ONLINE TWO-SIDED MARKETS: MANY BUYERS ENHANCE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

We study repeated trading problems involving a *single seller* and *multiple buyers*. We generalize classical bilateral trade mechanisms by introducing a second-price auction among the buyers before proposing a price to the seller. This setup increases competition and reveals more information about buyers' valuations, resulting in an *informational advantage* over bilateral trade. However, the seller still makes a binary accept/reject decision, resulting in *asymmetric feedback*. We show that this richer feedback model enables us to overcome known lower bounds in the single-buyer case. Specifically, we achieve a $\tilde{O}(T^{2/3})$ regret bound against the best *strong budget-balanced* mechanism, without assumptions on value distributions. Furthermore, when either *(i)* the valuations of the buyers and the seller are independent or *(ii)* they are sampled from bounded density distributions, we achieve a $\tilde{O}(T^{2/3})$ regret bound against *global budget-balanced* mechanisms. Along the way, we also derive new techniques for constrained bandits with feedback graphs.

1 INTRODUCTION

Facilitating trades between sellers and buyers is fundamental in several economic systems. In this work, we study two-sided markets with a *single seller* and *multiple buyers*, where each agent seeks to trade a single item. These settings are typically tackled as mechanism design problems, where agents have private valuations for the item and aim to maximize their own utilities. Ideally, a mechanism should ensure *efficiency*, *i.e.*, maximize the sum of agents' utilities. A standard tool to address this problem is the VCG mechanism (Nisan et al., 2007). However, the VCG mechanism fails to achieve *budget balance*, as it requires the mechanism designer to subsidize the market and incur financial losses. Indeed, a seminal result by Myerson & Satterthwaite (1983) shows that no efficient mechanism can be simultaneously incentive compatible, individually rational, and budget balanced.

A recent line of work initiated by Cesa-Bianchi et al. (2023) tries to circumvent this impossibility result through the lens of online learning. In our two-sided market setting, at each round t, a new seller and n new buyers arrive, with their valuations being drawn from an unknown (possibly joint) distribution. The learner (mechanism designer) first runs a second-price auction among the buyers, after setting a reserve price q_t. If the highest buyer valuation $\overline{b}_t$ is such that $\overline{b}_t \geq q_t$, then the corresponding buyer is willing to trade. Similarly, a price p_t is proposed to the seller, who accepts the trade if their valuation s_t is such that $s_t \leq p_t$. If both parties are willing to trade, then the resulting *gain from trade* is defined as:

$$\mathsf{GFT}_t(p_t, q_t) \coloneqq (\overline{b}_t - s_t)\mathbb{I}(s_t \leq p_t)\mathbb{I}(\overline{b}_t \geq q_t).$$

The learner's objective is to maximize the cumulative gain from trade. Instead of aiming for full efficiency, this approach relaxes the baseline and focuses on the less demanding task of achieving performance comparable to the *best fixed-price mechanism in hindsight*. In doing so, the mechanism must ensure some budget balance constraints during learning. Such constraints guarantee that the mechanism designer is *not* subsidizing the market, and, thus, they are deeply connected with its revenue. When a trade occurs, the buyer with the highest valuation pays the maximum between the reserve price q_t and the second-highest buyer valuation $\underline{b}_t$ (as in a classical second-price auction with reserve price), while the seller receives payment p_t. Thus, the revenue of the mechanism is

$$\mathsf{Rev}_t(p_t, q_t) = (\max\{q_t, \underline{b}_t\} - p_t)\mathbb{I}(s_t \leq p_t)\mathbb{I}(\overline{b}_t \geq q_t).$$

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Related Works Previous works focus on bilateral trade problems Cesa-Bianchi et al. (2023; 2024); Bernasconi et al. (2024a); Azar et al. (2022), which involve a single seller and a *single* buyer. Cesa-Bianchi et al. (2024) provide some positive results under certain assumptions on the underlying distribution of valuations. Specifically, they provide a tight $\tilde{\mathcal{O}}\left(T^{2/3}\right)$ regret bound when seller and buyer's valuations are drawn independently of each other, and i.i.d. from distributions with a bounded density. Moreover, they provide a $\tilde{\mathcal{O}}\left(T^{3/4}\right)$ regret bound when only the second assumption holds. Cesa-Bianchi et al. (2023) show that this regret rate is tight. Azar et al. (2022) study weak budget-balanced mechanisms and show how to attain (a tight) sublinear 2-regret bound, when the valuations are adversarial. Cesa-Bianchi et al. (2023) prove that sublinear regret can be attained against a smoothed adversary. Finally, Bernasconi et al. (2024a) show that sublinear regret is achievable without any assumption on the adversary by using *global budget balance* (GBB) mechanisms, *i.e.*, when the mechanism only ensures that $\sum_t \mathsf{Rev}_t(p_t, q_t) \geq 0$. Despite using GBB mechanisms, Bernasconi et al. (2024a) still compare mechanisms against an optimal *strong budget balance* (SBB) one. In a very recent work, Babaioff et al. (2024) study the sample complexity of learning optimal mechanisms in two-sided markets with a single seller and two buyers, focusing only on full feedback. We refer to Appendix A for a complete detailed discussion of related works.

Original Contributions Although our framework is a generalization of bilateral trade in many aspects, our work is based on the observation that *the presence of multiple buyers provide an informational advantage, which enhances learning*. Indeed, while in bilateral trade the buyer simply reveals whether they want to accept the price or not, the presence of multiple buyers increases competition: buyers have to reveal their valuations in order to win the second-price auction. However, the seller only decides whether to accept or reject the proposed price. Hence, our feedback is still much less powerful than full feedback. In this paper, we show that *such an asymmetric feedback can be exploited to circumvent impossibility results holding in bilateral trade with two-bit feedback*. Our main results are summarized in the following two bullet points.

- Our first result is an assumption-free $\tilde{\mathcal{O}}\left(T^{2/3}\right)$ regret bound when competing against SBB mechanisms. The main technical challenge is to circumvent the $\Omega(T)$ lower bound of Cesa-Bianchi et al. (2024). This is proven by showing that the mechanism must find a price that is a "needle in a haystack". This prevents the use of (non-adaptive) discretization techniques commonly used in continuous scenarios. We show that our asymmetric feedback can be used to design an *adaptive* grid to work on. Interestingly, it is sufficient to use an adaptive grid only on the buyers' price, on which we have a stronger feedback. We remark that finding optimal SBB mechanisms is essentially a single-dimensional problem, since the buyers' price q_t determines the seller price, as it is equal to $\max\{q_t, \underline{b}_t\}$.
- Our second main result is the first sublinear regret bound against GBB mechanisms. Specifically, we obtain $\tilde{\mathcal{O}}\left(T^{2/3}\right)$ regret when either the valuations are independent or their distributions admit a bounded density. The only work that considers a similar baseline (but with randomization) is Bernasconi et al. (2024a). However, its focus is on an adversarial setting, where linear regret is unavoidable. This impossibility result is based on the observation that learning optimal GBB mechanisms incorporates an online learning problem with unknown constraints. We show that this challenge can be overcome in a stochastic environment. As an additional challenge, we have to deal with a "huge set of arms". If each price lies on a grid of size K, we have to optimize over K^2 pairs, and this makes impossible learning without sharing the feedback across arms. Our proof relies on two main components: (i) an adaptive grid that works when either buyers and seller' distributions are independent or the joint distribution has bounded density, and (ii) a tool to share feedback among arms.

To address point (ii), we provide a novel technique for dealing with constrained feedback graphs where constraints are monotonically structured, based on successive arm elimination with precedence. We restrict to graphs that satisfy two properties: (i) they consist of n sets of nodes (represented by indexes), in which each node in a set provides feedback on all the nodes in the same set with higher index, and (ii) constraints are monotone, *i.e.*, if an arm B is unfeasible and there is an edge from A to B, then A is also unfeasible. For this family of problems we provide $\tilde{O}(\sqrt{Tn})$ regret bound (where we omitted a logarithmic dependence from the number of arms) and $\tilde{O}(\sqrt{Tn})$ positive constraint violations. This result is provided in a general setting and it may be of independent interest. Table 1 in the appendix shows a summary of our results, compared with those in the most related works.

The Role of Competition: Revenue and Feedback A primary advantage of increasing the number of buyers from one to many is the ability to extract greater revenue, which in turn provides more "slack" in the GBB constraint. This additional revenue is a key driver for improving overall performance and achieving higher efficiency benchmarks. The role of competition in revenue maximization is a cornerstone of auction design Krishna (2009). Specifically, in regular symmetric environments, a *second-price auction* (SPA) with an optimal reserve price is known to be a revenue-maximizing mechanism. While this setting does not encompass all possible scenarios, this class of mechanisms offers several crucial advantages in our context. First, even in asymmetric settings, the SPA guarantees an efficient allocation among buyers, which is a prerequisite for maximizing the gain from trade. Second, and more critically, SPAs provide a superior feedback structure for online learning. Unlike the minimal information revealed in ascending-bid or take-it-or-leave-it formats, the additional feedback inherent in SPAs—specifically observing the highest bid—serves as the primary lever that enables us to achieve $\tilde{\mathcal{O}}(T^{2/3})$ regret. One might question whether the complexity of learning a reserve price is justified given the power of competition. A seminal result by Bulow & Klemperer (1996) addresses this by showing that, under independent valuations, the expected revenue from an auction with $n+1$ symmetric bidders and no reserve price exceeds the revenue of any optimal mechanism (including those with a finely-tuned reserve) with only n bidders. This suggests that attracting an additional participant may be more impactful than optimizing the mechanism. However, it is important to note that regret is defined relative to the number of available buyers n. As a result, while increased competition affects the achievable revenue, it does not eliminate the fundamental challenges of online learning. Intuitively, adding additional buyers merely shifts the benchmark against which the performance is measured, while leaving the underlying learning problem essentially unchanged. This intuition is reflected in the resulting regret bounds, which notably do *not* depend on the number of buyers. Ultimately, although our framework generalizes bilateral trade in several dimensions, our contribution hinges on the observation that the presence of multiple buyers provides a distinct informational advantage. Consequently, we also provide tight bounds on bilateral trade settings under this enhanced feedback, which is related to the notion of semi-feedback in bilateral trade Chen et al. (2025); Jin (2026).

2 Repeaded Two-Sided Markets With Many Buyers and a Single Seller

We study online learning in a *two-sided* trading problem, where *many buyers* interact repeatedly with a *single seller* over $T \in \mathbb{N}$ rounds. We focus on a *stochastic* setting where the valuations of both the buyers and the seller are jointly drawn from a fixed distribution. At each round $t \in [T]$,[1] the market consists of $n \in \mathbb{N}$ buyers that are potentially interested in buying an item from a single seller. The buyers are characterized by some private valuations $\mathbf{b}_t = (b_{t,1}, \dots, b_{t,n}) \in [0,1]^n$ for the item, while the seller has their own private valuation $s_t \in [0,1]$. The buyers' and seller's valuations are sampled from a joint probability distribution $\mathcal{P}$ supported on $[0,1]^{n+1}$. Specifically, at each round $t \in [T]$, the pair $(\mathbf{b}_t, s_t) \sim \mathcal{P}$ is drawn independently from $\mathcal{P}$.

Our goal is to design mechanisms that maximize the gain from trade, so as to *optimize the efficiency of the market*. As customary in the related literature on bilateral trade, we focus on *fixed-price* mechanisms, by considering a straightforward generalization of the mechanism commonly adopted in bilateral trade (see, *e.g.*, Cesa-Bianchi et al. (2023; 2024); Bernasconi et al. (2024a)). At each round $t \in [T]$, the mechanism first implements a second-price auction among the buyers, by setting a reserve price $q_t \in [0,1]$. Let $\overline{b}_t \coloneqq \max_{b_{t,i} \in \mathbf{b}_t} b_{t,i}$ and $\underline{b}_t \coloneqq \max_{b_{t,i} \in \mathbf{b}_t : b_{t,i} \neq \overline{b}_t} b_{t,i}$ be the highest and the second-highest buyer valuations at round t, respectively, assuming ties are broken arbitrarily. If the highest buyer valuation exceeds the reserve price, *i.e.*, $\overline{b}_t \geq q_t$, the highest-valuation buyer wins the possibility of buying the item at a price equal to $\max\{q_t, \underline{b}_t\}$. Otherwise, no buyer wins the item. Then, the mechanism proposes a selling price $p_t \in [0,1]$ to the seller, who accepts to sell the item only if $s_t \leq p_t$. Finally, if the buyer with the highest valuation wins the item and the seller accepts to sell it, then the trade is completed and the resulting *gain from trade* is $\mathsf{GFT}_t(p_t, q_t) \coloneqq (\overline{b}_t - s_t)\mathbb{I}(s_t \leq p_t)\mathbb{I}(\overline{b}_t \geq q_t)$, where $\mathbb{I}(\cdot)$ denotes the indicator function. In the following, we also use $\mathsf{GFT}(p,q) \coloneqq \mathbb{E}_{(\mathbf{b},s)\sim\mathcal{P}}\left[(\overline{b} - s)\mathbb{I}(s \leq p)\mathbb{I}(\overline{b} \geq q)\right]$ to denote the *expected* gain from trade of a mechanism defined by a pair of fixed prices $(p,q) \in [0,1]^2$, where we let

[1] In this paper, we use $[x]$ to denote the set $\{1, \dots, x\}$ of the first $x \in \mathbb{N}$ natural numbers.

$\overline{b} \coloneqq \max_{b_i \in \mathbf{b}} b_i$ be the highest buyer valuation in the tuple $\mathbf{b} = (b_1, \dots, b_n) \in [0,1]^n$ sampled from $\mathcal{P}$. Beside being simple, it is easy to see that our mechanism enjoys the desirable property of being *dominant-strategy incentive compatible* (DSIC) for both the buyers and the seller.

While the primary objective in trading problems is to maximize the gain from trade, mechanisms typically need to satisfy an additional *budget balance* property, which could be either *weak/strong* (see, *e.g.*, Cesa-Bianchi et al. (2023; 2024)) or *global* (see, *e.g.*, Bernasconi et al. (2024a)). We consider two distinct classes of mechanisms based on the specific budget balance property being examined.

Per-Round Budget-Balanced Mechanisms: WBB and SBB Budget balance properties are concerned with the *revenue* of the mechanism $(p_t, q_t) \in [0,1]^2$ implemented at round $t \in [T]$, defined as: $\mathsf{Rev}_t(p_t, q_t) \coloneqq (\max\{q_t, \underline{b}_t\} - p_t)\mathbb{I}(s_t \leq p_t)\mathbb{I}(\overline{b}_t \geq q_t)$, where $\max\{q_t, \underline{b}_t\} - p_t$ represents the revenue earned by the mechanism in the event that the trade occurs. As for the gain from trade, in the following we denote by $\mathsf{Rev}(p,q) \coloneqq \mathbb{E}_{(s,\mathbf{b})\sim\mathcal{P}}\left[(\max\{q, \underline{b}\} - p)\mathbb{I}(s \leq p)\mathbb{I}(\overline{b} \geq q)\right]$ the *expected* revenue of a pair of prices $(p,q) \in [0,1]^2$, where we let $\underline{b} \coloneqq \max_{b_i \in \mathbf{b}: b_i \neq \overline{b}} b_i$. Budget balance generally guarantees that the mechanism is *not* subsidizing the market. In particular, *weak budget balance* (WBB) ensures that at each round the mechanism is *not* losing money. Formally:

$$\mathsf{Rev}_t(p_t, q_t) \geq 0 \quad \forall t \in [T].$$

As it will become clear in the following, WBB mechanisns are useful in learning scenarios. However, their gain from trade is dominated by the one of SBB mechanisms. Indeed, the SBB property guarantees that the mechanism is *not* gaining any profit from the market, and, thus, SBB mechanisms are the proper tool to maximize the gain from trade. Formally, SBB requires that $\mathsf{Rev}_t(p_t, q_t) = 0 \quad \forall t \in [T]$. The SBB property removes a degree of freedom from the mechanism, by forcing $p_t = \max\{q_t, \underline{b}_t\}$. Hence, we can identify a SBB mechanism by only specifying its reserve price $q_t \in [0,1]$. Since SBB mechanisms are both simpler and as powerful as WBB ones, similar to previous work on bilateral trade (see, *e.g.*, Cesa-Bianchi et al. (2023; 2024)), one of our goals will be to achieve performance close to that of an optimal SBB mechanism through the use of WBB mechanisms. Namely, the regret is defined as: $R_T \coloneqq T \cdot \max_{q\in[0,1]} \mathsf{GFT}(q) - \mathbb{E}\left[\sum_{t\in[T]} \mathsf{GFT}(p_t, q_t)\right]$, where q_t is the reserve price decided by the learning algorithm at each round t, $\mathsf{GFT}(q)$ and $\mathsf{GFT}_t(q_t)$ are used to denote $\mathsf{GFT}(\max\{\underline{b}_t, q\}, q)$ and $\mathsf{GFT}_t(\max\{\underline{b}_t, q_t\}, q_t)$, respectively, and the expectation is with respect to the randomness of the learner.

Global Budget Balance (GBB) This is a looser requirement that can be applied to mechanisms. Specifically, it does *not* force the revenue to be null at each round, but it only requires that the mechanism is *not* subsiding the market (*i.e.*, losing money) overall. Formally:

$$\sum_{t\in[T]} \mathsf{Rev}_t(p_t, q_t) \geq 0.$$

Thus, under the GBB property, the mechanism can choose the prices p_t with the only requirement that the budget balance constraint holds globally over the T rounds. We define our baseline as the fixed prices (p,q) that maximizes the gain from trade and guarantees positive revenue. Formally, $\mathsf{Opt} \coloneqq \max_{p,q\in[0,1]} \mathsf{GFT}(p,q)$ subject to the constraint $\mathsf{Rev}(p,q) \geq 0$. We define the regret accordingly as $R_T = T \cdot \mathsf{Opt} - \mathbb{E}\left[\sum_{t\in[T]} \mathsf{GFT}_t(p_t, q_t)\right]$, where (p_t, q_t) are the prices decided by the learning algorithm at each round t, and the expectation is with respect to the randomness of the learner.

Feedback After posting the reserve price q_t, the learner observes the bids of all the buyers with valuation larger than q_t. Moreover, the learner observes if $s_t \leq p_t$. This feedback is equivalent to the two-bit feedback (that is observing each agent decision separately) used in bilateral trade for the seller, but it is stronger for the buyer. Such asymmetry is motivated by the second-price auction with reserved price to which buyers participate. We refer to such feedback as *asymmetric* feedback.

Environments Most of our results hold for an arbitrary distribution $\mathcal{P}$. In some settings, problems with general distributions are non-learnable and we have to restrict to more structured distributions:

- *Independent:* The distribution of the buyers' valuations $\mathcal{P}^B$ and the one of the seller's valuations $\mathcal{P}^S$ and are independent, and $\mathcal{P} := \mathcal{P}^B \times \mathcal{P}^S$.
- *Bounded density:* $\mathcal{P}$ admits a density bounded by M.

Relation to bilateral trade Our setting generalizes bilateral trade to the case of multiple buyers. Indeed, by setting the number of buyers to n=1, we recover the online bilateral trade problem with semi-feedback investigated by Chen et al. (2025); Jin (2026). However, while the semi-feedback model assumes that the learner directly observes the buyers' private valuations, our feedback model is more restrictive: we observe valuations only if they exceed the chosen reserve price.

3 GFT Decomposition

Our first result breaks down the $\mathsf{GFT}(\cdot,\cdot)$ function into simpler components that allow to effectively exploit our feedback structure. A similar result for bilateral trade is presented in (Cesa-Bianchi et al., 2024). We decompose $\mathsf{GFT}(p,q)$ into two parts (see Appendix B for details):

$$\mathsf{GFT}(p,q) = \mathsf{GFT}_1(p,q) + \mathsf{GFT}_2(p,q),$$

where $\mathsf{GFT}_1(p,q) := \mathbb{E}_{(\mathbf{b},s)\sim\mathcal{P}}[(p-s)\mathbb{I}(\bar{b}\geq q)\mathbb{I}(s\leq p)]$ and $\mathsf{GFT}_2(p,q) := \mathbb{E}_{(\mathbf{b},s)\sim\mathcal{P}}[(\bar{b}-p)\mathbb{I}(\bar{b}\geq q)\mathbb{I}(s\leq p)]$. While GFT_1 is difficult to estimate due to limited feedback on s_t, we can recover full feedback on it by posting $(U,0)$ with $U\sim\mathcal{U}([0,1])$, allowing complete estimation of GFT_1 at the cost of using some rounds for exploration. In contrast, GFT_2 is easier to estimate via $\bar{b}_t$, which is either observed or known to be below q. Moreover, GFT_2 feedback is stronger than bandit: playing (p,q) informs us about all $\mathsf{GFT}_2(p,q')$ for $q'\geq q$. However, feedback remains bandit-like with respect to the seller price p.

A similar decomposition holds when fixing $p=\max\{q,\underline{b}\}$, useful in strong budget balance settings. This yields a single-price form $\mathsf{GFT}(q)$, split into two analogous components $\mathsf{GFT}_1(q)$ and $\mathsf{GFT}_2(q)$.

4 From Continuous to Discrete Prices

In this section, we deal with the challenge of dealing with a continuous action set of possible prices $[0,1]^2$. As it is standard in such settings, we would like to work with a discretization of the action set including finitely many actions. However, the function $\mathsf{GFT}(\cdot,\cdot)$ is not continuous in the prices, making it impossible to design a finite grid $\mathcal{B}$ that approximates the price space without knowledge of the distribution $\mathcal{P}$—as the locations of the discontinuities are unknown. Indeed, Cesa-Bianchi et al. (2024) show that such grids provide a good approximation only under the specific assumption of bounded density, even in the simpler bilateral trade setting. We address this by building an adaptive grid using samples from $\mathcal{P}$, refining it in high-density regions where the function changes most, and coarsening it where density is low. The goal is to ensure that intervals between grid points have low probability mass, echoing the bounded density assumption in prior work Cesa-Bianchi et al. (2023). Our method leverages asymmetric feedback: committing to $q_t=0$ gives full information about buyer valuations, allowing us to apply adaptivity only to the buyer's price grid. This would not work under standard two-bit feedback.

Algorithm 1 `GridEstimation`

Input: type $\in\{\mathsf{SBB},\mathsf{GBB}\}, T_0, K$
Initialize $\mathcal{B}\leftarrow\{0\}$,$\mathcal{B}^B\leftarrow\{0\}$, $k\leftarrow 0$, $p_0\leftarrow 0$
▷ $\mathcal{B}$ is a multiset, *i.e.*, it contains duplicates
for $t\in[T_0]$ **do**
  Play $q_t=0$ and observe $\bar{b}_t,\underline{b}_t$
  Play $p_t=\underline{b}_t$
  $\mathcal{B}\leftarrow\mathcal{B}\cup\{\bar{b}_t\}$
end for
while $k<K$ **do**
  $p_{k+1}\leftarrow$ max. $q\in\mathcal{B}$ s.t. $|\{b\in\mathcal{B}: p_k<b<q\}|\leq\frac{T_0}{K}$
  ▷ $p_{k+1}\leftarrow 1$ if such q doesn't exist
  $\mathcal{B}^B\leftarrow\mathcal{B}^B\cup\{p_{k+1}\}$
  $k\leftarrow k+1$
end while
if type $=\mathsf{SBB}$ **then**
  $\mathcal{B}\leftarrow\mathcal{B}^B\cup\left\{q\in[0,1]:\ q=\frac{n}{K}, n\in\{0\}\cup[K]\right\}$
else if type $=\mathsf{GBB}$ **then**
  $\mathcal{B}^S\leftarrow\left\{p\in[0,1]:\ p=\frac{n}{K}, n\in\{0\}\cup[K]\right\}$
  $\mathcal{B}^B\leftarrow\mathcal{B}^B\cup\left\{q\in[0,1]:\ q=\frac{n}{K}, n\in\{0\}\cup[K]\right\}$
  $\mathcal{B}\leftarrow\mathcal{B}^S\times\mathcal{B}^B$
end if
Return $\mathcal{B}$

Technically, we rely on the DKW inequality (Dvoretzky et al., 1956) to bound, with high probability, the error between true and empirical probabilities. Using T_0 samples from valuations (via $q_t = 0$ and $p_t = \underline{b}_t$), we build a grid such that each interval contains at most T_0/K points. Algorithm 1 details the procedure. In the strong budget balance case, we discretize only buyer prices using a union of the adaptive and uniform grids. Under global budget balance, we apply the same to buyers and use a uniform grid for sellers.

Our first result shows that the gain from trade of the optimal price on the grid is close to the optimal gain from trade of prices $q \in [0, 1]$. Formally:

Lemma 1. *Fix T_0, K and let $\mathcal{B} =$* `GridEstimation`(SBB, T_0, K). *For any $\delta \in (0, 1)$, with probability at least $1 - \delta$, it holds $\max_{q\in[0,1]} \mathsf{GFT}(q) - \max_{q\in\mathcal{B}} \mathsf{GFT}(q) \leq \frac{1}{K} + 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$.*

We can prove a similar result for the global budget balance setting and couples of prices (p, q). Let $(p^*, q^*) \in [0, 1]$ be an expected gain from trade maximizing couple of prices which satisfies GBB. Moreover, we let $(p_{k^*}, q_{j^*}) \in \mathcal{B}$ be the couple of prices where p_{k^*} is the smallest element in $\mathcal{B}^S$ such that $p^* \leq p_{k^*}$ and q_{j^*} is the smallest element in $\mathcal{B}^B$ such that $q^* \leq q_{j^*}$. Then, we obtain the following result.

Lemma 2. *Fix T_0, K and let $\mathcal{B} =$* `GridEstimation`(GBB, T_0, K). *If $\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$, then for any $\delta \in (0, 1)$, with probability at least $1-\delta$, it holds $\mathsf{GFT}(p^*, q^*) - \mathsf{GFT}(p_{k^*}, q_{j^*}) \leq \frac{2}{K} + \sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$.*

Moreover, if $\mathcal{P}$ admits density bounded by M, then for any $\delta \in (0, 1)$, with probability at least $1 - \delta$, it holds $\mathsf{GFT}(p^*, q^*) - \mathsf{GFT}(p_{k^*}, q_{j^*}) \leq \frac{M+1}{K} + 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$.

Finally, we show that (p_{k^*}, q_{j^*}) is almost budget balance, *i.e.*, $\zeta := -\mathsf{Rev}(p_{k^*}, q_{j^*})$ is not too large.

Lemma 3. *Fix T_0, K and let $\mathcal{B} =$* `GridEstimation`(GBB, T_0, K). *If $\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$, then for any $\delta \in (0, 1)$, with probability at least $1 - \delta$, it holds $\zeta \leq \frac{2}{K} + 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$. Moreover, if $\mathcal{P}$ has density bounded by M, then for any $\delta \in (0, 1)$, with probability at least $1 - \delta$, it holds that $\zeta \leq \frac{2M}{K}$.*

Notice that all our results in the global budget balance setting hold only when either seller and buyers' valuations are independent, namely $\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$, or the joint distribution $\mathcal{P}$ admits bounded density. This is essentially unavoidable, as we show in the lower bound presented in Appendix H. Finally, similarly to Bernasconi et al. (2024a) we are able to provide an algorithm performing Revenue maximization that guarantees the following result. We refer the reader to Appendix C for a more detailed discussion.

Lemma 4. *Fix $T_0 = \tilde{\mathcal{O}}\left(T^{2/3}\right), K = \tilde{\mathcal{O}}\left(T^{1/3}\right)$. There exists a learning algorithm* `maxREV` *that, for any $\delta \in (0, 1)$, with probability at least $1 - 2\delta$, guarantees:*

$$\tau \cdot \mathsf{GFT}(p_{k^*}, q_{j^*}) \leq 6 \log(T) \sum_{t\in[\tau]} \mathsf{Rev}_t(p_t, q_t) + \tilde{\mathcal{O}}\left(T^{2/3}\right) + \tau\zeta \quad \forall \tau \in [T].$$

5 Achieving No-Regret With Respect to SBB Mechanisms

In this section we design an algorithm that, using weakly budget balance mechanisms, achieves regret of order $\tilde{\mathcal{O}}\left(T^{\frac{2}{3}}\right)$ with respect to the optimal strongly (or weakly) budget balanced mechanism.

Such result is optimal in two aspect. First, while from a gain from trade perspective weakly budget balance mechanisms are dominated by strong budget balanced ones, they are necessary in learning settings. Indeed, Cesa-Bianchi et al. (2024) shows that it is impossible to learn using the limited feedback of SBB mechanisms even in bilateral trade setting. Moreover, our regret rate is optimal even in the simpler bilateral trade setting Cesa-Bianchi et al. (2024). Such lowerbound easily extends to our setting with multiple buyers. As we show in the following remark, it is sufficient to observe that in both our model and bilateral trade the feedback on the seller valuation is the same and that in the instances designed by Cesa-Bianchi et al. (2024) the challenging task is to estimate the seller valuation distribution.

Remark 1. *The $\Omega(T^{2/3})$ regret lower bound for* SBB *mechanisms in bilateral trade Cesa-Bianchi et al. (2024) extends to our setting. Their hard instances rely on the limited feedback from the seller, which is identical in our case—only the accept/reject response is observed. Moreover, the price rule $\max\{q, \underline{b}\}$ aligns with their setting when only one buyer has non-zero valuation.*

We now describe our algorithm, which can be conceptually divided into three phases:

Grid Construction Run Algorithm 1 for T_0 rounds to get a price grid $\mathcal{B}$ of size K. Thank to the results in section 4, we know that this grid contains a $\tilde{\mathcal{O}}\left(1/K + 1/\sqrt{T_0}\right)$-approximation of the optimal fixed price.

Estimating GFT_1 For T_0 rounds, sample $U_t \sim \mathcal{U}([0,1])$ and play $(p,q) = (U_t, U_t)$. This phase, using WBB mechanisms, allows unbiased estimation of $\mathsf{GFT}_1(q_k)$, to which we incorporate some optimism building $\overline{\mathsf{GFT}}_1(q_k)$.

Bandit Learning of GFT_2 For $t > 2T_0$, perform a UCB-like algorithm selecting q_t that maximize $\overline{\mathsf{GFT}}_1 + \overline{\mathsf{GFT}}_{2,t}$, while updating the estimate of GFT_2 using its bandit feedback nature.

Algorithm 2 Regret Minimizer for SBB Mechanisms

Input: T_0, K, δ
$\mathcal{B} \leftarrow$ `GridEstimation`(SBB, T_0, K)
$\widehat{\mathsf{GFT}}_1(q_k) \leftarrow 0 \ \forall p_k \in \mathcal{B}$
for $t = T_0 + 1, \dots, 2T_0$ **do**
  $U_t \sim \mathcal{U}([0,1])$ ▷ sample U_t from uniform distribution over $[0,1]$
  Play $(p = U_t, q = U_t)$
  for $q_k \in \mathcal{B}$ **do**
    $\widehat{\mathsf{GFT}}_1(q_k) \leftarrow \widehat{\mathsf{GFT}}_1(q_k) + \frac{1}{T_0}\mathbb{I}(s_t \le U_t \le \max\{q_k, \underline{b}_t\})\mathbb{I}(\bar{b}_t \ge q_k)$
  end for
end for
$\overline{\mathsf{GFT}}_1(q_k) \leftarrow \widehat{\mathsf{GFT}}_1(q_k) + \sqrt{\frac{\ln(\frac{2K}{\delta})}{2T_0}} \quad \forall q_k \in \mathcal{B}$
$\widehat{\mathsf{GFT}}_{2,2T_0}(q_k) \leftarrow 0, \overline{\mathsf{GFT}}_{2,2T_0}(q_k) \leftarrow 1$
for $t = 2T_0 + 1, \dots, T$ **do**
  play q_t
  play $p_t = \max\{q_t, \underline{b}_t\}$ ▷ price is posted only if $\bar{b}_t \ge q_t$
  $\widehat{\mathsf{GFT}}_{2,t}(q_t) \leftarrow \frac{1}{n_t(q_t)} \sum_{i=2T_0+1}^{t} (\bar{b}_i - p_i)\mathbb{I}(\bar{b}_i \ge q_i, s_i \le p_i)\mathbb{I}(q_i = q_t)$
  $\overline{\mathsf{GFT}}_{2,t}(q_t) \leftarrow \widehat{GFT}_{2,t}(q_t) + \sqrt{\frac{\ln(\frac{2TK}{\delta})}{n_t(q_t)}}$
  $q_{t+1} \leftarrow \arg\max_{q_k \in \mathcal{B}} \overline{\mathsf{GFT}}_1(q_k) + \overline{\mathsf{GFT}}_{2,t}(q_k)$
end for

The full procedure is detailed in Algorithm 2.

Theorem 1. *For any $\delta > 0$, Algorithm 2 with $T_0 = \tilde{\mathcal{O}}\left(T^{2/3}\right)$ and $K = \tilde{\mathcal{O}}\left(T^{1/3}\right)$ achieves: $R_T = \tilde{\mathcal{O}}\left(T^{2/3}\right)$, with probability at least $1 - 3\delta$.*

6 Achieving No-Regret With Respect to GBB Mechanisms

In this section, we focus on the more challenging global budget balance setting. Our results hold when the buyers' and seller's distribution are independent or with bounded density, as required by Lemma 2 and Lemma 3. As we show in Appendix H, one of these properties is essential for a broad class of approaches. This section is organized as follows. In Section 6.1, we provide an algorithm for constrained MABs with feedback graph. While it uses some specific property of our problem, we believe that it could be of general interest and can be extended to other settings. In Section 6.2, we exploit such regret minimizer to design our GBB regret minimizer algorithm.

6.1 Constrained MABs With Feedback Graphs

We focus on a generalization of MABs with feedback graph to constrained settings. At each round t, the learner selects an arm a_t from a finite set $\mathcal{A}$. Then, random rewards $r_t(a) \in [-1,1]$ and costs $c_t(a) \in [-1,1]$ are sampled from a distribution $\mathcal{D}$. We let $r(a)$ and $c(a)$ be the expected reward and cost of arm a. Different from standard MABs, the learner observe the feedback of some arms Obs_t which depend on the pulled arm a_t. We will introduce later the specific feedback we will focus on.

Let a^* be an optimal solution of $\max_{a \in \mathcal{A}} r(a)$ s.t. $c(a) \le 0$. We aim at designing algorithms that guarantees sublinear regret and violation of the constraints. Differently from standard MABs, we aim at obtaining a better dependence than $\sqrt{|A|}$ with respect to the number of arms, exploiting the enhanced feedback.

Some general approaches for addressing our problem can be derived from the literature on online learning with long-term constraints. In particular, some works derive general primal-dual methods that uses a black-box regret minimizer for the *uncontrained* problem Castiglioni et al. (2022); Slivkins et al. (2023); Bernasconi et al. (2024b). In our specific setting, we could use any regret minimizer for MABs with feedback graphs (see, *e.g.*, (Lykouris et al., 2020)). Unfortunately, such approach provides optimal guarantees only assuming Slater condition. Such condition is not verified in our setting, since there does not exist a couple of prices that guarantees (constant) strictly-positive revenue. In the following, we design a regret minimizer for a more restricted setting but which does not require Slater condition. Moreover, differently from the mentioned primal-dual method, our algorithm bounds the *positive* violations, which can be of independent interest.

Multi-line Feedback Graphs with monotone Costs We consider a particular feedback graph that generalizes the one of our problem. The set of arms $\mathcal{A}$ includes an arm $a_{i,j}$ with $i \in [n]$ and $j \in [m]$, where $n, m \in \mathbb{N}$. When we pull arm $a_{i,j}$, we observe the reward and costs of all arms $a_{i,k}$ for all $k \geq j$. We call a set of arms $\{a_{i,j}\}_{j \in [m]}$ a *line*. Notice that such graph has independent number n (see (Alon et al., 2015) for a definition). Hence, in standard unconstrained problem it is possible to obtain regret $\sqrt{Tn}$ with standard techniques. Of course, such results do not extend to constrained settings.

In the following, we show how to extend the result to constrained setting when the following *monotonicity* assumption on the cost holds: $c(a_{i,j}) \geq 0 \implies c(a_{i,k}) \geq 0 \quad \forall i,\ j < k$. As we will see in the following, this assumption is satisfied by our setting.

Algorithm 3 Successive Arm Elimination with Precedences (SAE-P)

Input: Set of arm $\mathcal{A}$, probability δ
$\mathcal{S}_0(i) \leftarrow \{a_{i',j} \in \mathcal{A} : i' = i\}$ for each $i \in [n]$
$\mathcal{S}_0 \leftarrow \mathcal{A}$
$\bar{r}_0(a), \underline{r}_0(a), \underline{c}_0(a) \leftarrow 0$ for each $a \in \mathcal{A}$
for $t \in [T]$ **do**
  $a_{i_t,\bar{j}_t} \leftarrow \arg\max_{a \in \mathcal{S}_{t-1}} \bar{r}_{t-1}(a)$
  $j_t \leftarrow$ minimum j s.t. $a_{i_t,j} \in \mathcal{S}_{t-1} \wedge \big(\bar{r}_{t-1}(a_{i_t,j}) \geq \underline{r}_{t-1}(a)\ \forall a \in \mathcal{S}_{t-1}(i_t)\big)$
  Play a_{i_t,j_t} and observe feedback $(r_t(a_{i_t,j})$ for each $j \geq j_t$
  Compute $\bar{r}_t$ and $\bar{c}_t$ for all elements in $\mathcal{S}_{t-1}$ according to Equation (1)
  $\mathcal{S}_t(i_t) \leftarrow \{a_{i_t,j} : j \geq j_t \wedge \underline{c}_t(a_{i_t,j}) \leq 0$
  $\mathcal{S}_t(i) \leftarrow S_{t-1}(i)$ for all $i \neq i_t$
  $\mathcal{S}_t \leftarrow \cup_{i \in [n]} S_t(i)$
end for

Algorithm 3 provides a pseudo-code of our algorithm. It essentially implements a successive arm elimination algorithm with *precedences*. It keeps track of the arms that are feasible and not dominated into a set $\mathcal{S}_t$. Then, it chooses the line with the better upper bound on the reward and it plays the arm in such line (and in $\mathcal{S}_t$) which provides stronger feedback. In particular, at each round it chooses the arm $a_{i_t,\bar{j}_t} \in \mathcal{S}_t$ that maximizes the upper bound on the reward. Then, chooses the arm $a_{i_t,j_t} \in \mathcal{S}_t$ which is not provably suboptimal. This is done taking track of a lower and upper confidence bound on each arm as follows:

$$\bar{r}_t(a) = \frac{1}{n_t(a)} \sum_{i=1}^{t} r_i(a) \mathbb{I}(a \in \text{Obs}_i) + 2\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_t(a)}} \quad \forall a \in \mathcal{A} \tag{1a}$$

$$\underline{r}_t(a) = \frac{1}{n_t(a)} \sum_{i=1}^{t} r_i(a) \mathbb{I}(a \in \text{Obs}_i) - 2\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_t(a)}} \quad \forall a \in \mathcal{A} \tag{1b}$$

$$\underline{c}_t(a) = \frac{1}{n_t(a)} \sum_{i=1}^{t} c_i(a) \mathbb{I}(a \in \text{Obs}_i) - 2\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_t(a)}} \quad \forall a \in \mathcal{A} \tag{1c}$$

where Obs_t is the set of arms whose feedback is observed at round t and $n_t(a) = \sum_{i=1}^{t} \mathbb{I}(a \in \text{Obs}_i)$. At the end of each round, the algorithm removes from $\mathcal{S}_t$ all arms that are dominated or unfeasible.

The main property of our algorithm is that we always observe the reward of all the non-removed arms in the played line. This can be done safely thanks to our monotonicity assumption. Indeed, monotone costs play a crucial role in avoiding the removal of an optimal arm. Intuitively, if we remove an arm it means that there is an arm with larger index with higher utility. If it turns out that such arm is unfeasible, by the monotonicity of the costs it must be the case that also the first arm was unfeasible. Finally, we highlight that our algorithm provides a stronger guarantees on the constraints than the

one actually required. In particular, it guarantees that $\sum_{t=1}^{\tau}[c(a_t)]^+$ is sublinear, where we define $[x]^+ = \max\{x, 0\}$. Formally:

Theorem 2. *Algorithm 3 is such that with probability at least* $1 - 2\delta$*:*

$$\tau \cdot r(a^*) - \sum_{t=1}^{\tau} r(a_t) \leq \tilde{\mathcal{O}}\left(\sqrt{nT}\log(m/\delta)\right) \forall \tau \in [T]$$

and $\sum_{t=1}^{\tau}[c(a_t)]^+ \leq \tilde{\mathcal{O}}\left(\sqrt{nT}\log(m/\delta)\right) \forall \tau \in [T]$.

6.2 A GBB Regret-Minimizing Algorithm

In this section, we present a no-regret algorithm for learning optimal GBB mechanisms. As previously noted, identifying which pairs of prices satisfy the budget balance constraint is not straightforward. While we partially address this challenge using Constrained Multi-Armed Bandits (MABs), this approach still incurs a sublinear constraint violation. To compensate for the potential negative revenue in the initial rounds, we begin by accumulating revenue in the first rounds. Specifically, Algorithm 4 starts with an initial phase dedicated solely to maximizing revenue. This allows the algorithm to reach a predefined budget threshold β, an input parameter, before transitioning to the exploration phase. As shown in in appendix C, maximizing revenue can be formulated as a stochastic MAB, and setting $T_0 = T^{2/3}$ and $K = T^{1/3}$, it exists an algorithm `RevMax` that accumulates budget β, while suffering regret at most $\tilde{\mathcal{O}}\left(\beta + T^{\frac{2}{3}}\right)$ thanks to Lemma 7. In the second phase, we estimate the first component of the revenue, GFT_1, similarly to the SBB case. In the final phase, we apply Algorithm 3. We give as input to `SAE-P` the cost $c_t(p, q)$ and reward $r_t(p, q)$ for each arm for where feedback is available. To ensure the feasibility of p_{k^*}, q_{j^*} within the Constrained MAB framework, we shift the revenue, *i.e.*, minus the cost, by an upper bound $\bar{\zeta}$ which depends on the underlying distributional assumption (see Lemma 3)). Finally, we set the reward r_t to the sum of the upper bound on the already estimated component $\overline{\mathsf{GFT}}_1$ and $\mathsf{GFT}_{2,t}$ at the current round.

Algorithm 4 Regret Minimizer for GBB Mechanisms

Input: $K, T_0, \delta, \beta, \bar{\zeta}$
$\mathcal{B} \leftarrow$ `GridEstimation`(GBB, T_0, K)
Run `maxREV`(F_K, β)
if `maxREV` terminates at time $\tau < T - T_0$ **then**
  $\widehat{GFT}_1(p_k, q_j) \leftarrow 0$ for all $(p_k, q_j) \in \mathcal{B}$
  for $t = \tau + 1, \ldots, \tau + T_0$ **do**
    Sample $U_t \sim \mathcal{U}([0, 1])$
    Play $(p_t = U_t, q_t = 0)$
    for $p_k, q_j \in \mathcal{B}$ **do**
      $\widehat{GFT}_1(p_k, q_j) \leftarrow \widehat{GFT}_1(p_k, q_j) + \frac{1}{T_0}\mathbb{I}(s_t \leq U_t \leq p_k)\mathbb{I}(\bar{b}_t \geq q_j)$
    end for
  end for
  $\overline{\mathsf{GFT}}_1(p, q) \leftarrow \widehat{\mathsf{GFT}}_1(p, q) + 2\sqrt{\frac{\ln(\frac{|\mathcal{B}|}{\delta})}{2T_0}} \quad \forall (p, q) \in \mathcal{B}$
  Instantiate an instance of Algorithm `SAE-P`$(\mathcal{A} = \mathcal{B}, \delta)$
  for $t = \tau + T_0 + 1, \ldots, T$ **do**
    Play prices (p_t, q_t) proposed by `SAE-P`
    Observe feedback and build for all $q \in \mathcal{B}^B : q \geq q_t$

$$\mathsf{GFT}_{2,t}(p_t, q) \leftarrow (\bar{b}_t - p_t)\mathbb{I}(\bar{b}_t \geq q)\mathbb{I}(s_t \leq p_t)$$

$$\mathsf{Rev}_t(p_t, q) \leftarrow (\max\{q, \underline{b}_t\} - p_t)\mathbb{I}(\bar{b}_t \geq q)\mathbb{I}(s_t \leq p_t)$$

    Build for all $(p, q) \in \mathcal{B}$ such that $q \geq q_t$ and $p = p_t$

$$c_t(p, q) \leftarrow \frac{1}{1 + \bar{\zeta}}\left(-\mathsf{Rev}_t(p, q) - \bar{\zeta}\right)$$

$$r_t(p, q) \leftarrow \frac{1}{2}\left(\overline{\mathsf{GFT}}_1(p, q) + \mathsf{GFT}_{2,t}(p, q)\right)$$

    Update `SAE-P` with c_t, r_t
  end for
end if

Combining all these components, we can prove that Algorithm 4 guarantees regret of order $\tilde{\mathcal{O}}\left(T^{\frac{2}{3}}\right)$ and satisfies the GBB contraint. Formally:

Theorem 3. *Assume either that* $\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$ *or that* $\mathcal{P}$ *has bounded density. Then, for any* $\delta > 0$*, Algorithm 4 with* $K = \tilde{\mathcal{O}}\left(T^{\frac{1}{3}}\right)$*,* $T_0 = \tilde{\mathcal{O}}\left(T^{\frac{2}{3}}\right)$*,* $\beta = \tilde{\mathcal{O}}\left(T^{2/3}\right)$*, and* $\bar{\zeta} = \tilde{\mathcal{O}}\left(T^{-1/3}\right)$ *guarantees:* $R_T = \tilde{\mathcal{O}}\left(T^{\frac{2}{3}}\right)$ *and* $\sum_{t=1}^{T}\mathsf{Rev}_t(p_t, q_t) \geq 0$*, with probability at least* $1 - 6\delta$*.*

AI USE STATEMENT

Generative AI tools were used for editing the manuscript, including checking the exposition, notation, and consistency of the proof presentation.

REFERENCES

Noga Alon, Nicolo Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *Conference on Learning Theory*, pp. 23–35. PMLR, 2015.

Noga Alon, Nicolo Cesa-Bianchi, Claudio Gentile, Shie Mannor, Yishay Mansour, and Ohad Shamir. Nonstochastic multi-armed bandits with graph-structured feedback. *SIAM Journal on Computing*, 46(6):1785–1826, 2017.

Yossi Azar, Amos Fiat, and Federico Fusco. An α-regret analysis of adversarial bilateral trade. *Advances in Neural Information Processing Systems*, 35:1685–1697, 2022.

Moshe Babaioff, Amitai Frey, and Noam Nisan. Learning to maximize gains from trade in small markets. *arXiv preprint arXiv:2401.11596*, 2024.

Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. Bandits with knapsacks. *Journal of the ACM (JACM)*, 65(3):1–55, 2018.

Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Federico Fusco. No-regret learning in bilateral trade via global budget balance. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pp. 247–258, 2024a.

Martino Bernasconi, Matteo Castiglioni, Andrea Celli, and Federico Fusco. Bandits with replenishable knapsacks: the best of both worlds. In *The Twelfth International Conference on Learning Representations*, 2024b.

Liad Blumrosen and Shahar Dobzinski. Reallocation mechanisms. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pp. 617–617, 2014.

Liad Blumrosen and Yehonatan Mizrahi. Approximating gains-from-trade in bilateral trading. In *Web and Internet Economics: 12th International Conference, WINE 2016, Montreal, Canada, December 11-14, 2016, Proceedings 12*, pp. 400–413. Springer, 2016.

Johannes Brustle, Yang Cai, Fa Wu, and Mingfei Zhao. Approximating gains from trade in two-sided markets via simple mechanisms. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pp. 589–590, 2017.

Jeremy Bulow and Paul Klemperer. Auctions versus negotiations. *American Economic Review*, 86 (1), 1996.

Stéphane Caron, Branislav Kveton, Marc Lelarge, and Smriti Bhagat. Leveraging side observations in stochastic bandits. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, pp. 142–151, 2012.

Matteo Castiglioni, Andrea Celli, Alberto Marchesi, Giulia Romano, and Nicola Gatti. A unifying framework for online optimization with long-term constraints. *Advances in Neural Information Processing Systems*, 35:33589–33602, 2022.

Matteo Castiglioni, Andrea Celli, and Christian Kroer. Online learning under budget and ROI constraints via weak adaptivity. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net, 2024. URL `https://openreview.net/forum?id=shzEkKPrsn`.

Nicolò Cesa-Bianchi, Tommaso R Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. Repeated bilateral trade against a smoothed adversary. In *The Thirty Sixth Annual Conference on Learning Theory*, pp. 1095–1130. PMLR, 2023.

Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. Bilateral trade: A regret minimization perspective. *Mathematics of Operations Research*, 49(1): 171–203, 2024.

Houshuang Chen, Yaonan Jin, Pinyan Lu, and Chihao Zhang. Tight regret bounds for fixed-price bilateral trade. *arXiv preprint arXiv:2504.04349*, 2025.

Alon Cohen, Tamir Hazan, and Tomer Koren. Online learning with feedback graphs without the graphs. In *International Conference on Machine Learning*, pp. 811–819. PMLR, 2016.

Yuan Deng, Jieming Mao, Balasubramanian Sivan, and Kangning Wang. Approximately efficient bilateral trade. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pp. 718–721, 2022.

Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pp. 642–669, 1956.

Yumou Fei. Improved approximation to first-best gains-from-trade. In *International Conference on Web and Internet Economics*, pp. 204–218. Springer, 2022.

Yaonan Jin. Tight regret bounds for bilateral trade under semi feedback. *arXiv preprint arXiv:2601.16412*, 2026.

Zi Yang Kang, Francisco Pernice, and Jan Vondrák. Fixed-price approximations in bilateral trade. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 2964–2985. SIAM, 2022.

Thomas Kesselheim and Sahil Singla. Online learning with vector costs and bandits with knapsacks. In *Conference on Learning Theory*, pp. 2286–2305. PMLR, 2020.

Tomás Kocák, Gergely Neu, and Michal Valko. Online learning with noisy side observations. In *International Conference on Artificial Intelligence and Statistics*, 2016. URL `https://api.semanticscholar.org/CorpusID:14362884`.

Tomáš Kocák, Gergely Neu, Michal Valko, and Remi Munos. Efficient learning by implicit exploration in bandit problems with side observations. *Advances in Neural Information Processing Systems*, 1, 12 2014.

Vijay Krishna. *Auction theory*. Academic press, 2009.

Raunak Kumar and Robert Kleinberg. Non-monotonic resource utilization in the bandits with knapsacks problem. *Advances in Neural Information Processing Systems*, 35:19248–19259, 2022.

Chung-Wei Lee, Haipeng Luo, and Mengxiao Zhang. A closer look at small-loss bounds for bandits with graph feedback. In Jacob Abernethy and Shivani Agarwal (eds.), *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pp. 2516–2564. PMLR, 09–12 Jul 2020.

Thodoris Lykouris, Karthik Sridharan, and Éva Tardos. Small-loss bounds for online learning with partial information. In *Conference on Learning Theory*, pp. 979–986. PMLR, 2018.

Thodoris Lykouris, Eva Tardos, and Drishti Wali. Feedback graph regret bounds for thompson sampling and ucb. In *Algorithmic Learning Theory*, pp. 592–614. PMLR, 2020.

Shie Mannor and Ohad Shamir. From bandits to experts: On the value of side-observations. *Advances in Neural Information Processing Systems*, 24, 2011.

R Preston McAfee. The gains from trade under fixed price mechanisms. *Applied economics research bulletin*, 1(1):1–10, 2008.

Roger B Myerson and Mark A Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of economic theory*, 29(2):265–281, 1983.

Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 2007.

Aleksandrs Slivkins, Karthik Abinav Sankararaman, and Dylan J Foster. Contextual bandits with packing and covering constraints: A modular lagrangian approach via regression. In *The Thirty Sixth Annual Conference on Learning Theory*, pp. 4633–4656. PMLR, 2023.

APPENDIX

The Appendix is structured as follows:

- In Section A, we provide additional related works
- In Section B, we provide the results Omitted from Section 3
- In Section C, we discuss an additive grid for revenue maximization
- In Section D, we provide the proofs Omitted from Section 4
- In Section E, we provide the proofs Omitted from Section 5
- In Section F, we provide the Proofs Omitted from Section 6.1
- In Section G, we provide the proofs Omitted from Section 6.2
- In Section H, we discuss an open problem related to assumption-free regret minimization with GBB constraints

A FURTHER RELATED WORKS

Table 1: Comparison with previous works. "**#b**" denotes the number of buyers, "**FB**" denotes the type of feedback, "**ind.**" denotes that the distribution of buyer(s) and seller valuations are independent, "**b.d.**" denotes that the joint distribution has bounded density, "**b.b.b.**" denotes the type of budget balance of the baseline, while "**b.b.a.**" denotes the one of the algorithm. We denote the semi-feedback by "SF", and we denote by "Asym" the asymmetric feedback. We use the notation "SF§" to facilitate the comparisons in the table, and to highlight that the results also hold for asymmetric feedback, which is slightly weaker than semi-feedback.

	#b	**FB**	**ind.**	**b.d.**	**b.b.b.**	**b.b.a.**	**regret**
Cesa-Bianchi et al. (2024)	1	1 bit	✓	✗	strong	SBB	$\Omega(T)$
			✓	✓	strong	SBB	$\Omega(T^{2/3})$ — $\tilde{\mathcal{O}}\left(T^{2/3}\right)$
			✗	✓	strong	WBB	$\tilde{\mathcal{O}}\left(T^{3/4}\right)$
Cesa-Bianchi et al. (2023)			✗	✓	strong	WBB	$\Omega(T^{3/4})$
Bernasconi et al. (2024a)			✗	✗	strong	GBB	$\Omega(T^{5/7})$ — $\tilde{\mathcal{O}}\left(T^{3/4}\right)$
Chen et al. (2025)		SF	✓	✗	strong	GBB	$\Omega(T^{2/3})$
Jin (2026)			✗	✗	strong	GBB	$\tilde{\mathcal{O}}\left(T^{2/3}\right)$
This paper	1	SF§	✗	✗	strong	WBB	$\tilde{\mathcal{O}}\left(T^{2/3}\right)$
	≥ 2	Asym	✗	✗	strong	WBB	$\Omega(T^{2/3})$ — $\tilde{\mathcal{O}}\left(T^{2/3}\right)$
			✗	✓	global	GBB	$\Omega(T^{2/3})$ — $\tilde{\mathcal{O}}\left(T^{2/3}\right)$
			✓	✗	global	GBB	$\Omega(T^{2/3})$ — $\tilde{\mathcal{O}}\left(T^{2/3}\right)$
			✗	✗	global	GBB	$\Omega(T^{2/3})$ — open

Bayesian mechanism design in two-sided markets Given the impossibility result by Myerson & Satterthwaite (1983), many works try to design approximation algorithms in Bayesian settings. There exist a large class of incentive compatible mechanisms that provide a constant-factor approximation to the social welfare Blumrosen & Dobzinski (2014); Kang et al. (2022) or the gain from trade McAfee (2008); Blumrosen & Mizrahi (2016); Brustle et al. (2017); Deng et al. (2022); Fei (2022).

Bandits with knapsack (and more general) constraints Another line of research relevant for two-sided markets with *global* budget constraints are bandits with knapsacks and, more generally, bandits

with long-term constraints. Indeed, similarly to long-term constraint, the constraint on the revenue must holds globally over all the rounds. Here, we focus on works with *stochastic* constraints, which are the most relevant to our work. The Bandit with knapsack model is introduced by Badanidiyuru et al. (2018), while Kesselheim & Singla (2020) improve the dependence from the number of resources. Subsequent works extend the framework in many directions like non-monotone resource utilization Kumar & Kleinberg (2022); Bernasconi et al. (2024b) and contextual bandits Bernasconi et al. (2024a); Slivkins et al. (2023). Some of these works (see, *e.g.*, Slivkins et al. (2023); Bernasconi et al. (2024a); Castiglioni et al. (2022; 2024)) provide a general framework for addressing bandit with knapsack assuming the existence of two suitable primal and dual regret minimizer. In principle, this approach could be applied to our constrained MAB with graph feedback by leveraging a black-box regret minimizer for standard (unconstrained) MABs with graph feedback Alon et al. (2015). However, these methods heavily rely on the Slater condition, which is essential for achieving optimal regret bounds. Since this condition does not hold in our setting (as we lack a lower bound on revenue), these approaches are ineffective in our setting. As a result, we must develop novel techniques.

MABs with feedback graphs Another line of research related to ours is the study of MABs with graph feedback, which focuses on feedbacks different from the full or bandit feedback. MABs with Feedback graphs are first introduced for adversarial bandits by Mannor & Shamir (2011), while Alon et al. (2017) improve the dependence of the regret bound to $\tilde{\mathcal{O}}\left(\sqrt{T\alpha\log(K)}\right)$ regret bound, where α is the independence number of the graph and K the number of arms. Adversarial MAB with feedback graphs have also been studied in relation to different aspects of the problem like the observability of the graph, the robustness to noise and the data-dependent guarantees (see, *e.g.* Alon et al. (2015); Kocák et al. (2014); Kocák et al. (2016); Lykouris et al. (2018); Cohen et al. (2016); Lee et al. (2020)). Caron et al. (2012) are the first to consider stochastic bandits with feedback graphs. In this setting, Lykouris et al. (2020) analyze the performance of active arm elimination, UCB and Thomson sampling techniques proving a worst-case regret of $\tilde{\mathcal{O}}\left(\sqrt{\alpha T\log(K)}\right)$. All these works focus only on *unconstrained* problems.

B Proofs Omitted from Section 3

Lemma 5. *Fix any couple of prices* $(p,q)\in[0,1]^2$. *Define the functions* $\mathsf{GFT}_1:[0,1]^2\to[-1,1]$ *and* $\mathsf{GFT}_2:[0,1]^2\to[-1,1]$ *as*

$$\mathsf{GFT}_1(p,q)\coloneqq\mathbb{E}_{(\mathbf{b},s)\sim\mathcal{P}}[(p-s)\mathbb{I}(\bar{b}\geq q)\mathbb{I}(s\leq p)],$$

and

$$\mathsf{GFT}_2(p,q)\coloneqq\mathbb{E}_{(\mathbf{b},s)\sim\mathcal{P}}[(\bar{b}-p)\mathbb{I}(\bar{b}\geq q)\mathbb{I}(s\leq p)].$$

Then

$$\mathsf{GFT}(p,q)=\mathsf{GFT}_1(p,q)+\mathsf{GFT}_2(p,q).$$

Moreover, let $\mathcal{U}$ *be the uniform distribution over* $[0,1]$. *Then*

$$\mathsf{GFT}_1(p,q)=\mathbb{P}_{U\sim\mathcal{U},(\mathbf{b},s)\sim\mathcal{P}}(s\leq U\leq p,\bar{b}\geq q).$$

Proof. For all posted couple of prices $(p,q)\in[0,1]^2$ and realized valuations $(\mathbf{b},s)$, the gain from trade can be rewritten as

$$\begin{aligned}(\bar{b}_t-s_t)\mathbb{I}(\bar{b}_t\geq q)\mathbb{I}(s_t\leq p)&=(\bar{b}_t-p+p-s_t)\mathbb{I}(\bar{b}_t\geq q)\mathbb{I}(s_t\leq p)\\&=(\bar{b}_t-p)\mathbb{I}(\bar{b}_t\geq q)\mathbb{I}(s_t\leq p)+(p-s_t)\mathbb{I}(\bar{b}_t\geq q)\mathbb{I}(s_t\leq p).\end{aligned}$$

Taking the expectation over $(\mathbf{b},s)\sim\mathcal{P}$ we get

$$\mathbb{E}[\mathsf{GFT}_t(p,q)]=\mathsf{GFT}_1(p,q)+\mathsf{GFT}_2(p,q).$$

This prove the first part of the statement.

Now, observe that the probability that a random variable U drawn uniformly from $[0,1]$ is in an interval $[x,y]$, $x,y\in[0,1]$, is

$$\mathbb{P}(x\leq U\leq y)=y-x.$$

Therefore, recalling that we defined as $\mathcal{U}$ the uniform distribution over $[0,1]$, for each valuations $(\mathbf{b}, s)$ it holds

$$\mathbb{P}_{U\sim\mathcal{U}}(s \le U \le p, \bar{b} \ge q) = (p-s)\mathbb{I}(s \le p, \bar{b} \ge q).$$

Taking the expectation over $(\mathbf{b}, s) \sim \mathcal{P}$ and recalling the definition of $\mathsf{GFT}_1(p,q)$ the statement easily follows. □

Lemma 6. *Fix any price $q \in [0,1]$. Define the functions $\mathsf{GFT}_1 : [0,1] \to [-1,1]$ and $\mathsf{GFT}_2 : [0,1] \to [-1,1]$ as*

$$\mathsf{GFT}_1(q) \coloneqq \mathbb{E}_{(\mathbf{b},s)\sim\mathcal{P}}[(\max\{q, \underline{b}\} - s)\mathbb{I}(\bar{b} \ge q)\mathbb{I}(s \le \max\{q, \underline{b}\})],$$

and

$$\mathsf{GFT}_2(q) \coloneqq \mathbb{E}_{(\mathbf{b},s)\sim\mathcal{P}}[(\bar{b} - \max\{q, \underline{b}\})\mathbb{I}(\bar{b} \ge q)\mathbb{I}(s \le \max\{q, \underline{b}\})].$$

Then

$$\mathsf{GFT}(q) = \mathsf{GFT}_1(q) + \mathsf{GFT}_2(q).$$

Moreover, let $\mathcal{U}$ be the uniform distribution over $[0,1]$. Then

$$\mathsf{GFT}_1(q) = \mathbb{P}_{U\sim\mathcal{U},(\mathbf{b},s)\sim\mathcal{P}}(s \le U \le \max\{\underline{b}, q\}, \bar{b} \ge q).$$

Proof. This result can be proven by a completely analogous proceeding to the proof of Lemma 5. □

C An Additive-Multiplicative Grid for Revenue

When working with global budget balance algorithms, it is essential to have a way to generate revenue, which will be useful to explore and deal with our approximate estimations. For instance, the process of exploring the decision space imposes a trade-off. In order to explore efficiently and gain useful information, the algorithm must sometimes sacrifice revenue (see Section 6.2). Specifically, it may be required to make decision with immediate negative revenue in order to gather critical information that will help in the future decisions. This creates a tension between the need of accumulating revenue and the necessity of exploring. We deal with this problem accumulating enough revenue in the first rounds to absorb the loss of revenue incurred during the exploration phase.

In order to accumulate revenue, we propose the construction of a finite class of mechanisms, denoted as F_K, similar to the one proposed by Bernasconi et al. (2024a). This class is defined as the union of two distinct families of mechanisms: F_K^+ and F_K^-, and it is designed to maximize the revenue. Formally, we define the two classes of mechanisms:

- *Mechanisms class F_K^+*: this class contains all the couples of prices $(p, p + 2^{-j})$, where $p \in \mathcal{B}^S$ and $j \in [\lceil \log_2(T) \rceil]$. These mechanisms set a reserve price $q = p + 2^{-j}$ to the buyers, while offering to the seller price p, thereby retaining a revenue of at least 2^{-j} if the trade happens.
- *Mechanisms class F_K^-*: this class includes all mechanisms of the form $(\max\{q, \underline{b}_t\} - 2^{-j}, q)$, where $q \in \mathcal{B}^B$ and $j \in [\lceil \log_2(T) \rceil]$. Here, the reserve price is set as q, while the seller is offered a price derived from the auction outcome (the maximum of the reserve price and the second-highest bid) minus 2^{-j}, ensuring a revenue of 2^{-j} if the trades happens.

Then, we let $F_K = F_K^+ \cup F_K^-$. We proceed showing that the optimal mechanism in F_K extracts revenue that is a $O(\log(T))$-approximation of the optimal gain from trade. Intuitively, this shows that we are not incurring too much regret while accumulating revenue (playing the revenue maximizing mechanism in F_K). In particular, we show that such result holds if the seller and buyer distributions are independent or the joint distribution $\mathcal{P}$ has bounded density.

Lemma 7. *Let $\mathcal{B} =$ `GridEstimation`(GBB, T_0, K, δ). Then, if $\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$ and/or $\mathcal{P}$ has density bounded by M, it holds*

$$\mathsf{GFT}(p_{k^*}, q_{j^*}) \le 6\log(T) \max_{(p,q)\in F_K} \mathsf{Rev}(p,q) + 2/T + \zeta.$$

Proof. In the following, to make the notation more agile we let $p = p_{k^*}$ and $q = q_{j^*}$. If $p \geq q$ then, we observe that

$$\begin{aligned}
\mathbb{E}[\left(\bar{b} - p\right) \mathbb{I}(\bar{b} \geq q)\mathbb{I}(s \leq p)] &\leq \mathbb{E}[\left(\bar{b} - p\right) \mathbb{I}(p > \bar{b} \geq q)\mathbb{I}(s \leq p)] + \mathbb{E}[\left(\bar{b} - p\right) \mathbb{I}(\bar{b} \geq p)\mathbb{I}(s \leq p)] \\
&\leq \mathbb{E}[\left(\bar{b} - p\right) \mathbb{I}(\bar{b} \geq p)\mathbb{I}(s \leq p)] \\
&\leq \sum_{j=0}^{\log T} \mathbb{E}\left[\left(\bar{b} - p\right) \mathbb{I}\left(p + \frac{1}{2^j} \leq \bar{b} \leq p + \frac{1}{2^{j-1}}\right) \mathbb{I}(s \leq p)\right] + \frac{1}{T} \\
&\leq 2 \sum_{j=0}^{\log T} \mathbb{E}\left[\frac{1}{2^j} \mathbb{I}\left(p + \frac{1}{2^j} \leq \bar{b} \leq p + \frac{1}{2^{j-1}}\right) \mathbb{I}(s \leq p)\right] + \frac{1}{T} \\
&\leq 3 \log(T) \max_{(p',q') \in F_K^+} \mathsf{Rev}(p', q') + \frac{1}{T},
\end{aligned}$$

where the last inequality holds by observing that $\mathbb{I}(p + \frac{1}{2^j} \leq \bar{b} \leq p + \frac{1}{2^{j-1}})\mathbb{I}(s \leq p) = 1$ implies that posting price $(p, p + 2^{-j})$, which belongs to F_K^+ by construction, the obtained revenue is 2^{-j}.

We apply an analogous reasoning for

$$\begin{aligned}
&\mathbb{E}[(\max\{q, \underline{b}\} - s)\mathbb{I}(\bar{b} \geq q)\mathbb{I}(s \leq p)] \\
&\qquad \leq \mathbb{E}[(\max\{q, \underline{b}\} - s) \mathbb{I}(\bar{b} \geq q)\mathbb{I}(\max\{q, \underline{b}\} < s \leq p)] \\
&\qquad\qquad + \mathbb{E}\left[(\max\{q, \underline{b}\} - s) \mathbb{I}(\bar{b} \geq q)\mathbb{I}(s \leq \max\{q, \underline{b}\})\right] \\
&\qquad \leq \sum_{j=0}^{\log T} \mathbb{E}\left[(\max\{q, \underline{b}\} - s) \mathbb{I}\left(\max\{q, \underline{b}\} - \frac{1}{2^{j-1}} \leq s \leq \max\{q, \underline{b}\} - \frac{1}{2^j}\right)\right] + \frac{1}{T} \\
&\qquad \leq 2 \sum_{j=0}^{\log T} \mathbb{E}\left[2^{-j} \mathbb{I}\left(\max\{q, \underline{b}\} - \frac{1}{2^{j-1}} \leq s \leq \max\{q, \underline{b}\} - \frac{1}{2^j}\right)\right] + \frac{1}{T} \\
&\qquad \leq 3 \log(T) \max_{(p',q') \in F_K^-} \mathsf{Rev}(p', q') + \frac{1}{T},
\end{aligned}$$

where the last inequality holds since if $\mathbb{I}(\max\{q, \underline{b}_t\} - \frac{1}{2^{j-1}} \leq s_t \leq \max\{q, \underline{b}_t\} - \frac{1}{2^j}) = 1$ the revenue achieved by posting prices $(\max\{q, \underline{b}_t\} - 2^{-j}, q)$, which belongs to F_K^-, is at least 2^{-j}.

Similarly, if it holds instead that $p < q$, then

$$\begin{aligned}
\mathbb{E}\left[\left(\bar{b} - p\right) \mathbb{I}(\bar{b} \geq q)\mathbb{I}(s \leq p)\right] &\leq \mathbb{E}\left[\left(\bar{b} - p\right) \mathbb{I}(\bar{b} \geq p)\mathbb{I}(s \leq p)\right] \\
&\leq \sum_{j=0}^{\log T} E\left[\left(\bar{b} - p\right) \mathbb{I}\left(p + \frac{1}{2^j} \leq \bar{b} \leq p + \frac{1}{2^{j-1}}\right) \mathbb{I}(s \leq p)\right] + \frac{1}{T} \\
&\leq 2 \sum_{j=0}^{\log T} \mathbb{E}\left[\frac{1}{2^j} \mathbb{I}\left(p + \frac{1}{2^j} \leq \bar{b}_t \leq p + \frac{1}{2^{j-1}}\right) \mathbb{I}(s_t \leq p)\right] + \frac{1}{T} \\
&\leq 3 \log(T) \max_{(p',q') \in F_K^+} \mathsf{Rev}(p', q') + \frac{1}{T},
\end{aligned}$$

as $p < q$ implies $\left(\bar{b} - p\right) \mathbb{I}(p \leq \bar{b} < q)\mathbb{I}(s \leq p) \geq 0$, and

$$\begin{aligned}
&\mathbb{E}[(\max\{q, \underline{b}\} - s)\mathbb{I}(\bar{b} \geq q)\mathbb{I}(s \leq p)] \\
&\qquad \leq \mathbb{E}[(\max\{q, \underline{b}\} - s) \mathbb{I}(\bar{b} \geq q)\mathbb{I}(s \leq \max\{q, \underline{b}\})] \\
&\qquad \leq \sum_{j=0}^{\log T} \mathbb{E}\left[(\max\{q, \underline{b}\} - s) \mathbb{I}\left(\max\{q, \underline{b}\} - \frac{1}{2^{j-1}} \leq s \leq q - \frac{1}{2^j}\right)\right] + \frac{1}{T} \\
&\qquad \leq 2 \sum_{j=0}^{\log T} \mathbb{E}\left[2^{-j} \mathbb{I}\left(\max\{q, \underline{b}\} - \frac{1}{2^{j-1}} \leq s \leq \max\{q, \underline{b}\} - \frac{1}{2^j}\right)\right] + \frac{1}{T}
\end{aligned}$$

$$\leq 3\log(T) \max_{(p',q')\in F_K^-} \mathsf{Rev}(p',q') + \frac{1}{T}.$$

Combining these results we get;

$$\begin{aligned}
\mathsf{GFT}(p,q) &= \mathbb{E}[\big(\bar{b}-s\big)\,\mathbb{I}(\bar{b}\geq q)\mathbb{I}(s\leq p)] \\
&= \mathbb{E}[\big(\bar{b}-p\big)\,\mathbb{I}(\bar{b}\geq q)\mathbb{I}(s\leq p)] + \mathbb{E}[(\max\{q,\underline{b}\}-s)\,\mathbb{I}(\bar{b}\geq q)\mathbb{I}(s\leq p)] \\
&\qquad + \mathbb{E}[p-\max\{q,\underline{b}\})\mathbb{I}(\bar{b}\geq q)\mathbb{I}(s\leq p)] \\
&\leq \max_{(p',q')\in F_K} 6\log(T)\mathsf{Rev}(p',q') + \frac{2}{T} - \mathsf{Rev}(p,q) \qquad (2a) \\
&\leq 6\log(T) \max_{(p',q')\in F_K} \mathsf{Rev}(p',q') + 2/T + \zeta, \qquad (2b)
\end{aligned}$$

where Inequality (2a) holds by definition of Revenue and the two previous results and Inequality (2b) holds by the definition of ζ. □

D Proofs Omitted from Section 4

Before proving the results on our discretization grids, we state this useful concentration bound from Dvoretzky et al. (1956) in the form presented by Cesa-Bianchi et al. (2024).

Theorem 4 (Theorem 14 in Cesa-Bianchi et al. (2024))**.** *If $(\Omega,\mathcal{F},\mathbb{P})$ is a probability space and $(X_n)_{n\in\mathbb{N}}$ is a $\mathbb{P}$-i.i.d. sequence of random variables, then , for any $\epsilon>0$ and all $m\in\mathbb{N}$, it holds*

$$\mathbb{P}\left[\sup_{x\in\mathbb{R}}\left|\frac{1}{m}\sum_{k=1}^{m}\mathbb{I}\{X_k\leq x\}-\mathbb{P}[X_1\leq x]\right|>\epsilon\right]\leq 2\exp(-2m\epsilon^2).$$

Lemma 1. *Fix T_0, K and let $\mathcal{B}$ =*`GridEstimation`*(SBB, T_0, K). For any $\delta\in(0,1)$, with probability at least $1-\delta$, it holds $\max_{q\in[0,1]}\mathsf{GFT}(q)-\max_{q\in\mathcal{B}}\mathsf{GFT}(q)\leq\frac{1}{K}+2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$.*

Proof. Let $q^*\in[0,1]$ be the price that maximizes the gain from trade. If q^* belong to the grid $\mathcal{B}^B$ the statement trivially follows. Otherwise, by the definition of $\mathcal{B}^B$, there exists a $q_k\in\mathcal{B}^B$ such that $\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{I}(q^*\leq\bar{b}_t<q_k)\leq\frac{1}{T_0}\frac{T_0}{K}=\frac{1}{K}$. Then, applying Theorem 4 with random variable $X_t=\bar{b}_t$ we obtain:

$$\left|\mathcal{P}_{(\mathbf{b},s)\sim\mathcal{P}}(\bar{b}\leq q^*)-\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{I}(\bar{b}_t\leq q^*)\right|\leq\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$$

and

$$\left|\mathcal{P}_{(\mathbf{b},s)\sim\mathcal{P}}(\bar{b}\leq q_k)-\frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{I}(\bar{b}_t\leq q_k)\right|\leq\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$$

with probability at least $1-\delta$.

Hence,

$$\begin{aligned}
\mathcal{P}_{(\mathbf{b},s)\sim\mathcal{P}}(\bar{b}\leq q_k)-\mathcal{P}_{(\mathbf{b},s)\sim\mathcal{P}}(\bar{b}\leq q^*) &\leq \frac{1}{T_0}\sum_{t=1}^{T_0}\mathbb{I}(q^*\leq\bar{b}_t<q_k)+2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}} \\
&\leq \frac{1}{K}+2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}. \qquad (3)
\end{aligned}$$

Now, let's split the valuations $(\mathbf{b},s)$ into three sets:

- V_1 such that $\bar{b}\geq q_k$,
- V_2 such that $\bar{b}\in[q^*,q_k)$,

- V_3 such that $\bar{b} < q^*$.

It is easy to see that when the sampled valuations belongs to V_1 the gain from trade of q_k is larger or equal to the one of q^*. Moreover, when the sampled valuations belong to V_3 the gain from trade of both mechanisms is 0. Finally, when the sampled valuations belong to V_2 the mechanism q^* might get a positive gain from trade (at most 1), while with q_k the trade does not happen (with gain from trade 0. Hence, we can bound

$$\mathsf{GFT}(q^*) - \mathsf{GFT}(q_k) \leq P_{(\mathbf{b},s)\sim\mathcal{P}}((\mathbf{b}, s) \in V_2) \leq \frac{1}{K} + 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}},$$

where the last inequality follows from Equation (3)

□

Lemma 2. *Fix T_0, K and let $\mathcal{B}$ =*`GridEstimation`*(GBB, T_0, K). If $\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$, then for any $\delta \in (0,1)$, with probability at least $1-\delta$, it holds $\mathsf{GFT}(p^*, q^*) - \mathsf{GFT}(p_{k^*}, q_{j^*}) \leq \frac{2}{K} + \sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$. Moreover, if $\mathcal{P}$ admits density bounded by M, then for any $\delta \in (0,1)$, with probability at least $1-\delta$, it holds $\mathsf{GFT}(p^*, q^*) - \mathsf{GFT}(p_{k^*}, q_{j^*}) \leq \frac{M+1}{K} + 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$.*

Proof. We bound the gain from trade in two steps. First, we show how a (p_{k^*}, q^*) on the grid well approximate the optimal couple (p^*, q^*). The proof of this statement is split into two different proofs depending on the underlying assumption. Then, we provide a joint argument for the approximation of (p_{k^*}, q^*) with a (p_{k^*}, q_{j^*}) on the grid.

Independent distributions ($\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$) We focus on the non-trivial case in which it does *not* exists a couple of optimal prices p^*, q^* such that $p^* \in \mathcal{B}^S$. Consider then the optimal prices such that the probability density function of the valuations of the sellers is strictly greater than zero in p^*[2]. Observe that such p^* must exist if $p^* = 0 \in \mathcal{B}^S$ is **not** an optimal price. Then, to be optimal (p^*, q^*) must be such that there exist an $\epsilon' \in (0,1)$ such that $\forall \epsilon \in (0,1)$ such that $\epsilon < \epsilon'$

$$\int_{p^*-\epsilon}^{p^*} \mathbb{E}[(\bar{b} - s)\mathbb{I}(\bar{b} \geq q^*)|s] f_s(s) ds \geq 0,$$

since all prices $[0, p^*] \times \{q^*\}$ are feasible with expected revenue greater or equal to the expected revenue of p^*, q^*. This implies, as ϵ is arbitrarily small and the function $g : [0,1] \to [0,1]$ defined as $g(x) \coloneqq \mathbb{E}[(\bar{b} - x)\mathbb{I}(\bar{b}_t \geq q^*)]$ is continuous, that

$$\mathbb{E}[(\bar{b} - s)\mathbb{I}(\bar{b} \geq q^*)|s = p^*] f_s(p^*) \geq 0,$$

and finally, since $f_s(p^*) > 0$

$$\mathbb{E}[(\bar{b} - p^*)\mathbb{I}(\bar{b} \geq q^*)] \geq 0. \tag{4}$$

Hence:

$$\mathsf{GFT}(p^*, q^*) = \mathbb{E}_{(\mathbf{b},s)\sim\mathcal{P}}[(\bar{b} - s)\mathbb{I}(\bar{b} \geq q^*, s \leq p^*)] \geq 0.$$

Thus,

$$\begin{aligned}\mathbb{E}[\mathsf{GFT}(p^*, q^*) - \mathsf{GFT}(p_{k^*}, q^*)] \\ = -\mathbb{E}[(\bar{b} - s)\mathbb{I}(p^* \leq s < p_{k^*})\mathbb{I}(\bar{b} \geq q^*)]\end{aligned}$$

[2]By convention we say that a non-absolute random variable X that has non-zero probability in the set $\{x\}$ of null Lebesgue measure present a density function $f_X(x) = +\infty$

$$
\begin{aligned}
&\leq -\mathbb{E}[(\bar{b} - p_{k^*})\mathbb{I}(p^* \leq s < p_{k^*})\mathbb{I}(\bar{b} \geq q^*)] \\
&\leq -\mathbb{E}[(\bar{b} - p^*)\mathbb{I}(\bar{b}_t \geq q^*)]\mathbb{P}(p^* \leq s < p_{k^*}) \\
&- \mathbb{E}[(p^* - p_{k^*})\mathbb{I}(\bar{b}_t \geq q^*)]\mathbb{P}(p_{k^*} < s \leq p^*) \\
&\leq 0 + \frac{1}{K},
\end{aligned}
$$

where the first inequality follows since the function is non-zero only for $p^* \leq s < p_{k^*}$ and the last from Equation (4) and since $-(p^* - p_{k^*}) \leq \frac{1}{K}$ by construction.

$\mathcal{P}$ has density bounded by M Similarly to the independent distribution case, we observe that the gain from trade of (p^*, q^*) and (p_{k^*}, q^*) differs only if $p_{k^*} < s \leq p^*$. Here, we exploit the bounded density assumption to derive

$$\mathbb{P}_{(\mathbf{b},s)\sim\mathcal{P}}(p_{k^*} < s \leq p^*) \leq M/K.$$

Hence:

$$\mathbb{E}[\mathsf{GFT}_t(p^*, q^*) - \mathsf{GFT}_t(p_{k^*}, q^*)] \leq \mathbb{P}(p^* \leq s_t < p_{k^*}) \leq \frac{M}{K}.$$

Now, we show that (p_{k^*}, q_{j^*}) has a gain from trade similar to (p_{k^*}, q^*) concluding the proof. Consider the non-trivial case $q^* \notin \mathcal{B}^B$. Then, similarly as we proved in Equation (3) of Lemma 1, we can prove that with probability at least $1 - \delta$ it holds:

$$\mathcal{P}_{(\mathbf{b},s)\sim\mathcal{P}}(\bar{b}_t \leq q^*) - \mathcal{P}_{(\mathbf{b},s)\sim\mathcal{P}}(\bar{b}_t \leq q_{j^*}) \leq \frac{1}{K} + 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}.$$

Moreover, similarly to the proof of Lemma 1 we observe that the gain from trade of the two mechanisms differs only when $\bar{b} \in [q^*, q_{j^*})$, and such difference is at most 1. Hence,

$$\mathsf{GFT}(p_{k^*}, q^*) - \mathsf{GFT}_t(p_{k^*}, q_{j^*}) \leq P_{(\mathbf{b},s)\sim\mathcal{P}}(\bar{b} \in [q^*, q_{j^*})) \leq \frac{1}{K} + \sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}},$$

□

Next, we show that has (p_{k^*}, q_{j^*}) has a not too negative revenue, making it a viable option in the design of GBB mechanisms.

Lemma 3. *Fix T_0, K and let $\mathcal{B} = \texttt{GridEstimation}(GBB, T_0, K)$. If $\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$, then for any $\delta \in (0,1)$, with probability at least $1-\delta$, it holds $\zeta \leq \frac{2}{K} + 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}}$. Moreover, if $\mathcal{P}$ has density bounded by M, then for any $\delta \in (0,1)$, with probability at least $1-\delta$, it holds that $\zeta \leq \frac{2M}{K}$.*

Proof. We divide the proof in two parts depending on the assumption on the distribution.

Independent distributions ($\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$) First, we notice that, since (p^*, q^*) has expected non-negative revenue and sellers and buyers valuations are independent, then

$$
\begin{aligned}
0 \geq \mathsf{Rev}(p^*, q^*) &= \mathbb{E}_{(s,\mathbf{b})\sim\mathcal{P}}[(\max\{\underline{b}, q^*\} - p^*)\mathbb{I}(s \leq p^*)\mathbb{I}(\bar{b} \geq q^*)] \\
&= \mathbb{E}_{(s,\mathbf{b})\sim\mathcal{P}}[(\max\{\underline{b}, q^*\} - p^*)|s \leq p^*, \bar{b} \geq q^*]\mathbb{P}(s \leq p^*)\mathbb{P}(\bar{b} \geq q^*) \\
&= \mathbb{E}_{\mathbf{b}\sim\mathcal{P}^B}[(\max\{\underline{b}, q^*\} - p^*)|\bar{b} \geq q^*]\mathbb{P}(s \leq p^*)\mathbb{P}(\bar{b} \geq q^*)].
\end{aligned}
$$

which implies that either

$$\mathbb{E}_{\mathbf{b}\sim\mathcal{P}^B}\left[\max\{\underline{b}, q^*\} - p^*|\bar{b}_t \geq q^*\right] \geq 0, \tag{5}$$

or $\mathbb{P}(s_t \leq p^*)\mathbb{P}(\bar{b}_t \geq q^*) = 0$. Notice that the second case is trivial as it would imply $\mathbb{P}(s_t \leq p_{k^*})\mathbb{P}(\bar{b}_t \geq q_{j^*}) = 0$ and therefore $\mathsf{Rev}(p_{k^*}, q_{j^*}) = 0$.

Thus, considering the non-trivial case $\mathbb{P}(s_t \le p^*)\mathbb{P}(\bar{b}_t \ge q^*) > 0$, with probability at least $1-\delta$ it holds

$$
\begin{aligned}
&\mathsf{Rev}(p_{k^*}, q_{j^*}) - \mathsf{Rev}(p^*, q^*) \\
&= \mathbb{E}[(\max\{q_{j^*}, \underline{b}\} - p_{k^*})\mathbb{I}(\bar{b} \ge q_{j^*})\mathbb{I}(s \le p_{k^*})] - \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q^*)\mathbb{I}(s \le p^*)] \\
&\ge \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q_{j^*})\mathbb{I}(s \le p_{k^*})] + (p^* - p_{k^*}) - \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q^*)\mathbb{I}(s \le p^*)] \\
&\ge \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q^*)\mathbb{I}(s \le p_{k^*})] - \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(q^* \le \bar{b} < q_{j^*})\mathbb{I}(s \le p_{k^*})] \\
&\qquad + (p^* - p_{k^*}) - \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q^*)\mathbb{I}(s \le p^*)] \\
&\ge \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q^*)]\mathbb{P}(s \le p_{k^*}) - \mathbb{P}(q^* \le \bar{b} < q_{j^*}) \\
&\qquad + (p^* - p_{k^*}) - \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q^*)]\mathbb{P}(s \le p^*) \\
&= \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q^*)]\left(\mathbb{P}(s \le p_{k^*}) - \mathbb{P}(s \le p^*)\right) - \mathbb{P}(q^* \le \bar{b} < q_{j^*}) + (p^* - p_{k^*}) \\
&\ge 0 - \frac{1}{K} - 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}} - \frac{1}{K} \\
&= -\frac{2}{K} - 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2T_0}},
\end{aligned}
$$

where we used $p^* \le p_{k^*}$ and $q^* \le q_{j^*}$, and in the last inequality we used Equation (3) and Equation (5).

$\mathcal{P}$ has density bounded by M Here, we simply observe that

$$
\begin{aligned}
&\mathsf{Rev}(p_{k^*}, q_{j^*}) - \mathsf{Rev}(p^*, q^*) \\
&\quad = \mathbb{E}[(\max\{q_{j^*}, \underline{b}\} - p_{k^*})\mathbb{I}(\bar{b} \ge q_{j^*})\mathbb{I}(s \le p_{k^*})] - \mathbb{E}[(\max\{q^*, \underline{b}\} - p^*)\mathbb{I}(\bar{b} \ge q^*)\mathbb{I}(s \le p^*)] \\
&\quad \ge -\mathbb{P}(q^* \le \bar{b}_t \le q_{j^*}) - \mathbb{P}(p^* \le s_t \le p_{k^*}) \\
&\quad \ge -\frac{2M}{K}.
\end{aligned}
$$

□

Now, we observe that the revenue maximizing prices on the F_K defined in appendix C can be learned easily using a general purpose regret minimizer since: 1) we have bandit feedback on the revenue, and 2) the size of the mechanism set satisfies $|F_K| \le 3K\log(T)$.

Lemma 4. *Fix $T_0 = \tilde{\mathcal{O}}\left(T^{2/3}\right), K = \tilde{\mathcal{O}}\left(T^{1/3}\right)$. There exists a learning algorithm* `maxREV` *that, for any $\delta \in (0,1)$, with probability at least $1-2\delta$, guarantees:*

$$
\tau \cdot \mathsf{GFT}(p_{k^*}, q_{j^*}) \le 6\log(T) \sum_{t\in[\tau]} \mathsf{Rev}_t(p_t, q_t) + \tilde{\mathcal{O}}\left(T^{2/3}\right) + \tau\zeta \quad \forall \tau \in [T].
$$

Proof. We can use UCB1 on the set of arms F_K to implement the `maxREV` algorithm. This guarantees that with probability at least $1-\delta$.

$$
\sum_{t\in[\tau]} \mathsf{Rev}(p_t, q_t) \ge \tau \mathsf{Rev}(p_j^*, q_j^*) - \tilde{O}(T^{2/3}).
$$

Moreover, an Azuma-Hoeffding inequality implies:

$$
\sum_{t\in[\tau]} \mathsf{Rev}_t(p_t, q_t) \ge \sum_{t\in[\tau]} \mathsf{Rev}(p_t, q_t) - \tilde{O}(\sqrt{T}),
$$

with probability at least $1-\delta$.

Hence, taking an union bound and applying lemma 7, we get:

$$
\tau\mathsf{GFT}(p_k^*, q_j^*) \le 6\log(T)\tau \max_{(p',q')\in F_K} \mathsf{Rev}(p', q') + 2 + \tau\zeta
$$

$$\leq 6\log(T)\sum_{t\in[\tau]}\mathsf{Rev}(p_t,q_t)+\tilde{O}(T^{2/3})+\tau\zeta$$
$$\leq 6\log(T)\sum_{t\in[\tau]}\mathsf{Rev}_t(p_t,q_t)+\tilde{O}(T^{2/3})+\tau\zeta.$$

□

E PROOFS OMITTED FROM SECTION 5

Theorem 1. *For any $\delta > 0$, Algorithm 2 with $T_0 = \tilde{\mathcal{O}}\left(T^{2/3}\right)$ and $K = \tilde{\mathcal{O}}\left(T^{1/3}\right)$ achieves: $R_T = \tilde{\mathcal{O}}\left(T^{2/3}\right)$, with probability at least $1-3\delta$.*

Proof. In this proof we will use q^* to denote the optimum price, *i.e.* $q^* \in \arg\max_{q\in[0,1]}\mathsf{GFT}(q)$, and q_{k^*} to denote the optimum price over the grid $\mathcal{B}$, *i.e.* $q_{k^*} \in \arg\max_{q\in\mathcal{B}}\mathsf{GFT}(q)$. First we can decompose the regret in 5 different components representing respectively the regret induced by the approximation to the grid $\mathcal{B}$ (R_K), the regret generated in the pure exploration phase (component (1)), the regret induced by the error of estimation of the actual GFT (components(2) and (4)) and the regret induced by the algorithm over the optimistic estimate of GFT ((3)).

Formally:

$$\begin{aligned}
R_T &= T\cdot\mathsf{GFT}(q^*)-\sum_{t=1}^{T}\mathsf{GFT}(q_t)\\
&=\underbrace{\sum_{t=1}^{2T_0}(\mathsf{GFT}(q^*)-\mathsf{GFT}(q_t))}_{(1)}+\underbrace{\sum_{t=2T_0+1}^{T}(\mathsf{GFT}(q^*)-\mathsf{GFT}(q_{k^*}))}_{R_K}\\
&+\underbrace{\sum_{t=2T_0+1}^{T}(\mathsf{GFT}(q_{k^*})-\overline{\mathsf{GFT}}_t(q_{k^*}))}_{(2)}\\
&+\underbrace{\sum_{t=2T_0+1}^{T}(\overline{\mathsf{GFT}}_t(q_{k^*})-\overline{\mathsf{GFT}}_t(q_t))}_{(3)}+\underbrace{\sum_{t=2T_0+1}^{T}(\overline{\mathsf{GFT}}_t(q_t)-\mathsf{GFT}(q_t))}_{(4)}\\
&=R_K+(1)+(2)+(3)+(4).
\end{aligned}$$

In the following, we proceed to bound each component separately.

Bound (1) : It can be bounded trivially by $2T_0$.

Bound R_K: With probability at least $1-\delta$ by Lemma 1

$$R_K \leq \tilde{\mathcal{O}}\left(\frac{T}{K}+\frac{T}{\sqrt{T_0}}\right).$$

Bound (2)**:** To bound (2) we first prove that by construction with probability at least $1-2\delta$ the estimated game for trade $\overline{\mathsf{GFT}}_t(q)$ is greater or equal than $\mathsf{GFT}(q)$ for all rounds $t\in[2T_0+1,\ldots,T]$ and for all $q\in\mathcal{B}$.

First observe that $\widehat{\overline{\mathsf{GFT}}}_1$ is such that

$$\mathbb{E}[\widehat{\overline{\mathsf{GFT}}}_1(q_k)]=\mathbb{E}\left[\frac{1}{T_0}\sum_{t=T_0+1}^{2T_0}\mathbb{I}(s_t\leq U_t\leq\max\{q_k,\underline{b}_t\})\mathbb{I}(\bar{b}_t\geq q_k)\right]$$

$$= \mathsf{GFT}_1(q_k)$$

using Lemma 5.

In addition, it holds that $|\mathbb{I}(s_t \leq U_t \leq \max\{q_k, \underline{b}_t\})\mathbb{I}(\bar{b}_t \geq q_k)| \leq 1$.

Thus by Hoeffding inequality with probability at least $1-\delta$ for all $q_k \in \mathcal{B}$:

$$\left|\widehat{\mathsf{GFT}}_1(q_k) - \mathsf{GFT}_1(q_k)\right| \leq 2\sqrt{\frac{\ln(\frac{2K}{\delta})}{2T_0}}, \tag{6}$$

and therefore with probability at least $1-\delta$:

$$\overline{\mathsf{GFT}}_1(q_k) = \widehat{\mathsf{GFT}}_1(q_k) + 2\sqrt{\frac{\ln(\frac{2K}{\delta})}{2T_0}} \geq \mathsf{GFT}_1(q_k).$$

For what concern GFT_2 observe that $\widehat{\mathsf{GFT}}_{t,2}$ is, for all q_k in $\mathcal{B}$ and for all $t \in [2T_0+1, T]$, defined as

$$\widehat{\mathsf{GFT}}_{t,2}(q_k) = \frac{1}{n_t(q_k)} \sum_{t=2T_0+1}^{t} (-\max\{q_k, \underline{b}_t\} + \bar{b}_t)\mathbb{I}(\bar{b}_t \geq q_k)\mathbb{I}(s_t \leq \max\{\underline{b}_t, q_k\}).$$

Thus, by applying also in this case the Hoeffding inequality for all $q_k \in \mathcal{B}$, $t \in [2T_0+1, \ldots, T]$ with probability at least $1-\delta$ it holds:

$$\left|\widehat{\mathsf{GFT}}_{2,t}(q_k) - \mathsf{GFT}_2(q_k)\right| \leq 2\sqrt{\frac{\ln(\frac{2KT}{\delta})}{2n_t(q_k)}}, \tag{7}$$

and therefore

$$\overline{\mathsf{GFT}}_{t,2}(q_k) = \widehat{\mathsf{GFT}}_{2,t}(q_k) + \sqrt{\frac{\ln(\frac{2KT}{\delta})}{2n_t(q_k)}} \geq \mathsf{GFT}_2(q_k).$$

Thus, with probability at least $1-2\delta$

$$(2) \leq 0.$$

Bound (3)**:** By construction of the algorithm $(3) \leq 0$.

Bound (4)**:** To bound (4) we first observe that by Inequality (6) with probability at least $1-\delta$ for all $q \in \mathcal{B}$

$$\overline{\mathsf{GFT}}_1(q_k) - \mathsf{GFT}_1(q_k) \leq 2\left(2\sqrt{\frac{\ln(\frac{2K}{\delta})}{2T_0}}\right),$$

and by Inequality (7) it holds with probability at least $1-\delta$ for all $q_k \in \mathcal{B}$ and for all $t \in [2T_0+1, T]$

$$\overline{\mathsf{GFT}}_{t,2}(q_k) - \mathsf{GFT}_2(q_k) \leq 2\left(2\sqrt{\frac{\ln(\frac{2KT}{\delta})}{2n_t(q_k)}}\right),$$

where $n_t(q_k)$ is the counter of in how many rounds from $2T_0+1$ to t the algorithm played the price q_k. Therefore, with probability at least $1-2\delta$

$$\begin{aligned}
(4) &= \sum_{t=2T_0+1}^{T} (\overline{\mathsf{GFT}}_t(q_t) - \mathsf{GFT}(q_t)) \\
&\leq 4(T-2T_0)\sqrt{\frac{\ln(\frac{4KT}{\delta})}{2T_0}} + 4\sum_{t=2T_0+1}^{T}\sqrt{\frac{\ln(\frac{4KT}{\delta})}{n_t(q_t)}} \\
&= \tilde{\mathcal{O}}\left(\frac{T}{\sqrt{T_0}} + \sqrt{KT}\right).
\end{aligned}$$

To conclude the proof we join all the results and we observe that with probability at least $1 - 3\delta$

$$\begin{aligned} R_t &= R_K + (1) + (2) + (3) + (4) \\ &\leq \tilde{\mathcal{O}}\left(\frac{T}{K} + \frac{T}{\sqrt{T_0}}\right) + \tilde{\mathcal{O}}\left(T_0\right) + \tilde{\mathcal{O}}\left(\frac{T}{\sqrt{T_0}} + \sqrt{KT}\right) \\ &\leq \tilde{\mathcal{O}}\left(\frac{T}{K} + \frac{T}{\sqrt{T_0}} + T_0 + \sqrt{KT}\right) \\ &= \tilde{\mathcal{O}}\left(T^{2/3}\right), \end{aligned}$$

where the last equality holds if we fix $K = \tilde{\mathcal{O}}\left(T^{\frac{1}{3}}\right), T_0 = \tilde{\mathcal{O}}\left(T^{\frac{2}{3}}\right)$.

□

F Proofs Omitted from Section 6.1

Theorem 2. *Algorithm 3 is such that with probability at least* $1 - 2\delta$*:*

$$\tau \cdot r(a^*) - \sum_{t=1}^{\tau} r(a_t) \leq \tilde{\mathcal{O}}\left(\sqrt{nT}\log(m/\delta)\right) \forall \tau \in [T]$$

and $\sum_{t=1}^{\tau}[c(a_t)]^+ \leq \tilde{\mathcal{O}}\left(\sqrt{nT}\log(m/\delta)\right) \forall \tau \in [T]$.

Proof. The proof can be divided in two parts; in the first we prove that the optimal arm belongs with high probability to the safe set at each round and in the second we actually bound the regret between the rewards of the arms chosen and an arbitrary arm in the safe set.

Optimal arm belongs to $\bigcap_{t\in[T]} \mathcal{S}_t$ First we prove the following two auxiliary results:

1. with probability at least $1 - \delta$ for all $t \in [T]$ and for all $a_{i,j} \in \mathcal{A}$
$$r(a_{i,j}) \in [\underline{r}_t(a_{i,j}), \overline{r}_t(a_{i,j})].$$
Since $\mathbb{E}[\widehat{r}_t(a_{i,j})] = \mathbb{E}[\frac{1}{n_t(a_{i,j})}\sum_{\ell=1}^{t} r_\ell(a_{i,j})\mathbb{I}(a_{i,j} \in Obs_\ell)] = \frac{n_t(a_{i,j})r(a_{i,j})}{n_t(a_{i,j})} = r(a_{i,j})$, where $n_t(a_{i,j})$ is the number of times the reward of arm $a_{i,j}$ is observed up to round t and where $\mathbb{I}(a_{i,j} \in Obs_\ell) = 1$ means that at round ℓ, as consequence of playing arm a_{i_ℓ,j_ℓ} the reward of arm $a_{i,j}$ is also disclosed, and since $|r_\ell(a_{i,j})\mathbb{I}(a_{i,j} \in Obs_\ell)| \leq 1$ for all $\ell \in [T]$, then by Hoeffding inequality with probability at least $1 - \delta$
$$\left|\widehat{r}_t(a_{i,j}) - r(a_{i,j})\right| \leq 2\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_t(a_{i,j})}}.$$

2. With probability at least $1 - \delta$ for all $t \in [T]$ and for all $(a_{i,j}) \in \mathcal{A}$
$$c(a_{i,j}) \in [\underline{c}_t(a_{i,j}), \overline{c}_t(a_{i,j})].$$
We prove this point in an analogous way. By means of Hoeffding inequality, recalling the definition of $\widehat{c}_t(a_{i,j}) = \frac{1}{n_t(a_{i,j})}\sum_{\ell=1}^{t} c_\ell(a_{i,j})\mathbb{I}(a_{i,j} \in Obs_\ell)]$, it holds with probability at least $1 - \delta$ for all $t \in [T]$ and for all $a_{i,j} \in \mathcal{A}$
$$\left|\widehat{c}_t(a_{i,j}) - c(a_{i,j})\right| \leq 2\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_t(a_{i,j})}}.$$

Finally, we are ready to prove that $a_{i^*,j^*} \in \mathcal{S}_t$ for all $t \in [T]$. Indeed there are two possible mechanisms of arms elimination.

The first is straightforward and consists in eliminating all arms for which it exists a t such that $\underline{c}_t(a_{i,j}) \geq 0$, which, with probability at least $1-\delta$, does not eliminate a_{i^*,j^*} since by definition $c(a_{i^*,j^*}) \leq 0$.

The second method is less obvious and eliminates only arms $a_{i,j'}$ in which there exists $t \in [T]$ and $a_{i,j} \in \mathcal{S}_t$ such that $\overline{r}(a_{i,j'}) \leq \underline{r}(a_{i,j})$ and $j' \leq j$. We distinguish two possible cases, either $c(a_{i,j}) \leq 0$ or $c(a_{i,j}) > 0$. If $c(a_{i,j}) \leq 0$ then $a_{i,j}$ is feasible and $a_{i,j'}$ is with high probability sub-optimal, indeed with probability at least $1-\delta$ it holds $r(a_{i,j'}) \leq \overline{r}(a_{i,j'}) \leq \underline{r}(a_{i,j}) \leq r(a_{i,j})$. Otherwise, if $c(a_{i,j}) > 0$ then $c(a_{i,j'}) \geq c(a_{i,j}) > 0$ and $a_{i,j'}$ is not feasible.

Bounding regret The regret can be decomposed in four parts:

$$\begin{aligned}
\tau \cdot r(a_{i^*,j^*}) - \sum_{t=1}^{\tau} r(a_{i_t,j_t}) = \tau \cdot r(a_{i^*,j^*}) - \sum_{t=1}^{\tau} \overline{r}_{t-1}(a_{i^*,j^*}) \\
+ \sum_{t=1}^{\tau} (\overline{r}_{t-1}(a_{i^*,j^*}) - \overline{r}_{t-1}(a_{i_t,\bar{j}_t})) \\
+ \sum_{t=1}^{\tau} (\overline{r}_{t-1}(a_{i_t,\bar{j}_t}) - \overline{r}_{t-1}(a_{i_t,j_t})) \\
+ \sum_{t=1}^{\tau} (\overline{r}_{t-1}(a_{i_t,j_t}) - r(a_{i_t,j_t})).
\end{aligned}$$

The first term is with probability at least $1-\delta$ upper bounded by 0 by construction of $\overline{r}$. The second term is also upper bounded by 0 with probability at least $1-2\delta$ as $a_{i_t,\bar{j}_t}$ is the arm that maximize $\overline{r}_{t-1}$ in $\mathcal{S}_{t-1}$ and as we proved above a_{i^*,j^*} is in $\mathcal{S}_{t-1}$ with probability at least $1-2\delta$. The third term can be further decomposed as

$$\begin{aligned}
\sum_{t=1}^{\tau} (\overline{r}_{t-1}(a_{i_t,\bar{j}_t}) - \overline{r}_{t-1}(a_{i_t,j_t})) &= \sum_{t=1}^{\tau} \left(\underline{r}_{t-1}(a_{i_t,\bar{j}_t}) + 4\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_{t-1}(a_{i_t,\bar{j}_t})}} - \overline{r}_{t-1}(a_{i_t,j_t}) \right) \\
&\leq \sum_{t=1}^{\tau} 4\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_{t-1}(a_{i_t,\bar{j}_t})}} \\
&= \sum_{t=1}^{\tau} 4\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_{t-1}(i_t)}} \\
&= \tilde{\mathcal{O}}\left(\sqrt{\tau n \log(nmT/\delta)}\right),
\end{aligned}$$

since by construction $n_{t-1}(a_{i_t,\bar{j}_t}) = n_{t-1}(a_{i_t,j'})$ for all $a_{i_t,j'} \in \mathcal{S}_{t-1}$, and we indicated that counter as simply $n_{t-1}(i_t)$.

The fourth term can be treated similarly, hence with probability at least $1-\delta$

$$\begin{aligned}
\sum_{t=1}^{\tau} (\overline{r}_{t-1}(a_{i_t,j_t}) - r(a_{i_t,j_t})) &= \sum_{t=1}^{\tau} \left(\underline{r}_{t-1}(a_{i_t,j_t}) + 4\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_{t-1}(a_{i_t,j_t})}} - r(a_{i_t,j_t}) \right) \\
&\leq \sum_{t=1}^{\tau} \left(4\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_{t-1}(a_{i_t,j_t})}} \right) \\
&= \sum_{t=1}^{\tau} \left(4\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_{t-1}(i_t)}} \right) \\
&= \tilde{\mathcal{O}}\left(\sqrt{\tau n \log(nmT/\delta)}\right).
\end{aligned}$$

Thus, joining all the bounds we get that with probability at least $1-2\delta$ for all $\tau \in [T]$ it holds

$$\tau \cdot r(a_{i^*,j^*}) - \sum_{t=1}^{\tau} r(a_{i_t,j_t}) \leq \tilde{\mathcal{O}}\left(\sqrt{\tau n \log(nmT/\delta)}\right).$$

Costs constraints To conclude the proof we observe that with probability at least $1-\delta$ for all $\tau \in [T]$

$$\begin{aligned}
\sum_{t=1}^{\tau}[c(a_{i_t,j_t})]^+ &\leq \sum_{t=1}^{\tau}[\overline{c}_{t-1}(a_{i_t,j_t})]^+ \\
&= \sum_{t=1}^{\tau}\left[\left(\underline{c}_{t-1}(a_{i_t,j_t}) + 4\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_{t-1}(i_t)}}\right)\right]^+ \\
&\leq \sum_{t=1}^{\tau} 4\sqrt{\frac{\ln(\frac{2nmT}{\delta})}{2n_{t-1}(i_t)}} \\
&= \tilde{\mathcal{O}}\left(\sqrt{\tau n \log(nmT/\delta)}\right).
\end{aligned}$$

□

G Proofs Omitted from Section 6.2

Theorem 3. *Assume either that $\mathcal{P} = \mathcal{P}^S \times \mathcal{P}^B$ or that $\mathcal{P}$ has bounded density. Then, for any $\delta > 0$, Algorithm 4 with $K = \tilde{\mathcal{O}}\left(T^{\frac{1}{3}}\right)$, $T_0 = \tilde{\mathcal{O}}\left(T^{\frac{2}{3}}\right)$, $\beta = \tilde{\mathcal{O}}\left(T^{2/3}\right)$, and $\bar{\zeta} = \tilde{\mathcal{O}}\left(T^{-1/3}\right)$ guarantees: $R_T = \tilde{\mathcal{O}}\left(T^{\frac{2}{3}}\right)$ and $\sum_{t=1}^{T} \mathsf{Rev}_t(p_t, q_t) \geq 0$, with probability at least $1-6\delta$.*

Proof. We first prove the statement relative to the regret.

$$\begin{aligned}
R_T &= T \cdot \mathsf{GFT}(p^*, q^*) - \sum_{t=1}^{T} \mathbb{E}[\mathsf{GFT}(p_t, q_t)] \\
&= \underbrace{T_0 \cdot \mathsf{GFT}(p^*, q^*) - \sum_{t=\tau+1}^{\tau+T_0} \mathbb{E}[\mathsf{GFT}(p_t, q_t)]}_{(1)} + \underbrace{(T-T_0)\mathsf{GFT}(p^*, q^*) - (T-T_0)\mathsf{GFT}(p_{k^*}, q_{j^*})}_{R_K} \\
&\quad + \underbrace{(\tau)\mathsf{GFT}(p_{k^*}, q_{j^*}) - \sum_{t=1}^{\tau} \mathbb{E}[\mathsf{GFT}(p_t, q_t)]}_{R_{rev}} \\
&\quad + \underbrace{T_0 \cdot \mathsf{GFT}(p_{k^*}, q_{j^*}) - \sum_{t=1}^{T_0} \mathbb{E}[\mathsf{GFT}(p_t, q_t)]}_{(2)} \\
&\quad + \underbrace{(T-\tau-T_0)\mathsf{GFT}_t(p_{k^*}, q_{j^*}) - \sum_{t=\tau+T_0+1}^{T} \mathbb{E}\left[\mathsf{GFT}_t(p_t, q_t)\right]}_{(3)} \\
&= R_K + R_{rev} + (1) + (2) + (3).
\end{aligned}$$

Bound R_K: With probability at least $1-\delta$ by Lemma 2 if $\mathcal{P}$ is bounded and/or sellers and buyers valuations are independent

$$R_K \leq \tilde{\mathcal{O}}\left(\frac{T}{K} + \frac{T}{\sqrt{T_0}}\right) = \tilde{\mathcal{O}}\left(T^{2/3}\right).$$

Bound R_{rev}: By Lemma 4 with probability at least $1-2\delta$

$$R_{\mathsf{Rev}} \leq \tilde{\mathcal{O}}\left(\beta + \tau\zeta + T^{2/3}\right) = \tilde{\mathcal{O}}\left(T^{2/3}\right).$$

Bound (1): (1) can be bounded trivially by T_0

$$(1) \leq T_0 = \tilde{\mathcal{O}}\left(T^{2/3}\right).$$

Bound (2): (2) can be also bounded trivially by T_0

$$(2) \leq T_0 = \tilde{\mathcal{O}}\left(T^{2/3}\right).$$

Bound (3): If the distributions of sellers and buyers are independent and/or $\mathcal{P}$ is bounded it exists a choice of $\bar{\zeta} = \tilde{\mathcal{O}}\left(T^{-1/3}\right)$ for which with probability at least $1-\delta$ $\mathsf{Rev}(p_{k^*}, q_{j^*}) \geq -\bar{\zeta}$ by Lemma 3. If the algorithm is initiated with that $\bar{\zeta}$ then it holds with probability at least $1-\delta$ that $c(p_{k^*}, q_{j^*}) = \mathbb{E}[c_t(p_{k^*}, q_{j^*})] = \mathbb{E}[\frac{1}{\zeta+1}(-\mathsf{Rev}_t(p_{k^*}, q_{j^*}) - \bar{\zeta})] = \frac{1}{\zeta+1}(-\mathsf{Rev}(p_{k^*}, q_{j^*}) - \bar{\zeta}) \leq 0$. Hence, we can apply theorem 2 to bound (3) and with probability at least $1-3\delta$ it holds

$$\begin{aligned}
(3) &= (T-\tau-T_0)\mathsf{GFT}(p_{k^*}, q_{j^*}) - \sum_{t=\tau+T_0+1}^{T} \mathsf{GFT}(p_t, q_t) \\
&= \sum_{t=\tau+T_0+1}^{T} (\mathsf{GFT}_1(p_{k^*}, q_{j^*}) + \mathsf{GFT}_2(p_{k^*}, q_{j^*})) - \sum_{t=\tau+T_0+1}^{T} (\mathsf{GFT}_1(p_t, q_t) + \mathsf{GFT}_2(p_t, q_t)) \\
&= \sum_{t=\tau+T_0+1}^{T} \left(\mathsf{GFT}_1(p_{k^*}, q_{j^*}) - \overline{\mathsf{GFT}}_1(p_{k^*}, q_{j^*})\right) + \sum_{t=\tau+T_0+1}^{T} \left(\overline{\mathsf{GFT}}_1(p_t, q_t) - \mathsf{GFT}_1(p_t, q_t)\right) \\
&\qquad + 2\sum_{t=\tau+T_0+1}^{T} \left(\frac{1}{2}(\overline{\mathsf{GFT}}_1(p_{k^*}, q_{j^*}) + \mathsf{GFT}_2(p_{k^*}, q_{j^*})) - \frac{1}{2}(\overline{\mathsf{GFT}}_1(p_t, q_t) + \mathsf{GFT}_2(p_t, q_t))\right) \\
&\leq 8\sqrt{\frac{\ln(2|\mathcal{B}|/\delta)}{2T_0}}(T-T_0-\tau) + 2\sum_{t=\tau+T_0+1}^{T} (r(p_{k^*}, q_{j^*}) - r(p_t, q_t)) \\
&\leq \tilde{\mathcal{O}}\left(\frac{T}{\sqrt{T_0}}\right) + \tilde{\mathcal{O}}\left(\sqrt{KT\log(K/\delta)}\right) \\
&= \tilde{\mathcal{O}}\left(T^{2/3}\right),
\end{aligned}$$

where the first inequality holds by the application of Hoeffding inequality for which with probability at least $1-\delta$ for all $(p,q) \in \mathcal{B}$

$$\begin{aligned}
|\mathsf{GFT}_1(p,q) - \widehat{\mathsf{GFT}}_1(p,q)| &= \mathbb{E}[\mathsf{GFT}_1(p,q) - \widehat{\mathsf{GFT}}_1(p,q)] \\
&\quad + \left|(\mathsf{GFT}_1(p,q) - \widehat{\mathsf{GFT}}_1(p,q)) - \mathbb{E}[\mathsf{GFT}_1(p,q) - \widehat{\mathsf{GFT}}_1(p,q)]\right| \\
&\leq \mathbb{E}\left[\frac{1}{T_0}\sum_{t=\tau+1}^{T_0} \mathbb{I}\left(s_t \leq U_t \leq p\right)\mathbb{I}(\bar{b}_t \geq q) - \mathbb{P}(s_t \leq U_t \leq p, \bar{b}_t \geq q)\right] \\
&\quad + 4\sqrt{\frac{\ln(\frac{2|\mathcal{B}|}{\delta})}{2T_0}}
\end{aligned}$$

$$= 4\sqrt{\frac{\ln(\frac{2|\mathcal{B}|}{\delta})}{2T_0}},$$

and therefore

$$\mathsf{GFT}_1(p,q) \in \left[\underline{\mathsf{GFT}}_1, \overline{\mathsf{GFT}}_1\right] = \left[\widehat{\mathsf{GFT}}_1 - 4\sqrt{\frac{\ln(\frac{2|\mathcal{B}|}{\delta})}{2T_0}}, \widehat{\mathsf{GFT}}_1 + 4\sqrt{\frac{\ln(\frac{2|\mathcal{B}|}{\delta})}{2T_0}}\right].$$

We prove now the result relative to the revenue of the algorithm. First observe that applying Hoeffding inequality with probability at least $1-\delta$

$$\sum_{t=\tau+T_0+1}^{T} \mathsf{Rev}_t(p_t,q_t) \geq \sum_{t=\tau+T_0+1}^{T} \mathsf{Rev}(p_t,q_t) - 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2}(T-T_0-\tau)}.$$

With probability at least $1-3\delta$ using Theorem 2 and Lemma 3

$$\begin{aligned}
\sum_{t=1}^{T} \mathsf{Rev}_t(p_t,q_t) &= \sum_{t=1}^{T_0} \mathsf{Rev}_t(p_t,q_t) + \sum_{t=T_0+1}^{\tau} \mathsf{Rev}_t(p_t,q_t) + \sum_{t=\tau+1}^{\tau+T_0} \mathsf{Rev}_t(p_t,q_t) + \sum_{t=\tau+T_0+1}^{T} \mathsf{Rev}_t(p_t,q_t) \\
&\geq -2T_0 + \beta + \sum_{t=\tau+T_0+1}^{T} \mathsf{Rev}_t(p_t,q_t) \\
&\geq -2T_0 + \beta + \sum_{t=\tau+T_0+1}^{T} \mathsf{Rev}(p_t,q_t) - 2\sqrt{\frac{\ln(\frac{2}{\delta})}{2}(T-T_0-\tau)} \\
&= -2T_0 + \beta - \tilde{\mathcal{O}}\left(\sqrt{T}\right) - (1+\bar{\zeta}) \sum_{t=\tau+T_0+1}^{T} \frac{1}{1+\bar{\zeta}}(-\mathsf{Rev}(p_t,q_t) - \bar{\zeta}) + (T-\tau-T_0)\bar{\zeta} \\
&\geq -2T_0 + \beta - \tilde{\mathcal{O}}\left(\sqrt{T}\right) - 2\sum_{t=\tau+T_0+1}^{T} c(p_t,q_t) + \tilde{\mathcal{O}}\left(T\bar{\zeta}\right) \\
&\geq \beta - \tilde{\mathcal{O}}\left(T_0 + T\bar{\zeta} + \sqrt{KT}\log(K/\delta)\right) \\
&\geq 0,
\end{aligned}$$

where the last holds for some choice of $\beta = \tilde{\mathcal{O}}\left(T^{2/3}\right)$.

□

H Open Problem: Assumption-Free Regret Minimization With GBB Constraints

In the main paper, we design a GBB algorithm under some assumptions on the underling valuation distribution. However, it remains an open question whether a sublinear regret bound can be achieve without these assumptions.

We provide some evidence that our approach is doomed to fail, and designing such algorithm requires novel non-trivial techniques. Specifically, the following example shows that it is impossible to learn an approximately optimal couple of *fixed* prices (p,q). Since this is the aim of our algorithm (see Section 6.2), it is clear that our algorithm cannot be easily extended to settings without assumptions. Notice that this does not rule out the existence of an algorithm achieving sublinear regret. However, such an algorithms would be much more complex as it should dynamically adjust q_t based on the buyers valuations $\mathbf{b}_t$.

Example 1. *We construct a family of distributions where learning the optimal price pair (p^*,q^*) is as hard as finding an arbitrary small segment on a line. In particular, consider a family of seller/buyers*

distributions $\mathcal{P}_\epsilon^x$ *parametrized by* $x \in (7/16, 9/16)$ *and* $\epsilon \in (0, 1/16)$ *described as follows:*

$$\mathbb{P}((s_t, \bar{b}_t, \underline{b}_t) = (a, y, y')) = \begin{cases} 1/3 & \text{if } (a, y, y') = (x - \epsilon, 3/4, 0) \\ 1/3 & \text{if } (a, y, y') = (x + \epsilon, 1/4, 0) \\ 1/3 & \text{if } (a, y, y') = (0, 1/4, 0) \\ 0 & \text{otherwise} \end{cases}$$

For valuations $(x - \epsilon, 3/4, 0)$ *the trade happen if* $(p, q) \in [x - \epsilon, 1] \times [0, 3/4]$ *with* $\mathsf{GFT} = 3/4 - x + \epsilon > 3/16$*, for valuations* $(x + \epsilon, 1/4, 0)$ *the trade happens if* $(p, q) \in [x + \epsilon, 1] \times [0, 1/4]$ *with* $\mathsf{GFT} = 1/4 - x - \epsilon < -3/16$*, and for valuations* $(0, 1/4, 0)$ *the trade happens if* $(p, q) \in [0, 1] \times [0, 1/4]$ *with* $\mathsf{GFT} = 1/4$*. Thus, every couple of prices* $(p, q) \notin [x - \epsilon, x + \epsilon) \times [0, 1/4]$ *leads to an expected sub optimality gap of at least* $1/16$*. Hence, finding an approximately optimal mechanisms requires identify the arbitrarily small interval* $[x - \epsilon, x + \epsilon)$ *with one-bit feedback,* i.e.*, if the seller accept the trade. This is equivalent to a binary search problem, which requires* $\Omega\left(\log(\frac{1}{\epsilon})\right)$ *samples. Such problem requires an arbitrary large number of samples if* ϵ *is arbitrary small.*

We leave open the question of whether this problem remains hard for the broader class of mechanisms available to our regret minimizer or if more sophisticated mechanisms can circumvent this impossibility result.