as defined in Algorithm 2. Then the expected regret of Algorithm 2 against oblivious adversaries is upper bounded by:

$$\sum_{t=1}^{T} g(t) + \mathbb{E}_{\{x\}}\left[\sum_{x^\tau\in\{x\}} \left(f_-(x^\tau) + f_+(x^\tau)\right)\right] \tag{10}$$

Hence, the regret of Algorithm 1 against adaptive adversaries is bounded by the same amount.

We apply the perturbed leader approach to the online buyer's problem from Section IV. Then using Lemma 6 we can turn any such algorithm to a no-envy learning algorithm for the original bidding problem in second price auctions, when the valuations fall into the XOS class.

In the online buyer's problem the action space is the collection of sets $A = 2^m$, while the parameter set of the adversary is to pick a threshold θ_j for each item j, i.e. $\Theta = \mathbb{R}_+^m$. The reward $u(S,\theta)$, at each round from picking a set S, if the adversary picks a vector $\theta \in \Theta$ is given by Equation (6). We will instantiate Algorithm 2 for this problem and apply the generic approach of the previous section. We will specify the exact distribution over sample sequences that we will use and we will bound the functions $f_-(\cdot)$, $f_+(\cdot)$ and $g(\cdot)$. First, observe that the reward is bounded by a function of the threshold vector: $u(S,\theta) \in [-\|\theta\|_1, H]$, where H is an upper bound on the valuation function, i.e. $v([m]) < H$.

Optimization oracle: It is easy to see that the offline problem for a sequence of parameters $\theta^{1:t}$ is exactly a *demand oracle*, where the price on each item j is its average threshold $\hat{\theta}_j^t$ in hindsight.

Single-sample exponential perturbation: We will use the following sample perturbation: we will only add one sample $x \in \Theta$, where the coordinate x_i of the sample is distributed independently and according to an exponential distribution with parameter ϵ, i.e. for any $k \ge 0$ the density of x_i at k is $f(k) = \frac{1}{2}\epsilon e^{-\epsilon k}$, while it is 0 for $k < 0$.

In the full version of the paper we show that the latter perturbation leads to good stability properties for Algorithm 2, i.e. a good upper bound on $g(t)$. Given the stability bound we then apply Theorem 8 to get the following result:

Theorem 9. *Algorithm 1 when applied to the online buyers problem with a single-sample exponential perturbation with parameter $\epsilon = \sqrt{\frac{1}{HDT}}$, where D is the maximum threshold that the adversary can pick and H is the maximum value, runs in randomized polynomial time, assuming a demand oracle and achieves regret $O\left(m^2(D+H)\sqrt{T}\right)$.*

Theorem 9, Lemma 6 and the reduction form oblivious to adaptive adversaries, imply a poly-time no-envy algorithm, assuming access to demand and XOS oracles. For submodular valuations XOS oracles can be simulated in poly-time via demand oracles [21], thereby only requiring access to demand oracles. Thus we get Theorem 2.

VI. No-Envy Learning via Convex Rounding

In this section we show how to design efficient approximate no-envy learning algorithms via the use of the convex rounding technique. Though our techniques can be phrased more generally, we will cope with the concrete case where players have explicitly given coverage valuations. Answering value and XOS queries for such valuations can be done in polynomial time [25], [21].

Proving Theorem 3: Based on Lemma 6, in order to design an α-approximate no-envy algorithm for the online bidding problem, it suffices to design an efficient algorithm

for the online buyer's problem with guarantees as described in Lemma 6. In the remainder of the section we will design such an algorithm for the *online buyer's problem* with $\alpha = \frac{e}{e-1}$ and for explicit coverage valuations, thereby proving Theorem 3. Subsequently, by Theorem 7 the latter will imply a price of anarchy guarantee of $\frac{2e}{e-1}$ for such dynamics. Next we sketch how one can design an algorithm for the online buyer problem and defer the full details to the full version. Suppose that the buyer picks a set at each iteration at random from a distribution where each item j is included independently with probability x_j to the set. Then for any vector x, the expected utility of the buyer from such a choice is $\mathbb{E}_{S^t \sim x^t}\left[u(S^t, \theta^t)\right] = V(x^t) - \langle \theta^t, x^t \rangle$, where $V(\cdot)$ is the multi-linear extension of $v(\cdot)$ and $\langle x, y \rangle$ is the inner product between vectors x and y. If $V(\cdot)$ was concave we could invoke online convex optimization algorithms, such as the projected gradient descent of [28] and get a regret bound, which would imply a regret bound for the buyers problem. However, $V(\cdot)$ is not concave for most valuation classes. We will instead use a *convex rounding scheme*, which is a mapping from any vector x to a distribution over sets $D(x)$ such that $F(x) = \mathbb{E}_{S \sim D(x)}\left[v(S)\right]$ is a concave function of x. We also require that the marginal probability of each item be at most the original probability of that item in x. If the rounding scheme satisfies that for any integral x associated with set S, $F(x) \geq \frac{1}{\alpha} v(S)$, then we can call an online convex optimization algorithm on the concave function $F(x) - \langle \theta, x \rangle$. The latter yields an α-approximate no-envy algorithm for the online buyers problem.

REFERENCES

[1] G. Christodoulou, A. Kovács, and M. Schapira, "Bayesian combinatorial auctions," in *ICALP'08*, pp. 820–832.

[2] K. Bhawalkar and T. Roughgarden, "Welfare guarantees for combinatorial auctions with item bidding," in *SODA '11*.

[3] M. Feldman, H. Fu, N. Gravin, and B. Lucier, "Simultaneous auctions are (almost) efficient," in *STOC '13*, pp. 201–210.

[4] W. Vickrey, "Counterspeculation, auctions, and competitive sealed tenders," *The Journal of Finance*, vol. 16(1), pp. 8–37, 1961.

[5] E. Clarke, "Multipart pricing of public goods," *Public Choice*, vol. 11, no. 1, pp. 17–33, 1971.

[6] T. Groves, "Incentives in teams," *Econometrica*, vol. 41, no. 4, pp. pp. 617–631, 1973.

[7] L. M. Ausubel and P. Milgrom, "The lovely but lonely vickrey auction," *Combinatorial auctions*, vol. 17, pp. 22–26, 2006.

[8] S. Bikhchandani, "Auctions of heterogeneous objects," *Games and Economic Behavior*, vol. 26, no. 2, pp. 193 – 220, 1999.

[9] A. Hassidim, H. Kaplan, Y. Mansour, and N. Nisan, "Non-price equilibria in markets of discrete goods," in *EC '11*.

[10] H. Fu, R. Kleinberg, and R. Lavi, "Conditional equilibrium outcomes via ascending price processes with applications to combinatorial auctions with item bidding," in *EC'12*.

[11] V. Syrgkanis and E. Tardos, "Composable and efficient mechanisms," in *STOC '13*, pp. 211–220.

[12] S. Dobzinski, "An impossibility result for truthful combinatorial auctions with submodular valuations," in *STOC '11*.

[13] S. Dobzinski and J. Vondrak, "The computational complexity of truthfulness in combinatorial auctions," in *EC '12*.

[14] S. Dughmi and J. Vondrák, "Limitations of randomized mechanisms for combinatorial auctions," *Games and Economic Behavior*, vol. 92, pp. 370 – 400, 2015.

[15] Y. Cai and C. Papadimitriou, "Simultaneous bayesian auctions and computational complexity," in *EC '14*, pp. 895–910.

[16] T. Roughgarden and J. R. Wang, "Minimizing regret with multiple reserves," in *EC '16*, pp. 601–616.

[17] M. Braverman, J. Mao, and S. M. Weinberg, "Interpolating between truthful and non-truthful mechanisms for combinatorial auctions," in *SODA '16*, pp. 1444–1457.

[18] N. Devanur, J. Morgenstern, V. Syrgkanis, and S. M. Weinberg, "Simple auctions with simple strategies," in *EC '15*.

[19] S. Dobzinski and M. Schapira, "An improved approximation algorithm for combinatorial auctions with submodular bidders," in *SODA '06*, pp. 1064–1073.

[20] U. Feige, "On maximizing welfare when utility functions are subadditive," in *STOC '06*, pp. 41–50.

[21] S. Dobzinski, N. Nisan, and M. Schapira, "Approximation algorithms for combinatorial auctions with complement-free bidders," *Math. Oper. Res.*, vol. 35(1), pp. 1–13, Feb. 2010.

[22] A. Kalai and S. Vempala, "Efficient algorithms for online decision problems," *Journal of Computer and System Sciences*, vol. 71, no. 3, pp. 291 – 307, 2005.

[23] S. Bubeck and N. Cesa-Bianchi, "Regret analysis of stochastic and nonstochastic multi-armed bandit problems," *Machine Learning*, vol. 5, no. 1, pp. 1–122, 2012.

[24] M.-F. Balcan, A. Blum, N. Haghtalab, and A. D. Procaccia, "Commitment without regrets: Online learning in stackelberg security games," in *EC '15*, pp. 61–78.

[25] S. Dughmi, T. Roughgarden, and Q. Yan, "From convex optimization to randomized mechanisms: Toward optimal combinatorial auctions," in *STOC '11*, pp. 149–158.

[26] S. Dobzinski, "Breaking the logarithmic barrier for truthful combinatorial auctions with submodular bidders," in *STOC'16*.

[27] M. Hutter and J. Poland, "Adaptive online prediction by following the perturbed leader," *J. Mach. Learn. Res.*, vol. 6, pp. 639–660, Dec. 2005.

[28] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in *ICML'03*, pp. 928–936.