# Improved learning rates in multi-unit uniform price auctions

Marius Potfer, Dorian Baudry, Hugo Richard, Vianney Perchet, Cheng Wang

# Improved learning rates in multi-unit uniform price auctions

**Marius Potfer**[*1,2] **Dorian Baudry**[3] **Hugo Richard**[1] **Vianney Perchet**[1]

**Cheng Wan**[2]

[1] Joint team Fairplay, ENSAE, and Criteo AI LAB
[2] EDF R&D
[3] Department of Statistics, University of Oxford

## Abstract

Motivated by the strategic participation of electricity producers in electricity day-ahead market, we study the problem of online learning in repeated multi-unit uniform price auctions focusing on the adversarial opposing bid setting. The main contribution of this paper is the introduction of a new modeling of the bid space. Indeed, we prove that a learning algorithm leveraging the structure of this problem achieves a regret of $\tilde{O}(K^{4/3}T^{2/3})$ under bandit feedback, improving over the bound of $\tilde{O}(K^{7/4}T^{3/4})$ previously obtained in the literature. This improved regret rate is tight up to logarithmic terms. Inspired by electricity reserve markets, we further introduce a different feedback model under which all winning bids are revealed. This feedback interpolates between the full-information and bandit scenarios depending on the auctions' results. We prove that, under this feedback, the algorithm that we propose achieves regret $\tilde{O}(K^{5/2}\sqrt{T})$.

## 1 Introduction

The short-term electricity market, based on a wholesale market, is organized as an auction that determines the quantity each electricity producer needs to produce and the price at which electricity is sold. They participate, in this market by submitting prices for each kilowatt-hour they can produce. They have the opportunity to participate strategically, submitting prices that can deviate from their actual production cost. While several regulatory and practical constraints apply, this market is essentially a multi-unit auction of identical items (Willems & Yu, 2022). These auctions are extensively studied and utilized for resource allocation. Several pricing rules can be applied, the most common being discriminatory pricing, uniform pricing (Ausubel et al., 2014), and Vickrey–Clarke–Groves (VCG) auctions (Sessa et al., 2017). Although the VCG auction is known for its truthful bidding property, it is seldom implemented due to its complexity. Instead, uniform or discriminatory pricing are often preferred, particularly in treasury auctions (Khezr & Cumpston, 2022) and their procurement variations in electricity reserve markets (Viehmann et al., 2021).

The wholesale electricity market is held every day and, structurally, the electricity producers who participate in the mechanism remain the same for multiple years. This represents an opportunity to study how producers can be strategic in the way they adapt to other's bidding strategies. We therefore focus on the problem of online bidding in a repeated multi-unit auction. This setting allows us to model how an agent participating multiple times to an auction with the same other participants can leverage the information he has obtained during past auctions. This family of settings, of which a review is available in (Nedelec et al., 2022), was first investigated for learning from the point of view of the auctioneer and was then applied to bidders learning how to bid optimally.

[*]marius.potfer@ensae.fr

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

Online learning in multi-unit auctions with uniform pricing (every object is sold at the same price independently of the winner) is studied in (Branzei et al., 2024) while the case of discriminatory pricing is studied in Galgana and Golrezaei, 2023. When the bids of all bidders are revealed after each auction (full-information), known regret rates for uniform and discriminatory pricing are of the same order $\mathcal{O}(\sqrt{T})$ (Branzei et al., 2024; Galgana & Golrezaei, 2023) where $T$ is the time horizon. When bidders only observe the number of items they win and the price (bandit feedback) the regret upper bounds given in Galgana and Golrezaei, 2023 and Branzei et al., 2024 are of order $\tilde{\mathcal{O}}(T^{2/3})$ (for discriminatory pricing) and $\tilde{\mathcal{O}}(T^{3/4})$ (for uniform pricing) suggesting that bidding multi-unit auctions with uniform pricing is strictly harder than with discriminatory pricing. Our study shows that this is not the case as we present an algorithm achieving regret $\tilde{\mathcal{O}}(T^{2/3})$ with uniform pricing therefore closing the gap between the two settings.

**Auction rules** A decision-maker (i.e., *the bidder*) repeatedly bids in a uniform pricing $K$-unit auction. The single-shot version of the auction, from the perspective of any participant $i$ whose value of obtaining a $\text{k}^{\text{th}}$-item is denoted by $v_{i,k} \in [0,1]$, proceeds as follows.

1. Each participant submits a bid profile $(b_{i,k})_{k\in[K]} \in B$, where
$$B = \{(b_k)_{k\in[K]}, \text{ such that } 1 \geq b_1 \geq b_2 \geq \ldots \geq b_K \geq 0\}.$$
We call $B$ the action space and we denote by $\mathbf{b}_{-i}$ the bids from other participants.
2. The price per item $p(\mathbf{b}_i, \mathbf{b}_{-i})$ is set either as
   - the $K^{\text{th}}$ highest bid (Last Accepted Bid (LAB) pricing rule),
   - the $(K+1)^{\text{th}}$ highest bid (First Rejected Bid (FRB) pricing rule).
3. The participant receives the items they won and pays $p(\mathbf{b}_i, \mathbf{b}_{-i})$ for each item. Since items are identical, we call allocation $x_i \in [K] := \{1, 2, ..., K\}$ the number of items participant $i$ receives, formally defined as follows:
$$x_i(\mathbf{b}_i, \mathbf{b}_{-i}) := \begin{cases} |\{k \in [K] \text{ s.t. } b_{i,k} \geq p(\mathbf{b}_i, \mathbf{b}_{-i})\}| \text{ for the LAB rule} \\ |\{k \in [K] \text{ s.t. } b_{i,k} > p(\mathbf{b}_i, \mathbf{b}_{-i})\}| \text{ for the FRB rule} \end{cases}. \tag{1}$$

The focus is to design efficient learning algorithms for the decision-maker, i.e., one specific participant $i$; we can therefore aggregate bids from other participants as the bid of a single *adversary* $\boldsymbol{\beta} := \mathbf{b}_{-i}$ and omit the index $i$ of the learner denoting $\mathbf{b} := \mathbf{b}_i$. This setup gives rise to the quasi-linear utility $u(\mathbf{b}, \boldsymbol{\beta}) = \sum_{l=1}^{x(\mathbf{b},\boldsymbol{\beta})} [v_l - p(\mathbf{b}, \boldsymbol{\beta})]$.

In the remainder of this paper, we adopt the LAB pricing rule. The techniques and theoretical proofs can be adapted from one setup to the other with little change. In the absence of specific mention of a pricing rule, our results can be applied to both auction types.

**Repeated setting** As mentioned above, this auction is not played just once, but repeated many times (say, each day). We shall then denote a time horizon $T$, and assume a different auction is run at each time step $t \in [T]$ and the objective of the bidder is to maximize their cumulative utility. Quite naturally, the bidder should adjust their bids to the adversary's behavior, learned from the outcomes of the previous iterations. On the other hand, we assume that the bidder does not need to learn their own values, i.e., the valuations $(v_k)_{k\in[K]}$ are known to the bidder and do not change over time.

We denote by $(\mathbf{b}^t)_{t\in[T]}$ and $(\boldsymbol{\beta}^t)_{t\in[T]}$ respectively the sequences of bids of the player and of the adversary, and by $p^t := p\left(\mathbf{b}^t, \boldsymbol{\beta}^t\right)$ and $x^t := x\left(\mathbf{b}^t, \boldsymbol{\beta}^t\right)$ the price and allocation at time $t$. The utility of the bidder after the auction $t \in [T]$ is then defined as $u(\mathbf{b}^t, \boldsymbol{\beta}^t) = \sum_{l=1}^{x^t} (v_l - p^t)$. As standard in online learning, we evaluate the performance of a learning (bidding) strategy through its *regret*, defined as follows

$$R_T = \sup_{\mathbf{b}\in B} \sum_{t=1}^{T} u(\mathbf{b}, \boldsymbol{\beta}^t) - \mathbb{E}\left[\sum_{t=1}^{T} u(\mathbf{b}^t, \boldsymbol{\beta}^t)\right], \tag{2}$$

where the expectation is taken over the randomness of the algorithm generating the bids $\mathbf{b}^t$. Maximizing the utility of the bidder is equivalent to minimizing the regret.

**Feedback** The bidders can improve their strategy using the information they receive after each iteration of the auction. The type of *feedback* they receive represents their knowledge about the bids of the adversary. In the literature, two common types of feedback are considered (Cesa-Bianchi et al., 2023),

1. the *bandit* feedback where the bidder's allocation $x_t$ is revealed and the price $p_t$ is only revealed if $x_t > 0$, and
2. the *full information* feedback where all the bids emitted by all participants are revealed.

Inspired by the terms of commodity electricity markets in several European countries *including Germany and France*, similarly to Karaca et al., 2020, we shall introduce and study a third partial feedback specific to multi-unit auctions,

3. the *all-winner feedback*: the allocation, the price, and all the winning bids are revealed to the bidder.

**Remark 1.** *With a uniform discretization of the bidding space $B$, learning to bid in multi-unit uniform auctions can be recast as a special instance of a combinatorial bandit problem. In the latter, the decision maker sequentially selects multiple arms out of $N$ available, i.e., picking at each stage an action in some admissible subset of $\{0,1\}^N$. Using off-the-shelf combinatorial bandit algorithms, that do not leverage the relevant structure of repeated auctions, would end up in a highly inefficient and sub-optimal procedure (see Section 2). Our approach is different; in essence, we reduce the complex combinatorial problem by expressing the utility objective as a polynomial (in $K$ and $T$) sum of simpler functions.*

**Related Work** Multiple-unit auctions of indivisible identical items have been extensively studied in their static settings. In particular, how the pricing rules (discriminatory, uniform, VCG) influence revenue (Ausubel et al., 2014), social welfare (Birmpas et al., 2019; De Keijzer et al., 2013), or price stability (Anderson & Holmberg, 2018). Their use in the context of electricity markets is common and similar questions are being studied with this specific application in mind (Akbari-Dibavar et al., 2020; Cramton & Stoft, 2006; Fabra et al., 2006; Son et al., 2004)

The repeated setting of auctions, and specifically the use of online learning procedure inspired by Multi Armed Bandits has received lots of attention in the last decade. First studied from the point of view of the auctioneer: Blum et al., 2004 studied maximizing auction revenue, Cesa-Bianchi et al., 2014 and Kanoria and Nazerzadeh, 2014 specifically focused on learning reserve prices. Learning to bid, the bidder's problem, was considered later on, initially in single-item auctions. Second price auctions facing either adversarial or stochastic highest opposing bids were studied in Weed et al., 2016, and in a contextual, budget-constrained setting by Flajolet and Jaillet, 2017. Balseiro et al., 2019 considered the first price auction with adversarial opposing bids leading to optimal regret rates of $\tilde{\mathcal{O}}(T^{2/3})$ in the known valuation and contextual setting.

First mentioned in Feng et al., 2018 for unit demand, multiple unit auctions as online learning problems only recently started to be considered as their own topic of interest. Discriminatory pricing and uniform pricing respectively studied in Galgana and Golrezaei, 2023 and Branzei et al., 2024 can be learned with $\tilde{\mathcal{O}}(\sqrt{T})$ in the full-information setting. Under bandit feedback, the former achieves $\tilde{\mathcal{O}}(KT^{2/3})$ regret rates in discriminatory pricing and the latter $\tilde{\mathcal{O}}(K^{7/4}T^{3/4})$ regret rates with uniform pricing. Compared to single unit auctions, the combinatorial nature of the action space in $K$-unit auctions makes it a harder learning problem. Branzei et al., 2024 makes use of a cautiously designed equivalent action space represented as a Directed Acyclic Graph (DAG) to address the combinatorial limitation and to design an algorithm guaranteeing the aforementioned regret bounds.

The effects of specific feedback on the ability to achieve lower regret rates have also raised some interest. Feng et al., 2018 studied the effects of "Win Only" feedback in a more general auction setting. More recently, Cesa-Bianchi et al., 2023 focused on feedback transparency. They characterize gaps in the regret rates that can be achieved depending on the amount of feedback received, getting three separate rates $O(\sqrt{T})$, $O(T^{2/3})$ and $\Omega(T)$ depending on the feedback considered. The work of Karaca et al., 2020, similarly to the all-winner feedback, studied partial feedback, which lie in between bandit and full-information, motivated by electricity market auctions.

**Contribution** We introduce a novel representation of the action space that overcomes the combinatorial complexity introduced by the multiplicity of the bids in $K$-unit auction. Inspired by the properties of the equivalent action space used by Branzei et al., 2024, we introduce bid-gaps, to further decompose the utility into a sum of independent functions. This decomposition leads to improved regret rates of $\tilde{\mathcal{O}}(K^{4/3}T^{2/3})$ under bandit feedback, compared to the known upper bound of $\tilde{\mathcal{O}}(K^{7/4}T^{3/4})$. These improved bounds match, in terms of $T$, the rates $\tilde{\mathcal{O}}(KT^{2/3})$ achievable in discriminatory pricing. We notice a reduction to simpler auctions which bear an $\Omega(T^{2/3})$ lower bound on the regret, answering the open question of the optimal rates dependency in $T$ in the bandit setting. Motivated by the terms of bid revelation in electricity reserve markets in several European countries *including Germany and France*, a novel feedback structure is considered, which lies in between bandit feedback and full information. This feedback, which we call all-winner, reveals all the winning bids of the action. We propose an algorithm that achieves a $\tilde{\mathcal{O}}(K^{5/2}\sqrt{T})$ regret, almost matching the regret rates under full information up to a factor $K$, while the lower bound of $\Omega(K\sqrt{T})$ proved by Branzei et al., 2024 for the full-information feedback, naturally extend to this setting. We summarize our results in Table 1 below.

| Feedback | Literature | This work | Lower bound |
|---|---|---|---|
| Full information | $\tilde{\mathcal{O}}(K^{3/2}\sqrt{T})$ | $\tilde{\mathcal{O}}(K^{3/2}\sqrt{T})$ | $\Omega(K\sqrt{T})$ |
| All winner | | $\tilde{\mathcal{O}}(K^{5/2}\sqrt{T})$ | $\Omega(K\sqrt{T})$ |
| Bandit | $\tilde{\mathcal{O}}(K^{7/4}T^{3/4})$ | $\tilde{\mathcal{O}}(K^{4/3}T^{2/3})$ | $\Omega(T^{2/3})^\star$ |

Table 1: Regret Rates in multi-unit uniform price auction. $^\star$ holds in the LAB pricing rule setting

## 2 Action space

We first motivate the new cautiously designed action space, and provide intuitions on how it is constructed and its main properties. We then formalize its definition.

**Motivation for an alternative representation** Usual techniques such as uniform discretization of the action space as in (Feng et al., 2018) might lead to consider the subset of non-increasing sequences on this discretization, denoted by $B_\epsilon \subset \left\{0, \epsilon, 2\epsilon, ..., (\lfloor \frac{1}{\epsilon} \rfloor - 1)\epsilon, \lfloor \frac{1}{\epsilon} \rfloor \epsilon\right\}^K$. Without loss of generality, we shall assume in the following that $1/\epsilon$ is an integer. The main downside of this representation is that the size of $B_\epsilon$ is exponential and thus, without any further properties of the problem leveraged, this would lead to arbitrarily bad regret rates $\tilde{\mathcal{O}}(T^{\frac{K+1}{K+2}})$ in bandit setting. Even though we can restrict the available action to *reasonable ones* (ie undominated strategies 6) this isn't enough to achieve improved rates in general.

Branzei et al., 2024 proposed a Directed Acyclic Graph (DAG) equivalent of the action space $B_\epsilon$ to overcome this combinatorial limitation. They use the decomposition of the utility into a sum of independent functions, depending only on pairs of consecutive bids (the edges in their graphs), to reduce the combinatorial complexity to only 2 (instead of $K$), and thus achieved an $\tilde{O}(K^{7/4}T^{3/4})$ regret bound under bandit feedback. Motivated by the breakthrough enabled by such a representation of the action space, we consider a new equivalent action space $H_\epsilon$, introduced in Equation 5. This new action space allows to leverage more precisely the regularity of the utility with respect to the bidder's choice of bids, which in turn leads to improved regret bounds under bandit feedback, presented in Theorem 1.

### 2.1 Action space tailored to the outcomes

We now provide intuitions on the utility regularity that will be leveraged. We start by observing that, for a given auction, the price is either set by one of the bidder's bid or by a bid from the adversary.

Assume that the bidder bids $(b_1, ..., b_K) \in B_\epsilon$, and that the price is $b_k$ for some $k \in [K]$. Then, we claim that many bid profiles would have led to the same outcome. Indeed, any bid profile (from the bidder) with the same $k^{\text{th}}$ bid $b_k$, leads to the same outcome. In the alternative case, where the adversary sets the price, if the bidder wins $k$ items (i.e., the $K-k$ adversary's bid $\beta^t_{K-k}$ sets the price), then any bid profile satisfying $b_k \geq \beta^t_{K-k} \geq b_{k+1}$ leads to the same outcome.

Notice that in the two aforementioned cases of regularities of the utility, all bids $\mathbf{b} \in B_\epsilon$ which would lead to the same outcome share one of the following properties: for a specific $k$ and $j$,

- in the first case : $b_k = j\epsilon$,
- in the second case : $b_k \geq (j+1)\epsilon \geq \beta^t_{K-k} \geq j\epsilon \geq b_{k+1}$.

For simplicity, we shall assume that the bids of the decision-maker belong to the $\epsilon$-discretization, i.e., $(b_1, ..., b_K) \in B_\epsilon$ while the bids of the adversary do not belong to it, in order to avoid ties[2]. We denote this set $B_{\setminus\epsilon}$, the set of non-increasing sequences of $[0,1]^K$ without values in $\left[\frac{1}{\epsilon}\right]$.

We, introduce an alternative description of the bidding space $B_\epsilon$ whose structure closely matches the aforementioned regularities' in order to improve the bidder's strategy. It leverages new binary variables indicating which of the aforementioned properties a bid $\mathbf{b} \in B_\epsilon$ has, they are defined as follows: for any $k \in [K]$ and $j \in \left[\frac{1}{\epsilon}\right]$,

$$h_{k,j}(\mathbf{b}) = \mathbb{1}\{b_k = j\epsilon\} \tag{3}$$

$$h_{k+\frac{1}{2},j}(\mathbf{b}) = \mathbb{1}\{b_k \geq (j+1)\epsilon > j\epsilon \geq b_{k+1}\}. \tag{4}$$

Let $\mathcal{K} = \left\{1, \frac{3}{2}, 2, ..., K-1, \frac{2K-1}{2K}, K\right\}$ and $\mathcal{J}_\epsilon = \left[\frac{1}{\epsilon}\right]$. For any $\mathbf{b} \in B_\epsilon$, we define the pseudo-bid $\mathbf{h_b}$ to be the list of these binary variable $h_{k,j}$ with $k, j \in \mathcal{K} \times \mathcal{J}$, such that $h_{k,j}(\mathbf{b}) = 1$, ordered in lexicographic order, increasingly in $k$ and decreasingly in $j$. We naturally define $H_\epsilon$ the pseudo-bid space generated by the bid space $B_\epsilon$ :

$$H_\epsilon = \{\mathbf{h_b} | \mathbf{b} \in B_\epsilon\} \tag{5}$$

**Lemma 1.** *For each pseudo-bid $\mathbf{h} \in H_\epsilon$, there exists a unique $\mathbf{b} \in B_\epsilon$ such that $\mathbf{h} = \mathbf{h_b}$. This therefore defines a bijective mapping between $H_\epsilon$ and $B_\epsilon$.*

*Proof.* From the expression of $H_\epsilon$ in (5), it is clear that the mapping $\mathbf{b} \mapsto \mathbf{h_b}$ , is surjective.
Let $\mathbf{h} \in H_\epsilon$, there exists $\mathbf{b} = \{b_1, ..., b_K\} \in B_\epsilon$ such that $\mathbf{h} = \mathbf{h_b}$. Let $j_k \in \mathcal{J}_\epsilon$ such that $b_k = j_k\epsilon$, for all $k \in [K]$, we have $h_{k,j_k} \in \mathbf{h}$. If there exists another bid $\tilde{\mathbf{b}} = \{\tilde{b}_1, ..., \tilde{b}_K\} \in B_\epsilon$ such that $\mathbf{h} = \mathbf{h}_{\tilde{\mathbf{b}}}$, for all $k \in [K], h_{k,j_k}(\tilde{\mathbf{b}}) = 1$.Therefore (3) yields $\tilde{b}_k = j_k\epsilon = b_k$ for all $k \in [K]$. This proves unicity and therefore that the mapping is bijective. □

The following characterization of the pseudo-bid space directly follows from Lemma 1.

**Corollary 1.** *Given a bid $\mathbf{b} = (b_i)_{i\in[K]} \in B_\epsilon$ and the pseudo-bid $\mathbf{h_b} \in H_\epsilon$, we have the following: for all $k, j \in [K] \times \mathcal{J}_\epsilon$,*

$$b_k = j\epsilon \iff h_{k,j} \in \mathbf{h_b} \tag{6}$$

$$b_k \geq (j+1)\epsilon > j\epsilon \geq b_{k+1} \iff h_{k+\frac{1}{2},j} \in \mathbf{h_b} \tag{7}$$

To provide further intuition, Figure 1a and Figure 1b show how two corresponding bids might be represented in $B_\epsilon$ and $H_\epsilon$. The bids (6) are represented by circles, while the bid-gaps (7) are ellipses.

### 2.2 Utility decomposition

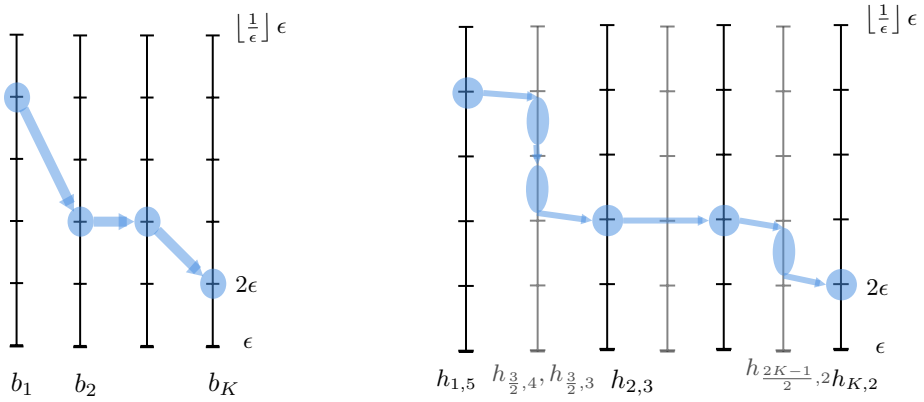
Leveraging the new action space, we define the utility, price, and allocation function on $H_\epsilon$ resulting from the bijective map with $B_\epsilon$. Let $\mathbf{h} \in H_\epsilon$ and $\mathbf{b} \in B_\epsilon$ the unique element of $B_\epsilon$ such that $\mathbf{h} = \mathbf{h_b}$. For all $\boldsymbol{\beta} \in B_{\setminus\epsilon}$, we define the utility as $u_H(\mathbf{h_b}, \boldsymbol{\beta}) := u(\mathbf{b}, \boldsymbol{\beta})$, the price $x_H(\mathbf{h_b}, \boldsymbol{\beta}) := x(\mathbf{b}, \boldsymbol{\beta})$ and the allocation $p_H(\mathbf{h_b}, \boldsymbol{\beta}) := p(\mathbf{b}, \boldsymbol{\beta})$

The following additional set notation, which matches a binary variable $h_{k,j}$ to its corresponding price range, allows for unified descriptions :

$$\mathcal{P}_\epsilon(h_{k,j}) := \begin{matrix} \{j\epsilon\} & \text{if } k \text{ is an integer} \\ (j\epsilon, (j+1)\epsilon) & \text{if } k \text{ is half integer} \end{matrix}, \tag{8}$$

We now explicitly show how the pseudo-bid space is *well suited* to capture the regularity of the outcomes of the auction (and therefore of the utility) mentioned above. To be more precise, the

[2]This assumption comes without loss of generality, since e.g. adding uniform noise with extremely small variance to the bidder's bid prevent ties a.s., see Lemma 8 in subsection A.3




(a) Bids in $B_\epsilon$ (b) Bids and corresponding bid-gaps in $B(\mathcal{P}_\epsilon)$

Figure 1: Graph representation of action spaces $B_\epsilon$ (Branzei et al., 2024) and $B(\mathcal{P}_\epsilon)$ (this paper)

following Lemma 2 states that within a pseudo-bid profile $\mathbf{h}$, a pseudo-bid $h_{k,j}$ can be *credited* for the outcome: any other pseudo-bid profile containing this pseudo bid would have lead to the same outcome.

**Lemma 2.** *Let $\boldsymbol{\beta} \in B_{\setminus\epsilon}$ and $(k,j) \in \mathcal{K} \times \mathcal{J}_\epsilon$. There exists $C \in \{0,1\}$, such that for all $\mathbf{h} \in H_\epsilon$ with $h_{k,j} \in \mathbf{h}$,*

$$\mathbb{1}\{p_H(\mathbf{h},\boldsymbol{\beta}) \in \mathcal{P}_\epsilon(h_{k,j})\} \cap \{x_H(\mathbf{h},\boldsymbol{\beta}) = \lfloor k \rfloor\} = C \tag{9}$$

*and if $C = 1$, $p_H(\mathbf{h},\boldsymbol{\beta})$ is also constant on $\{\mathbf{h} \in H_\epsilon : h_{k,j} \in \mathbf{h}\}$*

*Proof.* Let $\mathbf{b} \in B_\epsilon$ such that $h_{k,j} \in \mathbf{h_b}$. If $k$ is integer, then Corollary 1 yields $b_k = j\epsilon$, hence we get $\mathbb{1}\{p_H(\mathbf{h},\boldsymbol{\beta}) = j\epsilon\} \cap \{x_H(\mathbf{h},\boldsymbol{\beta}) = k\} = \mathbb{1}\{\beta_{K-k} > j\epsilon = b_k > \beta_{K-k+1}\}$ which only depends on $k, j$ and $\boldsymbol{\beta}$. If $k$ is half-integer, then Corollary 1 yields $b_{k+1} < j\epsilon < (j+1)\epsilon < b_k$, hence we get $\mathbb{1}\{p_H(\mathbf{h},\boldsymbol{\beta}) = \mathcal{P}_\epsilon(h_{k,j})\} \cap \{x_H(\mathbf{h},\boldsymbol{\beta}) = k\} = \mathbb{1}\{j\epsilon < \beta_{K-k} < (j+1)\epsilon\}$ which also only depends on $k, j$ and $\boldsymbol{\beta}$. It is straightforward to see that when the indicator function takes value one, the price is constant with value $j\epsilon$ for the integer case and $\beta_{K-k}$ in the half integer case. □

Lemma 2 allows us to exhibit a key property of the utility on the pseudo-bid space: it can be decomposed into a sum of sub-utilities (defined in Equation 11), each of which only depends on one of the components of $\mathbf{h}$.

**Lemma 3.** *Let $\mathbf{h} \in H_\epsilon$ and $\boldsymbol{\beta} \in B_{\setminus\epsilon}$. The utility of the bidder rewrites as a sum of sub-utilities:*

$$u_H(\mathbf{h},\boldsymbol{\beta}) = \sum_{h_{k,j} \in \mathbf{h}} w(h_{k,j},\boldsymbol{\beta}), \text{ with} \tag{10}$$

$$w(h_{k,j},\boldsymbol{\beta}) := \mathbb{1}\left\{\{p_H(\mathbf{h},\boldsymbol{\beta}) \in \mathcal{P}_\epsilon(h_{k,j})\} \cap \{x_H(\mathbf{h},\boldsymbol{\beta}) = \lfloor k \rfloor\}\right\} \sum_{l=1}^{\lfloor k \rfloor} (v_l - p_H(\mathbf{h},\boldsymbol{\beta})). \tag{11}$$

*Proof of Lemma 3.* $x(\mathbf{h},\boldsymbol{\beta}) = 0$ implies that $u$ as defined in Equation (10) is zero, as expected. For the case where $x(\mathbf{h},\boldsymbol{\beta}) > 0$, we use that the indicator functions in Equation 11 correspond to disjoint events. Thus, there exists a unique pair $(k,j) \in \mathcal{K} \times \mathcal{J}_\epsilon$ such that $w(h_{k,j},\boldsymbol{\beta}) > 0$, furthermore, this sub-utility $w(h_{k,j},\boldsymbol{\beta})$ matches the utility $u(\mathbf{h},\boldsymbol{\beta})$. This concludes the proof. □

While the expression of these indicators in Equation (11) involves the full action $\mathbf{h}$, Lemma 2 shows that they only depend either on the associated bid $h_{k,j}$. Furthermore, notice that within a given $\mathbf{h} \in H_\epsilon$, the events corresponding to each $h_{k,j} \in \mathbf{h}$ are disjoints and therefore only one can be realized. As a result there is at most one $h \in \mathbf{h}$ with positive sub-utility, we denote it $h_\star(\mathbf{h},\boldsymbol{\beta})$.

## 3 Learning algorithms and guarantees

We first present two algorithms and estimators of the utility corresponding to the different feedback settings. We then state the regret rate achieved by these algorithms when used with the introduced estimator, depending on the setting considered, as well as a lower bound on the regret in the bandit setting. To simplify notations, we shall further introduce $u_H^t(.) := u_H(., \boldsymbol{\beta}^t)$ and $w^t(.) := w(., \boldsymbol{\beta}^t)$.

### 3.1 Algorithm for online learning in K-unit uniform auction

The regret minimizing algorithm combines two separate procedures,

- An exponential weight update, detailed in Algorithm 1, which creates and updates the weights for each bid and bid-gap at each time step, akin to an EXP3 type algorithm (but non-normalized) (Cesa-Bianchi and Lugosi, 2006,Lattimore and Szepesvári, 2020).
- A sampling procedure, tailored to our action space, described in Algorithm 2, that re-normalizes the weights into a probability distribution by using a weight-pushing method, and uses an efficient procedure to sample an action. (Takimoto and Warmuth, 2002).

Because of the combinatorial structure of the problem, using directly the exponential weight algorithm would be highly inefficient, with a complexity of order $\mathcal{O}(\frac{1}{\epsilon^K})$ to store weights and compute probabilities directly on the action space. The second auxiliary sampling algorithm gets rid off this complexity burden.

Algorithm 1's pseudo-code details the exponential weight algorithm used as a building block of the no-regret procedure. It computes weights for each pseudo-bid based on the corresponding sub-utilities, or their estimated values. These weights are used to sample a pseudo-bid sequence $\mathbf{h} \in H_\epsilon$.

**Algorithm 1:** Component based exponential weighting

**Input:** time horizon $T$, parameters $\varepsilon > 0$ and $\eta > 0$
**Output:** actions for each time step $\left(\mathbf{h}^1, \mathbf{h}^2, \ldots, \mathbf{h}^{T-1}, \mathbf{h}^T\right) \in (H_\epsilon)^T$.
**Initialize:** for $(k, j) \in \mathcal{K} \times \mathcal{J}_\epsilon$, set $Q^0(h_{k,j}) = 1$
**for** $t = 1, 2, \ldots, T$ **do**
Sample $\mathbf{h}^t$ using the sampling procedure of Algorithm 2, with input parameters $Q^{t-1}$;
Receive the utility $u^t = u(\mathbf{h}^t, \boldsymbol{\beta}^t)$ and the feedback. Based on this feedback, define

$$v^t = \begin{cases} w^t = w(., \boldsymbol{\beta}^t) & \text{in the full-information feedback, see (11)} \\ \hat{w}^t & \text{with bandit feedback, see (19)} \\ \bar{w}^t & \text{with all-winner feedback, see (20)} \end{cases} ;$$

Update the weights based on the weights $v^t$ :
**for** $(k, j) \in \mathcal{K} \times \mathcal{J}_\epsilon$ **do**

$$Q^t(h_{k,j}) = Q^{t-1}(h_{k,j}) \exp\left(\eta v^t(h_{k,j})\right) \tag{12}$$

Algorithm 2 is a sampling procedure and uses a weight-pushing technique (Takimoto and Warmuth, 2002) to efficiently sample pseudo-bids and compute weights and probabilities on $H_\epsilon$. It is exploiting the lexicographical ordering of the pseudo bid (increasingly in $k$ and decreasingly in $j$) which creates a graph-like structure, as illustrated in Figure 1b. It is indeed possible to sample an element of $H_\epsilon$ by repeatedly sampling the next binary variable conditionally on the previous one. To explicit this graph-like structure, we define $s(.)$ the successor function, which given a pseudo-bid $h_{k,j}$ provides the set of possible next element of an action. For $k \in [K-1]$, $j \in \mathcal{J}_\epsilon \setminus \{0\}$,

$$s(h_{k+1/2,j}) := \left\{h_{k+1/2,j-1}, h_{k+1,j}\right\} \tag{13}$$

$$s(h_{k,j}) := \left\{h_{k+1/2,j-1}, h_{k+1,j}\right\} \tag{14}$$

$$s(h_{k,0}) := \{h_{k+1,0}\} \tag{15}$$

As well as the following, which serves as the stopping condition of the sampling:

$$\forall j \in \mathcal{J}_\epsilon,\ s(h_{K,j}) := \emptyset \tag{16}$$

**Algorithm 2:** Selection of the bids by a weight-pushing algorithm

**Input:** Weights $Q$ for every bids or bid-gap $(h_{k,j})_{(k,j)\in\mathcal{K}\times\mathcal{J}_\epsilon}$;
**Output:** bid vector $\mathbf{h} \in H_\epsilon$;
**Initialize :** $k \leftarrow K - \frac{1}{2}$. For all $j \in \mathcal{J}_\epsilon$, $\Gamma(h_{K,j}) \leftarrow 1$;
**while** $k \geq 1$ **do**
  **for** $j \in \mathcal{J}_\epsilon$ **do**
    $\Gamma(h_{k,j}) \leftarrow \sum_{h\in s(h_{k,j})} \Gamma(h)Q(h)$
  $k \leftarrow k - \frac{1}{2}$
$\Gamma_0 \leftarrow \sum_{j\in\left[\frac{1}{\epsilon}\right]} Q(h_{1,j})\Gamma(h_{1,j})$;
Sample $h$ according to the probabilities:
$\forall j \in \mathcal{J}_\epsilon, \mathbb{P}(h = h_{1,j}) = Q(h_{1,j})\frac{\Gamma(h_{1,j})}{\Gamma_0}$;
$\mathbf{h} \leftarrow \{h\}$;
**while** $s(h) \neq \emptyset$ **do**
  sample $h_+ \in s(h)$, the next element of the sequence $\mathbf{h}$, with probability:
  $\mathbb{P}(h_+ = h'|h) = Q(h')\frac{\Gamma(h')}{\Gamma(h)}$ for all $h' \in s(h)$;
  $\mathbf{h} \leftarrow \mathbf{h} \cup \{h_+\}$;
  $h \leftarrow h_+$;

The combination of Algorithm 1 and Algorithm 2 leads to the following probabilities, typical of an exponential weight algorithm, on $H_\epsilon$. For $\mathbf{h} \in H_\epsilon$,

$$\mathbb{P}^t(\mathbf{h}) = \frac{\exp\left(\sum_{n=0}^t \eta u_H^n(\mathbf{h})\right)}{\sum_{\mathbf{a}\in H_\epsilon} \exp\left(\sum_{n=0}^t \eta u_H^n(\mathbf{a})\right)} \tag{17}$$

as shown in Appendix B, in Equation 29.

**Estimators** With partial feedback (either bandits or all-winner), the bidder does not gather enough information to compute all of the sub-utilities, and it can only do it for a subset of pseudo-bids.They must therefore resort, as it is standard in multi-armed bandit literature, to estimators that should leverage all the information available. Under bandit feedback, only sub-utilities of binary variables which belong to the action played at time $t$ can be computed.- On the other hand, under all-winner feedback, the richer feedback allows to compute sub-utilities for a bigger set of binary variables $h$, we denote it $A$ and define it in Lemma 4.

**Lemma 4.** *With the all-winner feedback, the bidder can compute from its feedback the sub-utilities of any pseudo bid in $A(\mathbf{h}^t, \boldsymbol{\beta}^t)$, defined as:*

$$A(\mathbf{h}^t, \boldsymbol{\beta}^t) := \left\{h_{k,j}, (k,j) \in \mathcal{K}\times\mathcal{J}_\epsilon \middle| \text{ s.t. } \{k > x^t\} \text{ or } \{k = x^t \text{and } j\epsilon \geq p^t\}\right\} \tag{18}$$

*Where $x^t := x_H(\mathbf{h}^t, \boldsymbol{\beta}^t)$ and $p^t = p_H(\mathbf{h}^t, \boldsymbol{\beta}^t)$.*

The formal proof of Lemma 4 is in Appendix C.

As noted above, within a given pseudo-bid $\mathbf{h}$, only one sub-utility can be non-zero. We therefore also define the set of binary variables with non-zero sub-utilities $A_\star(\mathbf{h}^t, \boldsymbol{\beta}^t) := \{h_{k,j} \in A(\mathbf{h}^t, \boldsymbol{\beta}^t) | w(h_{k,j}, \boldsymbol{\beta}^t) > 0\}$.

We can now formally introduce the estimators used by the no-regret procedure.

**Definition 3.1** (Estimators)**.** *Let $\mathbf{h}^t \in H_\epsilon$ be the action played by the learner, and $\boldsymbol{\beta}^t \in B_{\setminus\epsilon}$ the bids of the adversary at time t,. For any bid or bid-gap $h$, we define the sub-utility estimators:*

$$\textit{Bandit feedback } \hat{w}^t(h) = \mathbb{1}(h = h_\star^t)\frac{w^t(h) - K}{\mathbb{P}^t(h)}, \tag{19}$$

$$\textit{All-winner feedback } \bar{w}^t(h) = \mathbb{1}\left(h \in A_\star^t(\mathbf{h}^t)\right)\frac{w^t(h) - K}{\mathbb{P}_{\mathbf{f}^t\sim\mathcal{B}^t}(h \in A_\star^t(\mathbf{f}^t))} \tag{20}$$

*where* $h_\star^t := h_\star(\mathbf{h}^t, \boldsymbol{\beta}^t)$ *is the sub/pseudo-bid played at time* $t$ *that has non-zero sub-utility,* $\mathcal{B}^t$ *is the probability distribution on* $H_\epsilon$ *as in* (17) *and* $\mathbb{P}^t(h) := \sum_{\mathbf{h} \in H_\epsilon : h \in \mathbf{h}} \mathbb{P}^t(\mathbf{h})$

Naturally, estimation of the utility of any action $\mathbf{h} \in H_\epsilon$ is done with a simple summation, over $h \in \mathbf{h}$, of these estimates.

### 3.2 Regret Analysis

Combining Algorithm 1 and the sampling Algorithm 2, we recover the regret guarantees obtained by (Branzei et al., 2024, Theorem 2) in the same setting for the full-information feedback. We provide a formal statement and proof of this result in the Appendix B, in Theorem 3. We now analyze the performance of the learning procedure for the two other types of feedback.

**Theorem 1.** *In the repeated* $K$*-unit auction with uniform pricing guarantees and under bandit feedback, Algorithm 1 incurs a regret of at most* $\mathcal{O}\left(K^{4/3}T^{2/3}\log(T)\right)$ *For any time horizon* $T$*, with the choices of* $\epsilon = \left(\frac{K}{T}\right)^{1/3}$ *and* $\eta = K^{-1/3}T^{-2/3}\sqrt{\log\left(\frac{T}{K}\right)/3}$.

*Proof sketch.* The aforementioned regret bounds are proved by using a similar analysis as the one in Lattimore and Szepesvári, 2020 to obtain regret bounds of EXP3 algorithm in the adversarial bandits case. We apply this analysis to the discretized action space $H_\epsilon$ and bound the additional cost of using a discretization separately. Then we choose a discretization size $\epsilon$ to minimize the total regret.

The improvement over known regret bounds in Branzei et al., 2024 results from the decomposition of the utility into sub-utilities. Since these sub-utilities only depend on one bid or bid-gap, the *variance* of the estimators (cf Lemma 9) only depend on the possible number of bids or bid-gaps ( of order $\frac{1}{\epsilon}$) not the number of bid profiles ( of order $\frac{1}{\epsilon^K}$). This is akin to why combinatorial bandits under semi-bandit feedback (Audibert et al., 2014) achieve better regret than under bandit-feedback. □

**Theorem 2.** *For any time horizon* $T$*, using Algorithm 1 in the repeated* $K$*-unit auction with uniform pricing guarantees, under all-winner feedback, a regret of at most* $\mathcal{O}\left(K^{5/2}\sqrt{T}\log(T)\right)$ *with* $\eta = K^{-1}T^{-1/2}$ *and* $\epsilon = K^{3/2}T^{1/2}$.

*Proof sketch.* The proof of these bounds in the all-winner feedback follows closely the analysis used for the bandit feedback. Better bounds are achieved in comparison to the bandit's feedback thanks to the lower *variance* of the estimator defined in (10), proved in Lemma 10. Intuitively, this comes from the ability to observe the realized utility more often, allowed by the richer feedback. We exhibit this by using tools used in bandits with graph feedback (Alon et al., 2017). □

**Regret lower bound** We provide a matching lower bound on the regret of any online learning algorithm (Lemma 5), in the bandit feedback setting, by extending a result from (Balseiro et al., 2019) in the context of single price auctions. This partially answers an open question raised by Branzei et al., 2024 regarding the achievable learning rate in the bandit setting for the problem that we consider.

**Lemma 5.** *Any online learning procedure must incur* $\Omega(T^{2/3})$ *regret in multi-unit uniform auction with the Last Accepted Bid rule under bandit feedback Bid pricing rule.*

This stems from the fact that against an adversary that only plays bids with value 1 except for its last bid, the auction is essentially a first-price auction.

*Proof.* We extend the lower bound on the regret of the first price auction in Balseiro et al., 2019. At time $t$, let $\boldsymbol{\beta}^t = \{1, 1, ..., 1, h^t\}$ be the bid of the adversary, let the valuation of the learner be $v = (1, 0, .., 0)$ and denote $\mathbf{b}^t = (b_1^t, ..., b_k^t) \in B$ the learner's bid. We only consider sensible bids, such that $b_i^t > 0 \iff i = 1$, because they are dominating strategies. The learner's utility is $u(\mathbf{h}, \boldsymbol{\beta}) = \mathbb{1}\{b_1^t > h^t\}(1 - b_1^t)$, and the bandits feedback, $(x^t, \mathbb{1}\{x^t > 0\}p^t)$ is : $(\mathbb{1}\{b_1^t > h^t\}, \mathbb{1}\{b_1^t > h^t\}b_1^t)$ which coincides with both the utility and the bandits feedback of the first price auction with value 1.

This specific instance of the repeated $K$-unit auction therefore coincides with the repeated first price auction with opposing bid $h^t$ at time $t$, which can be any instance of the first price auction. Therefore if no learning algorithm can guarantee better regret than $\mathcal{O}(T^{2/3})$ in the latter problem, no algorithm can guarantee better regret than $\mathcal{O}(T^{2/3})$ in the former.

□

# 4 Conclusion

We provided the first no-regret algorithm achieving optimal rates in $T$ for the $K$-unit uniform auction under bandit, full information, and all winner feedback. The techniques and theoretical tools presented can be applied to obtain similar regret guarantees in the adversarial bid setting with random valuation, under the assumption that the valuation and opposing bids are independent. An interesting open question is whether similar rates can be achieved in a contextual setting (when valuations changing at each round are observed before each play). The obtained regret rates match the ones obtained in the discriminatory price auction, a commonly compared auction mechanism, up to a factor $\mathcal{O}(K^{\frac{1}{3}})$. This raises the question of whether this gap can be closed or if a lower bound showing a separation in achievable regret rates exists.


## Acknowledgements

Dorian Baudry thanks the support of the French National Research Agency: ANR-19-CHIA-02 SCAI, ANR-22-SRSE-0009 Ocean, and ANR-23-CE23-0002 Doom. Dorian Baudry was also partially funded by UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee [grant number EP/Y028333/1].

This research was supported in part by the French National Research Agency (ANR) in the framework of the PEPR IA FOUNDRY project (ANR-23-PEIA-0003) and through the grant DOOM ANR-23-CE23-0002. It was also funded by the European Union (ERC, Ocean, 101071601). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.


## References


Akbari-Dibavar, A., Mohammadi-Ivatloo, B., & Zare, K. (2020). Electricity market pricing: Uniform pricing vs. pay-as-bid pricing. *Electricity Markets: New Players and Pricing Uncertainties*, 19–35.

Alon, N., Cesa-Bianchi, N., Gentile, C., Mannor, S., Mansour, Y., & Shamir, O. (2017). Nonstochastic multi-armed bandits with graph-structured feedback. *SIAM Journal on Computing*, *46*(6), 1785–1826. https://doi.org/10.1137/140989455

Anderson, E., & Holmberg, P. (2018). Price instability in multi-unit auctions. *Journal of Economic Theory*, *175*, 318–341.

Audibert, J.-Y., Bubeck, S., & Lugosi, G. (2014). Regret in online combinatorial optimization. *Mathematics of Operations Research*, *39*(1), 31–45.

Ausubel, L. M., Cramton, P., Pycia, M., Rostek, M., & Weretka, M. (2014). Demand Reduction and Inefficiency in Multi-Unit Auctions. *The Review of Economic Studies*, *81*(4), 1366–1400. https://doi.org/10.1093/restud/rdu023

Balseiro, S., Golrezaei, N., Mahdian, M., Mirrokni, V., & Schneider, J. (2019). Contextual bandits with cross-learning. *Advances in Neural Information Processing Systems*, *32*.

Birmpas, G., Markakis, E., Telelis, O., & Tsikiridis, A. (2019). Tight welfare guarantees for pure nash equilibria of the uniform price auction. *Theory of Computing Systems*, *63*, 1451–1469.

Blum, A., Kumar, V., Rudra, A., & Wu, F. (2004). Online learning in online auctions. *Theoretical Computer Science*, *324*(2-3), 137–146.

Branzei, S., Derakhshan, M., Golrezaei, N., & Han, Y. (2024). Learning and collusion in multi-unit auctions. *Advances in Neural Information Processing Systems*, *36*.

Cesa-Bianchi, N., Cesari, T., Colomboni, R., Fusco, F., & Leonardi, S. (2023). The role of transparency in repeated first-price auctions with unknown valuations. *arXiv preprint arXiv:2307.09478*.

Cesa-Bianchi, N., Gentile, C., & Mansour, Y. (2014). Regret minimization for reserve prices in second-price auctions. *IEEE Transactions on Information Theory*, *61*(1), 549–564.

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, learning, and games*. Cambridge university press.

Cramton, P., & Stoft, S. (2006). *Uniform-price auctions in electricity markets*. Elsevier Science.

De Keijzer, B., Markakis, E., Schäfer, G., & Telelis, O. (2013). Inefficiency of standard multi-unit auctions. *European Symposium on Algorithms*, 385–396.

Fabra, N., von der Fehr, N.-H., & Harbord, D. (2006). Designing electricity auctions. *The RAND Journal of Economics*, *37*(1), 23–46.

Feng, Z., Podimata, C., & Syrgkanis, V. (2018). Learning to bid without knowing your value. *Proceedings of the 2018 ACM Conference on Economics and Computation*, 505–522. https://doi.org/10.1145/3219166.3219208

Flajolet, A., & Jaillet, P. (2017). Real-time bidding with side information. *Neural Information Processing Systems*. https://api.semanticscholar.org/CorpusID:1863793

Galgana, R., & Golrezaei, N. (2023). Learning in repeated multi-unit pay-as-bid auctions. *arXiv preprint arXiv:2307.15193*.

Kanoria, Y., & Nazerzadeh, H. (2014). Dynamic reserve prices for repeated auctions: Learning from bids. *Web and Internet Economics: 10th International Conference, WINE 2014, Beijing, China, December 14-17, 2014. Proceedings 10*, 232–232.

Karaca, O., Sessa, P. G., Leidi, A., & Kamgarpour, M. (2020). No-regret learning from partially observed data in repeated auctions. *IFAC-PapersOnLine*, *53*(2), 14–19.

Khezr, P., & Cumpston, A. (2022). A review of multiunit auctions with homogeneous goods. *Journal of Economic Surveys*, *36*(4), 1225–1247.

Lattimore, T., & Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Nedelec, T., Calauzènes, C., El Karoui, N., Perchet, V., et al. (2022). Learning in repeated auctions. *Foundations and Trends® in Machine Learning*, *15*(3), 176–334.

Sessa, P. G., Walton, N., & Kamgarpour, M. (2017). Exploring the vickrey-clarke-groves mechanism for electricity markets. *IFAC-PapersOnLine*, *50*(1), 189–194.

Son, Y. S., Baldick, R., Lee, K.-H., & Siddiqi, S. (2004). Short-term electricity market auction game analysis: Uniform and pay-as-bid pricing. *IEEE Transactions on Power Systems*, *19*(4), 1990–1998.

Takimoto, E., & Warmuth, M. K. (2002). Path kernels and multiplicative updates. *International Conference on Computational Learning Theory*, 74–89.

Viehmann, J., Lorenczik, S., & Malischek, R. (2021). Multi-unit multiple bid auctions in balancing markets: An agent-based q-learning approach. *Energy economics*, *93*, 105035.

Weed, J., Perchet, V., & Rigollet, P. (2016). Online learning in repeated auctions. *Conference on Learning Theory*, 1562–1583.

Willems, B., & Yu, Y. (2022). Bidding and investment in wholesale electricity markets: Pay-as-bid versus uniform-price auctions. *Tilburg University, Toulouse School of Economics and CERRE*.

# A Problem-specific simplifications

This section focuses on characterizing undominated strategies and showing that when a learner plays on an $\epsilon$-discretization of the bid interval $[0, 1]$ assuming ties never occur is without loss of generality.

## A.1 Dominated strategies

Since the scale of the bid space is a deciding factor in the rates of regret we obtain, it is useful to analyze the utility functions of the learner. Indeed, for the learning procedure, it can be useful, to restrict ourselves from the start to bids which can potentially be optimal. We show next that, under certain condition on the values of the learner, we can restrict the bid space $B$.

**Lemma 6.** *Let $\{v_1, v_2, ..., v_k\}$ be the valuations of the learner. Then for any bids $\mathbf{b} = \{b_1, ..., b_K\} \in B$ such that there exists $i \in [K], b_i > v_i$, there exists $\tilde{\mathbf{b}}$ a non-increasing sequence of $[0, v_1] \times [0, v_2] \times ... \times [0, v_K]$ such that :*

$$\forall \boldsymbol{\beta} \in B, u(\tilde{\mathbf{b}}, \boldsymbol{\beta}) \geq u(\mathbf{b}, \boldsymbol{\beta})$$

*Proof.* Let $\{v_1, v_2, ..., v_k\}$ be the valuations of the learner. Let $\mathbf{b} = \{b_1, ..., b_K\} \in B$ a bid such that there exists $i \in [K], b_i > v_i$. We define $\tilde{\mathbf{b}}$ as follows:

$$\forall i \in [K], \tilde{\mathbf{b}}_i = \min(v_i, b_i)$$

Let $\boldsymbol{\beta} \in B$ be the bids of the adversary. As a max, $p(.)$ is increasing in its arguments hence

$$p(\tilde{\mathbf{b}}, \boldsymbol{\beta}) \leq p(\mathbf{b}, \boldsymbol{\beta}). \tag{21}$$

There are two possible cases : Either the allocation remains the same $x(\tilde{\mathbf{b}}, \boldsymbol{\beta}) = x(\mathbf{b}, \boldsymbol{\beta})$, or it decreases $x(\tilde{\mathbf{b}}, \boldsymbol{\beta}) \leq x(\mathbf{b}, \boldsymbol{\beta})$ (decreasing bids cannot result in a increased allocation).

If the allocation remains the same, then (21) implies $u(\tilde{\mathbf{b}}, \boldsymbol{\beta}) \geq u(\mathbf{b}, \boldsymbol{\beta})$.

If the allocation decreases, the items obtained when bidding $\mathbf{b}$ but not obtained when bidding $\tilde{\mathbf{b}}$ necessarily corresponds to bids which have been lowered (other bids remain higher than the price). Let $j \in [K]$ such that $b_j$ is one of these bids, since the item used to be won $b_j \geq p(\mathbf{b}, \boldsymbol{\beta})$. Since it is not won by the learner playing $\tilde{\mathbf{b}}$, we have $\tilde{b}_j \leq p(\tilde{\mathbf{b}}, \boldsymbol{\beta})$. Hence $v_j \leq p(\mathbf{b}, \boldsymbol{\beta})$.

Since this is the case for all items $j$ won under $\mathbf{b}$ and not under bidding $p(\mathbf{b}, \boldsymbol{\beta})$, $u(\tilde{\mathbf{b}}, \boldsymbol{\beta}) \geq u(\mathbf{b}, \boldsymbol{\beta})$.

It is therefore always the case that

$$u(\tilde{\mathbf{b}}, \boldsymbol{\beta}) \geq u(\mathbf{b}, \boldsymbol{\beta}) \tag{22}$$

Since it is true for all $\boldsymbol{\beta} \in B$, this concludes the proof. □

A consequence of Lemma 6 is that we can restrict our learning procedure to non-increasing sequence of $[0, v_1] \times [0, v_2] \times ... \times [0, v_K]$.

In the following of the paper for simplicity we use $B$ as the bidding space. This covers the *worst case* which corresponds to the case when for all $i \in [K]$, $v_i = 1$.

## A.2 Discretization error

To use online learning techniques in our instance of multi-unit uniform price auction, we use a discretization $B_\epsilon$ of the bid space $B$ which is continuous. We bound here the added regret incurred because of this discretization, that is the additional regret suffered when comparing the best action in hindsight of $B_\epsilon$ to the best of $B$.

**Definition A.1** (Discretized regret)**.** *Let $(\mathbf{b}^t)_{t\in[T]} \in B_\epsilon^T$ be the action played at time $t \in [T]$, against the opposing bids $\left(\boldsymbol{\beta}^t\right)_{t\in[T]} \in B^T$. The discretized regret is defined as follows:*

$$R_{T,\epsilon} = \max_{\mathbf{b}\in B_\epsilon} \sum_{t=1}^{T} u(\mathbf{b}, \boldsymbol{\beta}^t) - \mathbb{E}\left[\sum_{t=1}^{T} u(\mathbf{b}^t, \boldsymbol{\beta}^t)\right]. \tag{23}$$

We bound the cost of this discretization as follows :

**Lemma 7.** *Let $(\mathbf{b}^t)_{t\in[T]} \in B_\epsilon^T$ be the action played at time $t \in [T]$, against the opposing bids $\left(\boldsymbol{\beta}^t\right)_{t\in[T]} \in B^T$. With $R_{T,\epsilon}$ the dicretized regret and $R_T$ the regret, we have the following inequality:*

$$R_T \leq R_{T,\epsilon} + KT\epsilon\,.$$

*Proof of Lemma 7.* Let $\left(\boldsymbol{\beta}^t\right)_{t\in[T]} \in \left([0,1]^K\right)^T$ the adversary bids played up to time $T$.
Let $(\mathbf{b}^t)_{t\in[T]} \in B_\epsilon^T$.

$$\begin{aligned}
R_T &= \sup_{\mathbf{b}\in B} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) - \mathbb{E}\left[\sum_{t=1}^T u(\mathbf{b}^t, \boldsymbol{\beta}^t)\right] \\
&= \sup_{\mathbf{b}\in B} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) - \sup_{\mathbf{b}\in B_\epsilon} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) + \sup_{\mathbf{b}\in B_\epsilon} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) - \mathbb{E}\left[\sum_{t=1}^T u(\mathbf{b}^t, \boldsymbol{\beta}^t)\right] \\
&= \sup_{\mathbf{b}\in B} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) - \max_{\mathbf{b}\in B_\epsilon} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) + R_{T,\epsilon}\,.
\end{aligned}$$

Let $\mu > 0$, there exists $\mathbf{b}_{opt} \in B$ such that $\sum_{t=1}^T u(\mathbf{b}_{opt}, \boldsymbol{\beta}^t) + \mu \geq \sup_{\mathbf{b}\in B} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t)$.

We define its closest discretized bid from above

$$\mathbf{b}_{opt,\epsilon} := \left( \begin{cases} b_{opt,i} & \text{if } \frac{b_{opt,i}}{\epsilon} \in \mathbb{N} \\ 1 & \text{if } b_{opt,i} > \left\lfloor \frac{1}{\epsilon} \right\rfloor \epsilon \\ \left\lceil \frac{b_{opt,i}}{\epsilon} \right\rceil \epsilon & \text{else} \end{cases} \right)_{i\in[K]} . \tag{24}$$

Since $\max(.)$ cannot increase more than its arguments and $\forall i \in [K]$, $b_{opt,\epsilon,i} \leq b_{opt,i} + \epsilon$,

$$\forall t \in [T],\ p(\mathbf{b}_{opt,\epsilon}, \boldsymbol{\beta}^t) \leq p(\mathbf{b}_{opt}, \boldsymbol{\beta}^t) + \epsilon\,, \tag{25}$$

and since $\forall i \in [K],\ b_{opt,\epsilon,i} \geq b_{opt,i}$,

$$\forall t \in [T],\ x(\mathbf{b}_{opt,\epsilon}, \boldsymbol{\beta}^t) \geq x(\mathbf{b}_{opt}, \boldsymbol{\beta}^t) \tag{26}$$

therefore,

$$\sum_{t=1}^T u(\mathbf{b}_{opt,\epsilon}, \boldsymbol{\beta}^t) \geq \sum_{t=1}^T u(\mathbf{b}_{opt}, \boldsymbol{\beta}^t) - KT\epsilon \geq \sup_{\mathbf{b}\in B} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) - KT\epsilon - \mu\,.$$

Hence

$$\begin{aligned}
R_T &= \sup_{\mathbf{b}\in B} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) - \max_{\mathbf{b}\in B_\epsilon} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) + R_{T,\epsilon} \\
&\leq \sup_{\mathbf{b}\in B} \sum_{t=1}^T u(\mathbf{b}, \boldsymbol{\beta}^t) - \sum_{t=1}^T u(\mathbf{b}_{opt,\epsilon}, \boldsymbol{\beta}^t) + R_{T,\epsilon} \\
&\leq KT\epsilon + \mu + R_{T,\epsilon}.
\end{aligned}$$

Since the previous inequality is true for any $\mu > 0$, we get

$$R_T \leq KT\epsilon + R_{T,\epsilon}.$$

□

### A.3 Avoiding ties

In the latter analysis, we assume that ties never occur. We show here how this assumption, for our regret analysis, is equivalent to using a small perturbation of the bids of the learner. Let $\delta \in (0, \epsilon)$ and $X \sim \mathcal{U}[0, \delta]$ the random perturbation of the bids of the learner. We define $B_\epsilon^X$ the set of non-increasing sequences of $\{X, \epsilon + X, 2\epsilon + X, ..., 1 - \epsilon + X\}^K$, the set in which the perturbed bids of learners take value.

This perturbation of the discretized set we use as bid space for the learner comes at no costs in terms of added regrets. The previous Lemma 7 can straightforwardly be applied to the perturbed set $B_\epsilon^X$, as the key reason for the additional regret is the discretization step $\epsilon$, which remains unchanged here.

**Lemma 8.** *Let $\boldsymbol{\beta}^T \in B^T$ be the bid of the adversary up to time $T$, for any bid sequence of the learner $\mathbf{b}^T \in B_\epsilon^{X \times T}$, there is almost surely never a tie.*

*Proof.* Let $\boldsymbol{\beta}^T \in B^T$,and for all $t \in T$ denote $\boldsymbol{\beta}^t = \{\beta_1^t, ..., \beta_K^t\}$. For any bid sequence $\mathbf{b}^T \in B_\epsilon^{X \times T}$, a necessary condition for ties to occur is that there exists $(t, j) \in [T] \times [K]$ such that $\beta_j^t \in \{\epsilon + X, 2\epsilon + X, ..., 1 - \epsilon + X\}$. This is almost surely never the case, as it is the probability of $X$ to belong to a finite set. □

**Remark 2.** *For the sake of regret bounds, since we almost surely don't have any tie, the assumption that the adversary plays bids in $B_{\setminus \epsilon}$ is without loss of generality.*

## B Appendix: Regret analysis

### B.1 Full-information regret rates

**Theorem 3.** *In the full information feedback setting, Algorithm 1 coincides with the Hedge algorithm with parameter $\eta$ on $H_\epsilon$. It ensures a regret of at most $O\left(\sqrt{K^3 T \log(T)}\right)$ by taking $\epsilon = \sqrt{\frac{K}{T}}$ and $\eta = \sqrt{\frac{\log\left(\frac{T}{K}\right)}{2KT}}$.*

*Proof of Theorem 3.* In order to leverage classical results on the exponential weight algorithm in the expert setting, we aim to show that the combination of our algorithms leads to a probability update rule of the form:

$$\mathbb{P}^t(\mathbf{h}) = \frac{\mathbb{P}^{t-1}(\mathbf{h}) \exp\left(\eta u^t(\mathbf{h})\right)}{\sum_{\mathbf{j} \in H_\epsilon} \mathbb{P}^{t-1}(\mathbf{j}) \exp \eta u^t(\mathbf{j})}\,. \tag{27}$$

Let $\mathbf{h}$ be bid profile in $H_\epsilon$, we denote $h_i$ its $i^{\text{th}}$ element ($i^{\text{th}}$ element of the sequence) and $\text{len}(\mathbf{h})$ the length of the sequence $\mathbf{h}$. Given the sampling Algorithm 2, we have by telescoping and the product of conditional probabilities

$$\begin{aligned}
\mathbb{P}^t(\mathbf{h}) &= \prod_{i \in [1, \text{len}(\mathbf{h})]} \mathbb{P}^t(h_i | h_{i-1}) \\
&= \prod_{i \in [1, \text{len}(\mathbf{h})]} Q^t(h_i) \frac{\Gamma^t(h_i)}{\Gamma^t(h_{i-1})} \\
&= \frac{\prod_{i \in [1, \text{len}(\mathbf{h})]} Q^t(h_i)}{\Gamma_0^t} \\
&= \frac{\prod_{i \in [1, \text{len}(\mathbf{h})]} Q^{t-1}(h_i) \exp\left(\eta w^t(h_i)\right)}{\Gamma_0^t} \\
&= \frac{\exp\left(\eta u^t(\mathbf{h})\right)}{\Gamma_0^t} \prod_{i \in [1, \text{len}(\mathbf{h})]} \mathbb{P}^{t-1}(h_i | h_{i-1}) \frac{\Gamma^{t-1}(h_{i-1})}{\Gamma^{t-1}(h_i)}\,. \\
\mathbb{P}^t(\mathbf{h}) &= \frac{\Gamma_0^{t-1}}{\Gamma_0^t} \exp\left(\eta u^t(\mathbf{h})\right) \mathbb{P}^{t-1}(\mathbf{h})\,,
\end{aligned} \tag{28}$$

where we used the probability update rule (12).

We can prove that $\Gamma_0^t = \sum_{\mathbf{h}\in H_\epsilon} \prod_{i\in[1,\text{len}(\mathbf{h})]} Q^t(h_i)$. An induction on $k \in \left[\max_{\mathbf{h}\in H_\epsilon} \text{len}(\mathbf{h})\right]$, where we denote $\mathbf{h}_{:k}$ the $k$ first component of the sequence $\mathbf{h}$:

$$\Gamma_0^t = \sum_{\mathbf{h}_{:k}:\mathbf{h}\in H_\epsilon} \left( \prod_{i\in[1,\min\{k,\text{len}(\mathbf{h})\}]} Q^t(h_i) \right) \Gamma^t(h_{\min\{k,\text{len}(\mathbf{h})\}})$$

This is true for $k = 1$ from the definition (2) of $\Gamma_0^t$.

For $k \geq 1$,

$$\begin{aligned}
\Gamma_0^t &= \sum_{\mathbf{h}_{:k}:\mathbf{h}\in H_\epsilon} \left( \prod_{i\in[1,\min\{k,\text{len}(\mathbf{h})\}]} Q^t(h_i) \right) \Gamma^t(h_{\min\{k,\text{len}(\mathbf{h})\}}) \\
&= \sum_{\mathbf{h}_{:k}:\mathbf{h}\in H_\epsilon} \left( \prod_{i\in[1,\min\{k,\text{len}(\mathbf{h})\}]} Q^t(h_i) \right) \sum_{h\in s(h_k)} Q^t(h)\Gamma^t(h) \\
&= \sum_{\mathbf{h}_{:k+1}:\mathbf{h}\in H_\epsilon} \left( \prod_{i\in[1,\min\{k+1,\text{len}(\mathbf{h})\}]} Q^t(h_i) \right) \Gamma^t(h_{\min\{k+1,\text{len}(\mathbf{h})\}}) \,,
\end{aligned}$$

where we used the fact that the function successor $s(.)$ provides all possible next element of sequence $\mathbf{h}$.

We then simplify $\frac{\Gamma_0^{t-1}}{\Gamma_0^t}$:

$$\begin{aligned}
\Gamma_0^t &= \sum_{\mathbf{h}\in H_\epsilon} \prod_{i\in[1,\text{len}(\mathbf{h})]} Q^t(h_i) \\
&= \sum_{\mathbf{h}\in H_\epsilon} \prod_{i\in[1,\text{len}(\mathbf{h})]} Q^{t-1}(h_i) \exp\left(\eta u^t(h_i)\right) \\
&= \sum_{\mathbf{h}\in H_\epsilon} \exp\left(\eta u^t(\mathbf{h})\right) \prod_{i\in[1,\text{len}(\mathbf{h})]} \mathbb{P}_{t-1}(h_i|\mathbf{h}_{i-1}) \frac{\Gamma^{t-1}(\mathbf{h}_{i-1})}{\Gamma^{t-1}(h_i)} \\
&= \Gamma_0^{t-1} \sum_{\mathbf{h}\in H_\epsilon} \exp\left(\eta u^t(\mathbf{h})\right) \mathbb{P}_{t-1}(\mathbf{h}) \,.
\end{aligned}$$

We can therefore write

$$\mathbb{P}^t(\mathbf{h}) = \frac{\mathbb{P}^{t-1}(\mathbf{h}) \exp\left(\eta u^t(\mathbf{h})\right)}{\sum_{\mathbf{l}\in H_\epsilon} \mathbb{P}^{t-1}(\mathbf{l}) \exp \eta u^t(\mathbf{l})} \,. \tag{29}$$

This is the update rule of the Hedge algorithm on the action space $H_\epsilon$. Therefore using Theorem 2.2 of Cesa-Bianchi and Lugosi, 2006, restated in the Appendix 4 leads to the following regret bound:

$$\begin{aligned}
R_{T,\epsilon} &\leq \frac{\log\left(|B_\epsilon|\right)}{\eta} + \frac{\eta T K^2}{8} \\
&\leq \frac{\log\left(1/\epsilon^K\right)}{\eta} + \frac{\eta T K^2}{8} \,,
\end{aligned} \tag{30}$$

where, to bound $\log\left(|B_\epsilon|\right)$ in (30), we used the fact that the action space $H_\epsilon$ is in bijection with the original discretized bid space : $K$ non-increasing elements of $\{1, 2, \ldots, \lfloor \frac{1}{\epsilon} \rfloor\}$, which cardinal is trivially smaller than $\frac{1}{\epsilon^K}$.

Taking $\eta = \sqrt{\frac{\log\left(\frac{1}{\epsilon}\right)}{KT}}$, we obtain

$$R_{T,\epsilon} \leq \frac{9}{8}\sqrt{TK^3 \log\left(\frac{1}{\epsilon}\right)}\,.$$

Finally, using lemma 7,

$$\begin{aligned} R_T &\leq R_{T,\epsilon} + KT\epsilon \\ &\leq \frac{9}{8}\sqrt{TK^3 \log\left(\frac{1}{\epsilon}\right)} + KT\epsilon \\ &\leq \frac{9}{8}\sqrt{TK^3}\left(\log\left(\sqrt{\frac{T}{K}}\right) + 1\right)\,, \end{aligned}$$

with $\epsilon = \sqrt{\frac{K}{T}}$. □

### B.2 Partial feedback regret rates

#### B.2.1 Bandit feedback

To prove the regret rates in the bandit feedback settings, we use Lemma 9 which bounds the necessary quantity for the standard EXP3 analysis.

**Lemma 9** (Bandit feedback estimator)**.** *The estimator $\hat{u}^t(\mathbf{h})$ defined by* (19) *has the following properties:*

- *The estimator $\hat{u}^t(\mathbf{h})$ has a fixed bias $-K$:*

$$\mathbb{E}\left[\hat{u}^t(\mathbf{h})\right] = u^t(\mathbf{h}) - K\,.$$

- *The square of the estimator can be upper bounded as follows:*

$$\sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h})\mathbb{E}\left[\hat{u}^t(\mathbf{h})^2\right] \leq 4K^2 \max\left(K^2, \frac{1}{\epsilon}\right)\,.$$

Since the estimator has a constant bias for every action, one can use it in the problem similarly to an unbiased estimator.

*Proof of Lemma 9.* Let $h$ be a bid or a bid-gap, then

$$\begin{aligned} \mathbb{E}\left[\hat{u}^t(\mathbf{h})\right] &= \sum_{\mathbf{h}^t\in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t) \sum_{h\in\mathbf{h}} \hat{w}^t(h) \\ &= \sum_{\mathbf{h}^t\in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t) \sum_{h\in\mathbf{h}} \mathbf{1}(h = h^t_\star)\frac{w^t(h) - K}{\mathbb{P}^t(h)} \\ &= \sum_{\mathbf{h}^t\in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t)\frac{w^t(h^t_\star(\mathbf{h})) - K}{\mathbb{P}^t(h^t_\star(\mathbf{h}))}\mathbf{1}(h^t_\star(\mathbf{h}) = h^t_\star) \\ &= \frac{w^t(h^t_\star(\mathbf{h})) - K}{\mathbb{P}^t(h^t_\star(\mathbf{h}))} \sum_{\mathbf{h}^t\in H_\epsilon : h^t_\star(\mathbf{h})\in\mathbf{h}^t} \mathbb{P}^t(\mathbf{h}^t) \\ &= w^t(h^t_\star(\mathbf{h})) - K = u^t(\mathbf{h}) - K\,. \end{aligned}$$

$$\sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h})\mathbb{E}\left[\hat{u}^t(\mathbf{h})^2\right] = \mathbb{E}\left[\sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h})^2\right] \tag{31}$$

$$= \sum_{\mathbf{h}^t\in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t) \sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \left(\frac{w^t(h^t_\star(\mathbf{h})) - K}{\mathbb{P}^t(h^t_\star(\mathbf{h}))}\right)^2 \mathbf{1}(h^t_\star(\mathbf{h}) = h^t_\star) \tag{32}$$

$$= \sum_{\mathbf{h}^t\in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t) \left(\frac{w^t(h^t_\star) - K}{\mathbb{P}^t(h^t_\star)}\right)^2 \sum_{\mathbf{h}\in H_\epsilon : h^t_\star\in\mathbf{h}} \mathbb{P}^t(\mathbf{h}) \tag{33}$$

$$= \sum_{\mathbf{h}^t\in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t) \frac{\left(w^t(h^t_\star) - K\right)^2}{\mathbb{P}^t(h^t_\star)} \tag{34}$$

$$\leq K^2 \sum_{h^t_\star\in\mathcal{O}^t} \frac{\sum_{\mathbf{h}^t\in H_\epsilon : h^t_\star\in\mathbf{h}^t} \mathbb{P}^t(\mathbf{h}^t)}{\mathbb{P}^t(h^t_\star)} \tag{35}$$

$$\leq K^2 \sum_{h^t_\star\in\mathcal{O}^t} 1 \tag{36}$$

$$\leq (K)^2 |\mathcal{O}^t|\,, \tag{37}$$

where

$$\mathcal{O}^t = \left\{ h^t_\star(\mathbf{h}) \middle| \mathbf{h}\in H_\epsilon \right\}.$$

There only remains to upper bound $|\mathcal{O}^t|$.

For any $p\in[0,1]$ we denote for this proof $j(p) = \lfloor \frac{p}{\epsilon} \rfloor$, which is the value $j$ such that $p\in[j\epsilon, j\epsilon+\epsilon)$. Let $\mathbf{h}\in H_\epsilon$, notice that $h^t_\star(\mathbf{h},\boldsymbol{\beta}^t) \in \{b_{x_t(\mathbf{h},\boldsymbol{\beta}^t),j(p(\mathbf{h},\boldsymbol{\beta}^t))},\ b_{x_t(\mathbf{h},\boldsymbol{\beta}^t)+\frac{1}{2},j(p(\mathbf{h},\boldsymbol{\beta}^t))}\}$ which directly results from the decomposition formula 10. To upper bound $|\mathcal{O}^t|$ we will therefore upper bound the different values the pair $\left(x_t(\,.\,,\boldsymbol{\beta}^t), p(\,.\,,\boldsymbol{\beta}^t)\right)$ can take.

Since the learner plays bids in $H_\epsilon$ (or equivalently $B_\epsilon$), $p(.,\boldsymbol{\beta}^t)$ can only take either the value of one of the components in $\boldsymbol{\beta}$ or one of the bids of its first argument.
Because $\boldsymbol{\beta}^t$ is a vector of size K we can write that: $\left|\{p(\mathbf{h},\boldsymbol{\beta}^t), \mathbf{h}\in H_\epsilon\}\right| \leq K + \lfloor\frac{1}{\epsilon}\rfloor$.

Furthermore, because units are either attributed to the player or the adversary, we can write $K - \left|\{\beta^t_i > p_t(\,.\,,\boldsymbol{\beta}^t)\}\right| \geq x_t(\,.\,,\boldsymbol{\beta}^t) \geq K - \left|\{\beta^t_i \geq p_t(\,.\,,\boldsymbol{\beta}^t)\}\right|$. The two cardinals can only differ if the price is set by an adversary bid because the no ties assumption implies almost surely for all $i\in[K], \beta^t_i \notin \left[\frac{1}{\epsilon}\right]$.

Therefore each possible value of $p_t(\,.\,,\boldsymbol{\beta}^t)$ only correspond to one value of $x_t(\,.\,,\boldsymbol{\beta}^t)$, except for the $K$ values set by the adversary, where $x_t(\,.\,,\boldsymbol{\beta}^t)$ can at most take $K$ values.

Therefore

$$2\left(K^2 + \left\lfloor\frac{1}{\epsilon}\right\rfloor\right) \geq 2\left|\left\{\left(x_t(\mathbf{h},\boldsymbol{\beta}^t), p(\mathbf{h},\boldsymbol{\beta}^t)\right), \mathbf{h}\in H_\epsilon\right\}\right| \tag{38}$$

$$\geq 2\left|\left\{\left(x_t(\mathbf{h},\boldsymbol{\beta}^t), j\left(p(\mathbf{h},\boldsymbol{\beta}^t)\right)\right), \mathbf{h}\in H_\epsilon\right\}\right| \tag{39}$$

$$\geq \left|\mathcal{O}^t\right|, \tag{40}$$

which leads to the needed upper bounds

□

We now restate Theorem 1 and then provide proof of the corresponding regret guarantees.

**Theorem 1.** *In the repeated $K$-unit auction with uniform pricing guarantees and under bandit feedback, Algorithm 1 incurs a regret of at most $\mathcal{O}\left(K^{4/3}T^{2/3}\log(T)\right)$. For any time horizon $T$, with the choices of $\epsilon = \left(\frac{K}{T}\right)^{1/3}$ and $\eta = K^{-1/3}T^{-2/3}\sqrt{\log\left(\frac{T}{K}\right)/3}$.*

*Proof of Theorem 1.* For $T \in \mathbf{N}$, we denote $(\mathbf{h}^t)_{t\in[T]} \in H_\epsilon^T$ the actions played at each time-steps, generated by Algorithm 1.
We can first notice that, by conducting the same analysis as in B.1 up to Equation (29), we obtain:

$$\mathbb{P}^t(\mathbf{h}) = \frac{\mathbb{P}_{t-1}(\mathbf{h}) \exp\left(\eta \hat{u}^t(\mathbf{h})\right)}{\sum_{\mathbf{l}\in H_\epsilon} \mathbb{P}_{t-1}(\mathbf{l}) \exp \eta \hat{u}^t(\mathbf{l})} \,, \tag{41}$$

which by simple induction allows us to obtain:

$$\mathbb{P}^t(\mathbf{h}) = \frac{\exp\left(\sum_{j=1}^t \eta \hat{u}^j(\mathbf{h})\right)}{\sum_{\mathbf{l}\in H_\epsilon} \exp\left(\sum_{j=1}^t \eta \hat{u}^j(\mathbf{l})\right)} \,. \tag{42}$$

We can now proceed to the regret analysis.
For any action $\mathbf{h} \in H_\epsilon$, we define :

$$R_{T,\mathbf{h}} = \sum_{t=1}^T u^t(\mathbf{h}) - \mathbb{E}\left[\sum_{t=1}^T u^t(\mathbf{h}^t)\right] \,,$$

which is the expected regret relative to playing $\mathbf{h}$ in all the rounds.

We have, because of Lemma 9, $\mathbb{E}\left[\sum_{t=1}^T \hat{u}^t(\mathbf{h})\right] = \sum_{t=1}^T u^t(\mathbf{h}^t) - KT$
and $\mathbb{E}_{t-1}\left[u^t(\mathbf{h}^t)\right] = \sum_{\mathbf{h}\in B} \mathbb{P}^t(\mathbf{h}) u^t(\mathbf{h}) = \sum_{\mathbf{h}\in B} \mathbb{P}^t(\mathbf{h}) \mathbb{E}_{t-1}\left[\hat{u}^t(\mathbf{h})\right] + K$.
Therefore

$$R_{T,\mathbf{h}} = \mathbb{E}\left[\sum_{t=1}^T \hat{u}^t(\mathbf{h})\right] - \mathbb{E}\left[\sum_{t=1}^T \sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \hat{u}^t(\mathbf{h})\right] \,. \tag{43}$$

We denote $W_n = \sum_{\mathbf{h}\in H_\epsilon} \exp(\eta \sum_{t=1}^n \hat{u}^t(\mathbf{h}))$.
Then we have for any $\mathbf{h} \in H_\epsilon$,

$$\exp\left(\eta \sum_{t=1}^T \hat{u}^t(\mathbf{h})\right) \leq \sum_{\mathbf{h}\in H_\epsilon} \exp(\eta \sum_{t=1}^T \hat{u}^t(\mathbf{h})) = W_T = W_0 \prod_{t=1}^T \frac{W_t}{W_{t-1}} \,.$$

We can then upper bound the terms of the product as follows :

$$\begin{aligned}
\frac{W_t}{W_{t-1}} &\leq \sum_{\mathbf{h}\in H_\epsilon} \frac{\exp\left(\eta \sum_{l=1}^{t-1} \hat{u}^l(\mathbf{h})\right)}{W_{t-1}} \exp\left(\eta \hat{u}^t(\mathbf{h})\right) \\
&\leq \sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \exp\left(\eta \hat{u}^t(\mathbf{h})\right) \,,
\end{aligned}$$

where the second inequality comes from 42.

We can then further bound this term using the inequalities

$$\forall\, x \leq 1,\ \exp(x) \leq 1 + x + x^2 \text{ and } \forall x \in \mathbb{R}, 1 + x \leq \exp(x) \,.$$

This gives

$$\frac{W_t}{W_{t-1}} \leq 1 + \eta \sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \hat{u}^t(\mathbf{h}) + \eta^2 \sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \hat{u}^t(\mathbf{h})^2 \tag{44}$$

$$\leq \exp\left(\eta \sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \hat{u}^t(\mathbf{h}) + \eta^2 \sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \hat{u}^t(\mathbf{h})^2\right) \,. \tag{45}$$

This in turn yields

$$\exp\left(\eta\sum_{t=1}^{T}\hat{u}^t(\mathbf{h})\right) \leq W_0\prod_{t=1}^{T}\exp\left(\eta\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h}) + \eta^2\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h})^2\right) \tag{46}$$

$$\leq W_0\exp\left(\eta\sum_{t=1}^{T}\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h}) + \eta^2\sum_{t=1}^{T}\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h})^2\right), \tag{47}$$

where by applying the log, simplifying, and taking the expectation we get

$$\sum_{t=1}^{T}\hat{u}^t(\mathbf{h}) - \sum_{t=1}^{T}\sum_{\mathbf{h}\in B}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h}) \leq \frac{\log(W_0)}{\eta} + \eta\sum_{t=1}^{T}\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h})^2 \tag{48}$$

$$\mathbb{E}\left[\sum_{t=1}^{T}\hat{u}^t(\mathbf{h}) - \sum_{t=1}^{T}\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h})\right] \leq \frac{\log(W_0)}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h})^2\right] \tag{49}$$

$$R_{T,\mathbf{h}} \leq \frac{\log(W_0)}{\eta} + \eta\mathbb{E}\left[\sum_{t=1}^{T}\sum_{\mathbf{h}\in B}\mathbb{P}^t(\mathbf{h})\hat{u}^t(\mathbf{h})^2\right]. \tag{50}$$

We recognize the expression of the regret from (43). Because it is true for all $\mathbf{h}\in H_\epsilon$, we can take the maximum and notice : $R_{T,\epsilon} = \max_{\mathbf{h}\in H_\epsilon} R_{T,\mathbf{h}}$. Noticing that $W_0 = |H_\epsilon| \leq \left(\frac{1}{\epsilon}\right)^K$ and using Lemma 9 concludes the bound on the discritzed regret as follows:

$$R_{T,\epsilon} \leq \frac{\log|H_\epsilon|}{\eta} + \eta\sum_{t=1}^{T}\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\mathbb{E}\left[\hat{u}^t(\mathbf{h})^2\right] \tag{51}$$

$$\leq K\frac{\log\left(\frac{1}{\epsilon}\right)}{\eta} + \eta 4K^2T\frac{1}{\epsilon} \tag{52}$$

$$\leq 5K\sqrt{\frac{KT}{\epsilon}\log\left(\frac{1}{\epsilon}\right)}, \tag{53}$$

with $\eta = \sqrt{\frac{\epsilon}{KT}\log\left(\frac{1}{\epsilon}\right)}$

Then using Lemma 7, we can bound the regret:

$$R_T = R_{T,disc} + KT\epsilon \tag{54}$$

$$\leq 5K^{3/2}\sqrt{\frac{T}{\epsilon}\log\left(\frac{1}{\epsilon}\right)} + KT\epsilon \tag{55}$$

$$\leq 5K^{4/3}T^{2/3}\left(1 + \frac{1}{3}\log\left(\frac{T}{K}\right)\right), \tag{56}$$

with the specific choice of $\epsilon = \left(\frac{K}{T}\right)^{1/3}$. □

#### B.2.2 All-winner feedback

As in the bandit feedback, to prove the regret rates, we use Lemma 10 which bounds the necessary quantity for the standard EXP3 analysis.

**Lemma 10.** *The estimator $\bar{u}^t(\mathbf{h})$, defined by* (20) *has the following properties:*

- *The estimator $\bar{u}^t(\mathbf{h})$ has a fixed bias $-K$:*

$$\mathbb{E}\left[\bar{u}^t(\mathbf{h})\right] = u^t(\mathbf{h}) - K. \tag{57}$$

- *The square of the estimator verifies:*

$$\sum_{\mathbf{h}\in H_\epsilon}\mathbb{P}^t(\mathbf{h})\mathbb{E}\left[\bar{u}^t(\mathbf{h})^2\right] \leq 8K^4\log(2). \tag{58}$$

*Proof of Lemma 10.* This proof is mostly based on the following careful computations.

$$\begin{aligned}
\mathbb{E}\left[\bar{u}^t(\mathbf{h})\right] &= \sum_{\mathbf{h}^t \in B} \mathbb{P}^t(\mathbf{h}^t) \sum_{h \in \mathbf{h}} \bar{w}^t(h) \\
&= \sum_{\mathbf{h}^t \in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t) \sum_{h_{k,j} \in \mathbf{h}} \mathbf{1}\left(h \in A^t_\star(\mathbf{h}^t)\right) \frac{w^t(h) - K}{\mathbb{P}_{\mathbf{l}^t \sim \mathcal{B}^t}(h \in A^t_\star(\mathbf{l}^t))} \qquad (59) \\
&= \sum_{\mathbf{h}^t \in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t) \frac{w^t(h^t_\star(\mathbf{h})) - K}{\mathbb{P}_{\mathbf{l}^t \sim \mathcal{B}^t}(h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{l}^t))} \mathbf{1}(h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{h}^t)) \\
&= \frac{w^t(h^t_\star(\mathbf{h})) - K}{\mathbb{P}_{\mathbf{l}^t \sim \mathcal{B}^t}(h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{l}^t))} \sum_{\mathbf{h}^t \in H_\epsilon : h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{h}^t)} \mathbb{P}^t(\mathbf{h}^t) \\
&= w^t(h^t_\star(\mathbf{h})) - K = u^t(\mathbf{h}) - K \,.
\end{aligned}$$

$$\begin{aligned}
\sum_{\mathbf{h} \in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \mathbb{E}\left[\hat{u}^t(\mathbf{h})^2\right] &= \mathbb{E}\left[\sum_{\mathbf{h} \in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \hat{u}^t(\mathbf{h})^2\right] \\
&= \sum_{\mathbf{h} \in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \sum_{\mathbf{h}^t \in H_\epsilon} \mathbb{P}^t(\mathbf{h}^t) \left(\frac{u^t(h^t_\star(\mathbf{h})) - K}{\mathbb{P}_{\mathbf{l}^t \sim \mathcal{B}^t}(h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{l}^t))}\right)^2 \mathbf{1}(h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{h}^t)) \\
&= \sum_{\mathbf{h} \in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \left(\frac{u^t(h^t_\star(\mathbf{h})) - K}{\mathbb{P}_{\mathbf{l}^t \sim \mathcal{B}^t}(h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{l}^t))}\right)^2 \sum_{\mathbf{h}^t \in H_\epsilon : h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{h}^t)} \mathbb{P}^t(\mathbf{h}^t) \\
&= \sum_{\mathbf{h} \in H_\epsilon} \mathbb{P}^t(\mathbf{h}) \frac{\left(u^t(h^t_\star(\mathbf{h})) - K\right)^2}{\mathbb{P}_{\mathbf{l}^t \sim \mathcal{B}^t}(h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{l}^t))} \\
&\leq K^2 \sum_{\mathbf{h} \in H_\epsilon} \frac{\mathbb{P}^t(\mathbf{h})}{\mathbb{P}_{\mathbf{l}^t \sim \mathcal{B}^t}(h^t_\star(\mathbf{h}) \in A^t_\star(\mathbf{l}^t))} \\
&\leq K^2 \sum_{h^t_\star \in \mathcal{O}^t} \frac{\mathbb{P}^t(h^t_\star)}{\mathbb{P}_{\mathbf{l}^t \sim \mathcal{B}^t}(h^t_\star \in A^t_\star(\mathbf{l}^t))} \\
&\leq K^2 \sum_{h^t_\star \in \mathcal{O}^t} \frac{\mathbb{P}^t(h^t_\star)}{\sum_{a \in \mathcal{O}^t : h^t_\star \in A^t_\star(a)} \mathbb{P}^t(a)} \qquad (60) \\
&\leq K^2 8K \log\left(2 \frac{1}{\epsilon^{2K} \alpha K}\right) \\
&\leq 8K^4 \log\left(\frac{2}{\epsilon}\right) \,.
\end{aligned}$$

Taking $\alpha = \frac{1}{K}$.

Where to bound (60), we use lemma 11 from Alon et al., 2017, restated in the Appendix.

We define a graph over the elements of $\mathcal{O}^t$, such that each element $o_1$ has an incoming edge from the other elements $o_2$ such that $o_1 \in A^t_\star(o_2)$. This graph matches (60) to the expression lemma 11 allows to bound.

It only remains to determine the independence number of this graph. First notice that, for each value of $k + \frac{1}{2}$ only one bid-gaps with this first index can belong to $\mathcal{O}^t$. Indeed, otherwise, since there exists a bid-profile $\mathbf{h}$ such that both belong to it, $h^t_\star(\mathbf{h})$ would have two values, which is impossible because only one bid or bid-gap per bid profile can have non-zero sub-utility.

Then notice that for two bids in $\mathcal{O}^t$, with the same first index $k$ an integer values, the observed set of the *lowest* one necessarily contains the other. This naturally arises from the definition of $A^t$.

These two observations ensure that, in an independent set of this graph, there is at most one element having each index $k \in \{1, \frac{3}{2}, 2, ...\frac{2K-1}{2K}, K\}$. This ensures the independence number of this is at most $2K$. Which using the lemma 11 from Alon et al., 2017, completes the proof.

□

We restate the regret guarantees in the all-winner feedback before the proof.

**Theorem 2.** *For any time horizon $T$, using Algorithm 1 in the repeated $K$-unit auction with uniform pricing guarantees, under all-winner feedback, a regret of at most $\mathcal{O}\left(K^{5/2}\sqrt{T}\log(T)\right)$ with $\eta = K^{-1}T^{-1/2}$ and $\epsilon = K^{3/2}T^{1/2}$.*

*Proof of Theorem 2.* The proof of this theorem is identical to the one of theorem 1, with only the need to replace $\hat{u}^t$ by $\bar{u}^t$ up to the point where we bound the regret in equation 51. The proof completes as follows. The discretized regret can be bounded as in B.2.1:

$$R_{T,disc} \leq \frac{\log|H_\epsilon|}{\eta} + \eta \sum_{t=1}^{T} \sum_{\mathbf{h}\in H_\epsilon} \mathbb{P}^t(\mathbf{h})\mathbb{E}\left[\bar{u}^t(\mathbf{h})^2\right] \tag{61}$$

$$\leq K\frac{\log\left(\frac{1}{\epsilon}\right)}{\eta} + \eta 8K^4 T \log\left(\frac{2}{\epsilon}\right) \tag{62}$$

$$\leq K^{5/2}\sqrt{T}\left(8\log(2) + 9\log\left(\frac{1}{\epsilon}\right)\right), \tag{63}$$

with $\eta = \frac{1}{K\sqrt{T}}$.

Then using Lemma 7, we can bound the regret:

$$R_T = R_{T,disc} + KT\epsilon \tag{64}$$

$$\leq K^{5/2}\sqrt{T}\left(8\log(2) + 9\log\left(\frac{1}{\epsilon}\right)\right) + KT\epsilon \tag{65}$$

$$\leq K^{5/2}\sqrt{T}\left(1 + 8\log(2) + \frac{9}{2}\log\left(\frac{T}{K^3}\right)\right), \tag{66}$$

with $\epsilon = \sqrt{\frac{K^3}{T}}$ □

# C Proof of technical lemmas

**Lemma 4.** *With the all-winner feedback, the bidder can compute from its feedback the sub-utilities of any pseudo bid in $A(\mathbf{h}^t, \boldsymbol{\beta}^t)$, defined as:*

$$A(\mathbf{h}^t, \boldsymbol{\beta}^t) := \left\{h_{k,j}, (k,j) \in \mathcal{K}\times\mathcal{J}_\epsilon \middle| \text{ s.t. } \{k > x^t\} \text{ or } \{k = x^t \text{and } j\epsilon \geq p^t\}\right\} \tag{18}$$

*Where $x^t := x_H(\mathbf{h}^t, \boldsymbol{\beta}^t)$ and $p^t = p_H(\mathbf{h}^t, \boldsymbol{\beta}^t)$.*

*Proof of Lemma 4.* Let $t \in [T], \mathbf{h}^t$ and $\boldsymbol{\beta}^t$ be the action of the player and the adversary at time $t$. Let $\mathbf{b}^t$ be the corresponding bid to the pseudo-bid $\mathbf{h}^t$. Under the all-winner feedback, all winning bids are revealed, hence the feedback reveals to the learner the $K - x(\mathbf{b}^t, \boldsymbol{\beta}^t)$ biggest bids of the adversary : $(\beta_i)_{i\leq K-x(\mathbf{b}^t,\boldsymbol{\beta}^t)}$. Furthermore, since the price is known, the learner can deduce from the rules of the auction that for all $i \geq K - x(\mathbf{b}^t, \boldsymbol{\beta}^t)$ :

$$\beta_i \leq p(\mathbf{b}^t, \boldsymbol{\beta}^t). \tag{67}$$

For any value $k, j \in \mathcal{K}\times\mathcal{J}_\epsilon$ such that $h_{k,j} \in A(\mathbf{h}, \boldsymbol{\beta})$, let's show that we can evaluate the corresponding sub-utilities.

We first look at the ability of the learner to evaluate the indicator functions in the sub-utilities defined in Lemma 3, for $h_{k,j} \in A(\mathbf{h}^t, \boldsymbol{\beta}^t)$.

For the integer values of $k$, we can rewrite the indicator function of the sub-utilities, as follows : $\mathbb{1}\{p_H(\mathbf{h},\boldsymbol{\beta}) = j\epsilon\} \cap \{x_H(\mathbf{h},\boldsymbol{\beta}) = k\} = \mathbb{1}\{\beta_{K-k} > j\epsilon > \beta_{K-k+1}\}$. When $K-k+1 \leq K - x(\mathbf{b}^t, \boldsymbol{\beta}^t)$ this can be evaluated for any value of $j$, because the adversary bids are known. When $K-k = K-x(\mathbf{b}^t,\boldsymbol{\beta}^t)$, the indicator function can still be evaluated if $j\epsilon > p(\mathbf{b}^t,\boldsymbol{\beta}^t)$ using (67).

For the half-integer values of $k$, we can rewrite the indicator function of the sub-utilities Lemma 2 as follows : $\mathbb{1}\{p_H(\mathbf{h},\boldsymbol{\beta}) \in (j\epsilon,(j+1)\epsilon)\} \cap \{x_H(\mathbf{h},\boldsymbol{\beta}) = k-1/2\} = \mathbb{1}\{j\epsilon < \beta_{K-k+1/2} < (j+1)\epsilon\}$. Therefore, when $K-k+1/2 \leq K - x(\mathbf{b}^t,\boldsymbol{\beta}^t)$ this indicator function can be evaluated. Hence when $k \geq x + 1/2$, and that regardless of the value of $j$.

Therefore, it is always possible for the learner to evaluate the indicator function.

Evaluating the remaining term of the sub-utilities $\sum_{l=1}^{\lfloor k \rfloor} v_l - p_H(\mathbf{h},\boldsymbol{\beta})$ is more straightforward since it only needs to be done when the indicator function takes value 1.

For the integer values of $k$, if the indicator function takes value 1, then $p_H(\mathbf{h},\boldsymbol{\beta}) = j\epsilon$, therefore the remaining term is known.

For the half-integer values of $k$, if the transformed indicator function takes value 1, then the price is set by $\beta_{K-k+1/2}$, therefore, the remaining term is also known.

This concludes the proof as the full sub-utilities can always be evaluated on $A(\mathbf{h},\boldsymbol{\beta})$. □

# D Restated results from the literature

## D.1 Exponential weight forecaster

In this problem of learning under expert advice, they are $N$ expert and at each time $t \in [N]$, the learner chooses a probability to play each expert $(y_i^t)_{i\in[N]} \in \mathcal{Y}$ and nature reveals the losses $(l_i^t)_{i\in[N]} \in [0,L]^N$.

$$R_n = \sum_{t=1}^{n}\sum_{i=1}^{N} y_i^t l_i^t - \min_{i\in[N]}\left(\sum_{t=1}^{n} l_i^t\right) .$$

**Theorem 4** (Theorem 2.2 Cesa-Bianchi and Lugosi, 2006)**.** *Assume that the losses $l$ take values in $[0,L]$. For any $n$ and $\eta > 0$, and for all $y_1,\ldots,y_n \in \mathcal{Y}$, the regret of the exponentially weighted average forecaster satisfies*

$$R_n \leq \frac{\log N}{\eta} + \frac{nL^2\eta}{8}.$$

*In particular, with $\eta = \sqrt{8\ln N/n}$, the upper bound becomes $\sqrt{(n/2)\ln N}$.*

This theorem, besides the changes in notations, is a slight variation from the original formulation as it allows for losses greater than 1. The resulting $L^2$ term in the upper bound is a well known extension and the steps to prove this extension to scaled losses are provided in the original work by Cesa-Bianchi and Lugosi, 2006.

## D.2 Lemma graph feedback

The following lemma is restated from Alon et al., 2017.

**Lemma 11.** *Let $G = (V,E)$ be a directed graph with $|V| = K$, in which each node $i \in V$ is assigned a positive weight $w_i$. Assume that $\sum_{i\in V} w_i \leq 1$, and that $w_i \geq \epsilon$ for all $i \in V$ for some constant $0 < \epsilon < \frac{1}{2}$. Then*

$$\sum_{i\in V} \frac{w_i}{w_i + \sum_{j\in N^{\text{in}}(i)} w_j} \leq 4\alpha \ln\frac{4K}{\alpha\epsilon} ,$$

*where $\alpha = \alpha(G)$ is the independence number of $G$.*

### D.3 First price auction lower bound

We restate Theorem 10 from (Balseiro et al., 2019) :

**Theorem 5** (Lower Bound for Learning to Bid)**.** *Any algorithm must incur $\Omega(T^{2/3})$ regret for the learning to bid in first-price auctions problem, even if the value of the bidder is fixed (i.e., there is only one context).*

## NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims made in the abstract are presented in the paper, specifically as the two main Theorem 1 and Theorem 2 and Lemma 5.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discuss and introduces the assumptions made in order for the stated results to hold, most of which are stated in the Introduction 1.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: While the full set of assumption is presented as part of the problem setting in the Introduction 1, proofs of the theorems are provided in the appendix B.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [NA] .

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer:[NA]

Justification: The paper does not include experiments.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://nips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [NA] .

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [NA] .

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)

- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [NA] .

   Justification: The paper does not include experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification:

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: This is a theoretical paper, its result are not tied to a specific field. While it might be the basis for further research into applying learning in auction, which would allow for participant in auction to better adapt to others strategies, it is unclear what societal impact this might have and how fit for practical use the techniques develloped here are.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [NA]

    Justification: The paper does not use existing assets.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.