An Online Learning Theory of Brokerage

Nataša Bolić
University of Ottawa
Ottawa, Canada
nboli039@uottawa.ca

Tommaso Cesari
University of Ottawa
Ottawa, Canada
tcesari@uottawa.ca

Roberto Colomboni
University of Milan & IIT
Milan & Genoa, Italy
roberto.colomboni@unimi.it

ABSTRACT

We investigate brokerage between traders from an online learning perspective. At any round t, two traders arrive with their private valuations, and the broker proposes a trading price. Unlike other bilateral trade problems already studied in the online learning literature, we focus on the case where there are no designated buyer and seller roles: each trader will attempt to either buy or sell depending on the current price of the good.

We assume the agents' valuations are drawn i.i.d. from a fixed but unknown distribution. If the distribution admits a density bounded by some constant M, then, for any time horizon T:

- If the agents' valuations are revealed after each interaction, we provide an algorithm achieving regret $M \log T$ and show this rate is optimal, up to constant factors.
- If only their willingness to sell or buy at the proposed price is revealed after each interaction, we provide an algorithm achieving regret $\sqrt{MT}$ and show this rate is optimal, up to constant factors.

Finally, if we drop the bounded density assumption, we show that the optimal rate degrades to $\sqrt{T}$ in the first case, and the problem becomes unlearnable in the second.

KEYWORDS

Regret minimization; Online learning; Two-sided markets

ACM Reference Format:
Nataša Bolić, Tommaso Cesari, and Roberto Colomboni. 2024. An Online Learning Theory of Brokerage. In *Proc.of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 9 pages.

An extended version of this paper, including the supplementary material, can be found in [6].

1 INTRODUCTION

Over-the-counter (OTC) markets offer a variety of decentralized alternatives to traditional financial exchanges and have gained prominence for their flexibility, diversity, and accessibility for participants. Paraphrasing the words of Tolstoy, "*All centralized markets are the same, but each OTC market is unique in its own way*" [21, 24]. In recent years, OTC markets have flourished, becoming an indispensable part of the global financial ecosystem, with a steady growth trend documented since 2016 [19] and the value of US assets traded in OTC markets surpassing a staggering 50,000 billion USD (exceeding centralized markets by over 20,000 billion USD) in 2020 [24]. Central to the functioning of decentralized OTC markets are brokers who, acting as intermediaries, bridge the gap between buyers and sellers, ensuring that trades are executed smoothly. Beyond mere intermediation, brokers play a significant role in price discovery, gauging demand and supply to determine optimal asset prices. However, the classical impossibility result of Myerson and Satterthwaite [22] highlights that the role of the broker is not without challenges. Inspired by a recent stream of literature [2, 8, 10, 11], we approach the bilateral trade problem of brokerage between traders through the lens of online learning. When viewed from a regret minimization perspective, bilateral trade has been explored over rounds of seller/buyer interactions with no prior knowledge of their private valuations, but only under rigid buyer and seller roles. In contrast, it's important to note that in many key OTC markets, traders are willing to either buy or sell, depending on the prevailing market conditions [23]. These markets encompass a wide array of asset trades, including stocks, derivatives, art, collectibles, precious metals and minerals, energy commodities like gas and oil, as well as digital currencies (cryptocurrencies), among others.

Motivated by brokerage between traders in these markets, we will investigate scenarios where traders' roles as buyers or sellers are not strictly defined.

1.1 Setting

We study the following problem. At each time $t \in \mathbb{N}$,

(1) Two traders arrive with private valuations V_{2t-1} and V_{2t}.
(2) The broker proposes a trading price P_t.
(3) If the price P_t falls between the lowest[1] $V_{2t-1} \wedge V_{2t}$ and highest $V_{2t-1} \vee V_{2t}$ valuations (i.e., if the trader with the smallest valuation is eager to sell at price P_t and the other is willing to buy at P_t), the trader with the highest valuation buys the item from the trader with the lowest valuation paying the brokerage price P_t.
(4) Some feedback is revealed.

Consistently with the existing bilateral trade literature, we assume valuations and prices belong to $[0, 1]$, and the reward associated with each interaction is the sum of the utilities of the traders, known as *gain from trade*. Formally, for any $p, v_1, v_2 \in [0, 1]$, the gain from trade of a price p when the valuations of the traders are v_1 and v_2 is

$$\mathrm{g}(p, v_1, v_2) := (v_1 \vee v_2 - v_1 \wedge v_2)\, \mathbb{I}\{v_1 \wedge v_2 \le p \le v_1 \vee v_2\} \ .$$

The aim of the learner is to minimize the *regret*, defined, for any time horizon $T \in \mathbb{N}$, as

$$R_T := \sup_{p \in [0,1]} \mathbb{E}\left[\sum_{t=1}^{T} \mathrm{GFT}_t(p)\right] - \mathbb{E}\left[\sum_{t=1}^{T} \mathrm{GFT}_t(P_t)\right] ,$$

[1] We denote the minimum (resp., maximum) of any two real numbers $x, y \in \mathbb{R}$ by $x \wedge y$ (resp., $x \vee y$).

Proc.of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

where we let $\mathrm{GFT}_t(q) := \mathrm{g}(q, V_{2t-1}, V_t)$ for all $q \in [0,1]$ and the expectations are taken with respect to the (possible) randomness of $(V_t)_{t\in\mathbb{N}}$ and $(P_t)_{t\in\mathbb{N}}$.

We study this problem under the assumption that the traders' valuations $V, V_1, V_2, \dots$ are generated i.i.d. from an unknown distribution ν, a natural assumption for large and stable markets.

Finally, we consider two different types of feedback:

- *Full feedback.* The valuations V_{2t-1} and V_{2t} of the two current traders are revealed at the end of every round t.
- *Two-bit feedback.* Only the indicator functions $\mathbb{I}\{P_t \le V_{2t-1}\}$ and $\mathbb{I}\{P_t \le V_{2t}\}$ are revealed at the end of every round t.

The information collected by the full feedback model corresponds to *direct revelation mechanisms*, where the traders communicate their valuations V_{2t-1} and V_{2t} before each round, and the price proposed by the mechanism at time t only depends on past bids $V_1, \dots, V_{2t-2}$. The two-bit feedback model corresponds to *posted price* mechanisms, where traders only communicate their willingness to buy or sell at the posted price,[2] and the valuations V_{2t-1} and V_{2t} are *never* revealed.

1.2 Overview of Our Contributions

If the distribution ν of the traders' valuations admits a density bounded by some constant M, then, for any time horizon T:

- In the full feedback case, we design an algorithm (Algorithm 1) achieving regret $O(M \log T)$ (Theorem 3.1) and provide a matching lower bound $\Omega(M \log T)$ (Theorem 3.2).
- In the two-bit feedback case, we design an algorithm (Algorithm 2) achieving regret $O(\sqrt{MT})$ (Theorem 4.2) and provide a matching lower bound $\Omega(\sqrt{MT})$ (Theorem 4.3).

We stress that ours is the first paper on online learning in bilateral trade where lower bounds have the optimal dependency on M.

If we drop the bounded density assumption, we show that the optimal rate degrades to $\Theta(\sqrt{T})$ in the full feedback case (Theorem 5.1 and 5.3), while the problem becomes unlearnable in the two-bit feedback case (Theorem 5.4).

1.3 Techniques and Challenges

The two feedback models we consider present different challenges.

Full Feedback. It can be proven that assuming ν has a bounded density implies the Lipschitzness of the expected gain from trade. The standard approach for Lipschitz objectives is to discretize the domain uniformly and run an optimal expert algorithm on the discretization, which yields straightforwardly an $\frac{M}{K}T + \sqrt{T \log K}$ regret guarantee, where K is the number of prices in the discretization and T is the time horizon. Alternatively, one could build a reduction to known bilateral trade problems to obtain an improved $\sqrt{T}$ bound (for further details, see the Related Work section). Both these natural choices are highly suboptimal. In contrast, we exploit the specific structure of our problem to achieve an exponential gain, resulting in an $M \log T$ regret bound. Regarding the lower bound, it's worth noting that determining the optimal dependence on M in such problems is remarkably challenging. In fact, no previous papers on online learning for bilateral trade problems showcased lower bounds displaying *any* dependence on M, let alone the optimal one. We manage to achieve this through two pivotal lemmas: an Approximation (Lemma 2.1) and a Representation (Lemma 2.2) lemma. These two technical results lead us to Theorem 2.3, which establishes two points: first, $\mathbb{E}[V]$ is always a maximizer of the expected gain from trade; second, posting a price close to the maximizer has a cost only quadratic in the distance. These two facts point to a peculiar similarity between ours and the statistical problem of estimating an expectation online with a quadratic loss, an observation that turns out to be crucial in a setting where we could not otherwise recycle any of the existing techniques of other online learning settings in bilateral trade (none of which features logarithmic rates). Still, following this path is made technically challenging by the fact that building a hard instance (in a setting where we cannot control the gain from trade directly but only indirectly through the distributions of traders' valuations) is far from being straightforward. We manage to circumvent this roadblock by carefully designing a 1-parameter family of hard distributions which are used to build a reduction to a Bayesian problem. By means of non-trivial information-theoretic and probabilistic arguments, we can exploit the quadratic loss intuition, to obtain a lower bound featuring the optimal $M \log T$ rate.

Beyond this, the Approximation and Representation Lemmas also suggest the best pathway to tackle the non-bounded-density setting. The idea of trying to approximate the representation given by the second lemma and the fact that this representation reduces to a simpler form in the bounded-density case allows us to design a simple algorithm that enjoys optimal regret guarantees both in the bounded-density ($M \log T$ regret) and the non-bounded density ($\sqrt{T}$ regret) cases while being completely oblivious to which of the two assumptions hold.

Two-Bit Feedback. A clear challenge of the two-bit feedback is posed by its low quality: at each time step, this feedback is not even sufficient to reconstruct the so-called bandit feedback (i.e., the reward $\mathrm{GFT}_t(P_t)$ of the posted price P_t), which is typically considered the bare minimum in online learning. Once more, the Approximation and Representation Lemmas point to the strategy of estimating $\mathbb{E}[V]$ with the available feedback. Leveraging a discretization argument (Lemma 4.1) and concentration inequalities, we obtain an upper bound of $\sqrt{MT}$ in the bounded-density case. To show that this rate is optimal, we build on the hard instances we designed for the full feedback case. The difference now is that the scarcity of feedback leads to a setting similar to the so-called "revealing action" problem [12]. A notable difference is that the regular revealing action problem has an optimal $T^{2/3}$ regret, but, cast in our problem, due to the shape of the reward around the maximizer, a careful analysis shows that the regret of the adapted revealing action is $\sqrt{MT}$. Again, we stress that this is the first lower bound in online learning with two-bit feedback for bilateral trade with any dependence on M, let alone the optimal one.

[2] For the nitpicker, the natural feedback model in this case would be to reveal the four bits $\mathbb{I}\{P_t < V_{2t-1}\}, \mathbb{I}\{P_t = V_{2t-1}\}, \mathbb{I}\{P_t < V_{2t}\}$, and $\mathbb{I}\{P_t = V_{2t}\}$. This feedback is more informative than the one we propose; hence our upper bounds hold *a fortiori* in the four-bit feedback model. For the lower bounds, in the bounded density case, given that the two and four-bit feedback give the same information with probability 1, the same results hold; in the general case, a straightforward adaptation of the exact same construction in our lower bounds gives the result for the four-bit feedback case. Given this equivalency between these two models, we opt for the two-bit feedback for the sake of conciseness.

As for the full feedback case, we show how the problem changes without the bounded density assumption. Dropping the assumption leads to a pathological phenomenon typical of bilateral trade problems (known as needle-in-a-haystack) leading us to the design of hard instances in which all-but-one prices suffer a high $\Omega(1)$-regret, and where it is essentially impossible to find the optimal price among the continuum amount of suboptimal ones given the small amount of information carried by the two-bit feedback.

1.4 Related Work

Since its inception in the seminal work of Myerson and Satterthwaite [22], a large body of literature on bilateral trade has emerged, mainly from a best-approximation/game-theoretic perspective [1, 3, 5, 7, 14–18, 20]. For a discussion about this literature, see [8].

In recent years, bilateral trade has also been studied in online learning settings. Given that these works are the most closely related to ours, we focus on discussing our relationship with them.

In [10], the authors studied a bilateral trade setting where the sequence of seller/buyer valuations $(S_t, B_t)_{t\in\mathbb{N}}$ is an i.i.d. process. This is also the case in our setting, via the reduction where we set $S_t \coloneqq V_{2t-1} \wedge V_{2t}$ and $B_t \coloneqq V_{2t-1} \vee V_{2t}$. In the full-feedback case, they obtain regret $\widetilde{O}(\sqrt{T})$ (improved to $O(\sqrt{T})$ in [8]) and show that any algorithm suffers worst-case[3] regret $\Omega(\sqrt{T})$, even under the bounded density assumption (shorthanded by BDA, in the rest of this section), i.e., when the joint distribution of seller/buyer valuations has a density bounded by some constant M. In contrast, in the full feedback case, we prove that this rate is suboptimal in our setting under BDA, and can be improved exponentially to an optimal $\Theta(M \log T)$ (Theorem 3.1 and 3.2). Without BDA, one could run their FBP algorithm in our setting, keeping its $O(\sqrt{T})$ guarantees. Instead, we elected to propose a different algorithm (Algorithm 3) with a simpler analysis, achieving a regret with the same dependence on T but with better numerical constants. We also remark that a new lower bound proof is required to prove this rate is optimal (Theorem 5.3), given that the family of instances to prove the lower bound in [10] cannot arise via the reduction above.

In the two-bit feedback i.i.d. setting, [10] show that any algorithm has to suffer linear worst-case[3] regret, even under BDA. In contrast, we prove that in the two-bit feedback case, our problem is learnable under BDA and obtain optimal $\Theta(\sqrt{MT})$ guarantees (Theorem 4.2 and 4.3). Without BDA, we show that our problem is also unlearnable (Theorem 5.4).

Under BDA and the additional assumption that S_t and B_t are independent of each other (which in not the case in our problem because of the correlation between the maximum and the minimum of the traders' valuations), [10] achieve a regret $\widetilde{O}(M^{1/3}T^{2/3})$ (improved to $O(M^{1/3}T^{2/3})$ in [8]) and show that any algorithm suffers worst-case regret $\Omega(T^{2/3})$ when M is sufficiently large.

The bilateral trade problem has also been studied under a *weak budget balance* assumption in [8]. In this setting, the learner can post two different prices: a selling price p (to the seller) and a buying price q (to the buyer), with $p \le q$. The authors prove that this bilateral trade problem is learnable even without assuming the independence of seller's S_t and buyer's B_t valuations, providing an $\widetilde{O}(MT^{3/4})$ upper bound. [9, 11] prove that this rate is optimal in T (up to logarithmic terms), providing a matching (in T) $\Omega(T^{3/4})$ lower bound. We remark that even if the broker were permitted to offer two distinct prices, $p \le q$ (with p as the selling price and q as the buying price), our results would still be optimal. This is because there's no reason to do so in the full feedback scenario[4], and our two-bit case lower bound remains valid with slight adjustments.

In the adversarial case, [10] show that learning is impossible. Their lower bound construction yields that learning in the adversarial case is also impossible in our setting. In fact, given that in the adversarial case the sequence of traders' valuations can be chosen arbitrarily, we can set $V_{2t-1} \coloneqq s_t$ and $V_{2t} \coloneqq b_t$ where s_t and b_t are defined as in the adversarial lower bound construction in the proof of Theorem 5.1 in [10], and the same proof applies verbatim.

To achieve learnability in the adversarial case, [2] weakened the notion of regret, setting as a benchmark a multiple of (and not exactly the) the cumulative reward of the best fixed price in hindsight. In the full-feedback case they obtained $\widetilde{\Theta}(\sqrt{T})$ guarantees on the 2-regret, while in the two-bit-feedback case they obtained a mismatching upper $\widetilde{O}(T^{3/4})$ and lower $\Omega(T^{2/3})$ bounds on the 2-regret (and closing this gap is still an open problem in the bilateral trade literature). Analogous considerations such as the ones made above for the standard adversarial setting apply also to this weakened notion of regret in our setting.

As a final note on the current online learning literature on bilateral trade, we remind the reader again that ours is the first paper in this literature that has the optimal dependency on both M and T.

2 STRUCTURAL RESULTS

We denote the Dirac measure based at $x \in \mathbb{R}$ by δ_x, i.e., δ_x is the measure defined via the equation $\delta_x[A] = \mathbb{I}\{x \in A\}$ for any set A. For any (signed) measure μ and any measurable set E, we will write μE rather than $\mu[E]$ whenever this does not cause confusion. For any measure μ over $[0,1]$, we let $\bar{\mu} \coloneqq \int_{[0,1]} x \,\mathrm{d}\mu(x)$, and we define the functions $\widetilde{\rho}(\mu)$ and $\rho(\mu)$, for all $p \in [0,1]$, by

$$\widetilde{\rho}(\mu)(p) \coloneqq \int_0^p \big(\mu[0,\lambda] + \mu[0,\lambda)\big)\,\mathrm{d}\lambda + \big(\mu[0,p] + \mu[0,p)\big)(\bar{\mu} - p)\,,$$

$$\rho(\mu)(p) \coloneqq \widetilde{\rho}(\mu)(p) + \mu\{p\}\left(\int_0^p \mu[0,\lambda]\,\mathrm{d}\lambda + \int_p^1 \mu[\lambda,1]\,\mathrm{d}\lambda\right).$$

The following lemma shows that $\bar{\mu}$ maximizes $\widetilde{\rho}(\mu)$ (in general) and $\rho(\mu)$ (if μ has a bounded density), and the cost of approximating $\bar{\mu}$.

Lemma 2.1 (Approximation). *If μ is a probability measure on $[0,1]$, then $\widetilde{\rho}(\mu)(\bar{\mu}) = \max_{p\in[0,1]} \widetilde{\rho}(\mu)(p)$ and, for any $p \in [0,1]$, $\widetilde{\rho}(\mu)(\bar{\mu}) - \widetilde{\rho}(\mu)(p) \le 2|\bar{\mu} - p|$. If μ has a density bounded by $M > 0$, then $\rho(\mu) = \widetilde{\rho}(\mu)$ and*

$$0 \le \rho(\mu)(\bar{\mu}) - \rho(\mu)(p) \le M\,|\bar{\mu} - p|^2\,, \qquad \forall p \in [0,1]\,.$$

Proof. For $\lambda \in [0,1]$, let $m(\lambda) \coloneqq \mu[0,\lambda] + \mu[0,\lambda)$ and note that m is a $[0,1]$-valued non-decreasing function of λ. For any $p \in [0,1]$,

$$\widetilde{\rho}(\mu)(\bar{\mu}) - \widetilde{\rho}(\mu)(p) = \int_p^{\bar{\mu}} \big(m(\lambda) - m(p)\big)\,\mathrm{d}\lambda \quad \begin{cases} \ge 0\,, \\ \le 2|\bar{\mu} - p|\,, \end{cases} \tag{1}$$

[3] Among all i.i.d. sequences $(S_t, B_t)_{t\in\mathbb{N}}$, not just those arising from the reduction $S_t \coloneqq V_{2t-1} \wedge V_{2t}$ and $B_t \coloneqq V_{2t-1} \vee V_{2t}$.

[4] As [11] remark, "The only reason for a budget-balanced algorithm to post two different prices is to obtain more information. A direct verification shows that the expected gain from trade can always be maximized by posting the same price to both the seller and the buyer"

which implies that $\widetilde{\rho}(\mu)(\bar{\mu}) = \max_{p\in[0,1]} \widetilde{\rho}(\mu)(p)$. Next, note that for all $p \in [0,1]$, $\left|\widetilde{\rho}(\mu)(p) - \rho(\mu)(p)\right| \le \mu\{p\}$, which, if μ has a density f bounded by a constant M, implies $\widetilde{\rho}(\mu)(p) = \rho(\mu)(p)$ and

$$\rho(\mu)(\bar{\mu}) - \rho(\mu)(p) = \widetilde{\rho}(\mu)(\bar{\mu}) - \widetilde{\rho}(\mu)(p) \overset{(1)}{=} \int_p^{\bar{\mu}} \big(m(\lambda) - m(p)\big)\,\mathrm{d}\lambda$$

$$= 2\int_p^{\bar{\mu}} \int_p^{\lambda} f(x)\,\mathrm{d}x\,\mathrm{d}\lambda \le 2M \left|\int_p^{\bar{\mu}} |\lambda - p|\,\mathrm{d}\lambda\right| = M\,|\bar{\mu} - p|^2 \ . \quad \square$$

The next lemma provides a crucial representation of the objective $p \mapsto \mathbb{E}\big[\mathrm{GFT}_t(p)\big]$. Its long and (somewhat) tedious proof is deferred to the supplementary material.

Lemma 2.2 (Representation). *For any $t \in \mathbb{N}$ and $p \in [0,1]$,*

$$\mathbb{E}\big[\mathrm{GFT}_t(p)\big] = \rho(\nu)(p) \ .$$

The following is an immediate corollary of Lemmas 2.1 and 2.2.

Theorem 2.3. *If ν admits a density bounded by some constant M, then for any $t \in \mathbb{N}$ and any $p \in [0,1]$, it holds that*

$$0 \le \mathbb{E}\big[\mathrm{GFT}_t\big(\mathbb{E}[V]\big)\big] - \mathbb{E}\big[\mathrm{GFT}_t(p)\big] \le M \cdot \big|\mathbb{E}[V] - p\big|^2 \ ,$$

and, in particular, $\max_{p\in[0,1]} \mathbb{E}\big[\mathrm{GFT}_t(p)\big] = \mathbb{E}\big[\mathrm{GFT}_t\big(\mathbb{E}[V]\big)\big]$.

The previous theorem gives much intuition on the problem under the bounded density assumption. It proves that the optimal action would be to post the (unknown) expected value $\mathbb{E}[V]$ of the valuations. Moreover, it suggests the strategy of approximating this value on the basis of the observed feedback, since posting a price close to expectation has only a quadratic cost in the approximation.

3 FULL FEEDBACK

We begin by studying the full feedback case (corresponding to direct revelation mechanisms) under the bounded density assumption.

3.1 Upper Bound

Following the intuition provided by Theorem 2.3, we introduce the Follow-the-Mean algorithm (FTM), which simply posts the empirical average of the past valuations (Algorithm 1).

Algorithm 1: Follow the Mean (FTM)

Post $P_1 := 1/2$, then receive feedback V_1, V_2;
for *time* $t = 2, 3, \ldots$ **do**
  Post $P_t := \frac{\sum_{s=1}^{2(t-1)} V_s}{2(t-1)}$, then receive feedback V_{2t-1}, V_{2t};

The next theorem shows that FTM enjoys an $M \log T$ regret.

Theorem 3.1. *If ν has density bounded by some constant $M > 0$, then the regret of FTM satisfies, for all time horizons $T \ge 2$*

$$R_T \le \frac{1}{2} + \frac{M}{4}\big(1 + \ln(T-1)\big) \ .$$

Proof. For all time horizons $T \ge 2$, we have

$$R_T - \frac{1}{2} \le \sum_{t=2}^{T} \Big(\mathbb{E}\big[\mathrm{GFT}_t\big(\mathbb{E}[V]\big)\big] - \mathbb{E}\big[\mathrm{GFT}_t(P_t)\big]\Big)$$

$$\overset{(l)}{=} \sum_{t=2}^{T} \mathbb{E}\left[\Big[\mathbb{E}\big[\mathrm{GFT}_t\big(\mathbb{E}[V]\big) - \mathrm{GFT}_t(p)\big]\Big]_{p=P_t}\right]$$

$$\overset{(t)}{\le} \sum_{t=2}^{T} \mathbb{E}\left[\Big[M\big|p - \mathbb{E}[V]\big|^2\Big]_{p=P_t}\right] = M\sum_{t=2}^{T} \mathbb{E}\Big[\big|P_t - \mathbb{E}[V]\big|^2\Big]$$

$$\overset{(f)}{=} M\sum_{t=2}^{T} \int_0^{\infty} \mathbb{P}\Big[\big|P_t - \mathbb{E}[V]\big|^2 \ge \varepsilon\Big]\,\mathrm{d}\varepsilon \overset{(h)}{\le} M\sum_{t=1}^{T-1} \int_0^{\infty} 2e^{-8t\varepsilon}\,\mathrm{d}\varepsilon$$

$$= \frac{M}{4}\sum_{t=1}^{T-1} \frac{1}{t} \le \frac{M}{4}\left(1 + \int_1^{T-1} \frac{1}{s}\,\mathrm{d}s\right) \le \frac{M}{4}\big(1 + \ln(T-1)\big) \ ,$$

where (l) follows from the Freezing Lemma (see, e.g., [13, Lemma 8]) after observing that $\mathrm{GFT}_t(P_t) = \mathrm{g}(P_t, V_{2t-1}, V_{2t})$ and P_t is independent of (V_{2t-1}, V_{2t}); (t) from Theorem 2.3; (f) follows from Fubini's Theorem; and (h) from Hoeffding's inequality. $\square$

3.2 Lower Bound

In this section, we prove the optimality of FTM by showing a matching $M \log T$ lower bound. This is the most technically challenging result of the paper. For a high-level overview of its proof, we refer the reader back to Section 1.3.

Theorem 3.2. *There exist two numerical constants $c_1, c_2 > 0$ such that, for any $M \ge 2$ and any time horizon $T \ge c_2 M^4$, the worst-case regret of any algorithm satisfies*

$$\sup_{\nu\in\mathcal{D}_M} R_T^{\nu} \ge c_1 M \log T \ ,$$

where R_T^{ν} is the regret at time T of the algorithm when the underlying i.i.d. sequence of traders' valuations follows the distribution ν, and $\mathcal{D}_M$ is the set of all distributions ν with density bounded by M.

Proof. Given that we are in a stochastic i.i.d. setting, we can restrict this proof to deterministic algorithms without loss of generality. Let $M \ge 2$, $J_M := \left[\frac{1}{2} - \frac{1}{14M}, \frac{1}{2} + \frac{1}{14M}\right]$, $f := \mathbb{I}_{[0,\frac{3}{7}]} + M\mathbb{I}_{J_M} + \mathbb{I}_{[\frac{4}{7},1]}$, and, for any $\varepsilon \in [-1,1]$, $g_\varepsilon := -\varepsilon\mathbb{I}_{[\frac{1}{7},\frac{3}{14}]} + \varepsilon\mathbb{I}_{(\frac{3}{14},\frac{2}{7}]}$ and $f_\varepsilon := f + g_\varepsilon$ (see Fig. 1, left). For any $\varepsilon \in [-1,1]$, note that $0 \le f_\varepsilon \le M$ and $\int_0^1 f_\varepsilon(x)\,\mathrm{d}x = 1$, hence f_ε is a valid density on $[0,1]$ bounded by M, and we will denote the corresponding probability measure by ν_ε. Consider for each $q \in [0,1]$, an i.i.d. sequence $(B_{q,t})_{t\in\mathbb{N}}$ of Bernoulli random variables of parameter q, an i.i.d. sequence $(\tilde{B}_t)_{t\in\mathbb{N}}$ of Bernoulli random variables of parameter $1/7$, an i.i.d. sequence $(U_t)_{t\in\mathbb{N}}$ of uniform random variables on $[0,1]$, a uniform random variable E on $[-\varepsilon_M, \varepsilon_M]$ where $\varepsilon_M := \frac{7}{M}$, such that $\Big((B_{q,t})_{t\in\mathbb{N},q\in[0,1]}, (\tilde{B}_t)_{t\in\mathbb{N}}, (U_t)_{t\in\mathbb{N}}, E\Big)$ is an independent family. Let $\varphi\colon [0,1] \to [0,1]$ be such that, if U is a uniform random variable on $[0,1]$, then the distribution of $\varphi(U)$ has density $\frac{7}{6} \cdot f \cdot \mathbb{I}_{[0,1]\setminus[1/7,2/7]}$ (which exists by the Skorokhod representation theorem [25, Section 17.3]). For each $\varepsilon \in [-1,1]$ and $t \in \mathbb{N}$, define

$$V_{\varepsilon,t} := \left(\frac{2+U_t}{14}(1 - B_{\frac{1+\varepsilon}{2},t}) + \frac{3+U_t}{14}B_{\frac{1+\varepsilon}{2},t}\right)\tilde{B}_t + \varphi(U_t)(1-\tilde{B}_t). \quad (2)$$

Straightforward computations show that, for each $\varepsilon \in [-1,1]$ the sequence $(V_{\varepsilon,t})_{t\in\mathbb{N}}$ is i.i.d. with common distribution given by ν_ε, and this sequence is independent of E. For any $\varepsilon \in [-1,1]$, $p \in [0,1]$, and $t \in \mathbb{N}$, let $\mathrm{GFT}_{\varepsilon,t}(p) := \mathrm{g}(p, V_{\varepsilon,2t-1}, V_{\varepsilon,2t})$ (for a qualitative

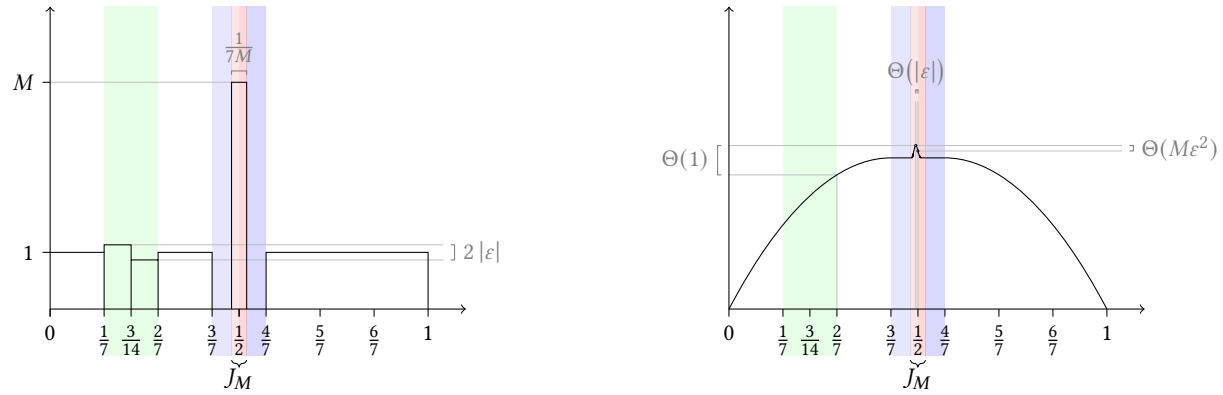

Figure 1: On the left, the density f_ε of a "hard" instance used to prove the lower bounds in Theorems 3.2 and 4.3. A base uniform distribution is warped in the intervals $[1/7, 2/7]$ (green) and $[3/7, 4/7]$ (blue+red). The density on $[1/7, 2/7]$ is split into two uneven parts, differing by ε from the original. The mass on $[3/7, 4/7]$ is concentrated in a small set J_M of size $\Theta(1/M)$ around $1/2$. The corresponding gain from trade, on the right, has a smooth spike of height $\Theta(M|\varepsilon|^2)$ situated in J_M, at a distance $\Theta(|\varepsilon|)$ from $1/2$. When $\varepsilon < 0$ (resp., $\varepsilon > 0$), the spike is left (resp., right) of $1/2$, and posting $1/2$ is better than posting any price after (resp., before) $1/2$. In the two-bit feedback lower bound, the only way to gather usable feedback is to post prices in $[1/7, 2/7]$, which give rewards $\Theta(1)$-away from the optimal one.

representation of its expectation, see Fig. 1, right). For any $\varepsilon \in [-1, 1]$ and $t \in \mathbb{N}$, a direct computation shows that $\bar{v}_\varepsilon = \mathbb{E}[V_{\varepsilon,t}] = \frac{1}{2} + \frac{\varepsilon}{196}$. By Lemmas 2.1 and 2.2, we have, for all $\varepsilon \in [-1, 1]$, $t \in \mathbb{N}$, and $p \in [0, 1]$,

$$\mathbb{E}\big[\mathrm{GFT}_{\varepsilon,t}(p)\big] = 2\int_0^p \int_0^\lambda f_\varepsilon(s)\,\mathrm{d}s\,\mathrm{d}\lambda + 2(\bar{v}_\varepsilon - p)\int_0^p f_\varepsilon(s)\,\mathrm{d}s\,,$$

which, together with the fundamental theorem of calculus —[4, Theorem 14.16], noting that $p \mapsto \mathbb{E}\big[\mathrm{GFT}_{\varepsilon,t}(p)\big]$ is absolutely continuous with derivative defined a.e. by $p \mapsto 2(\bar{v}_\varepsilon - p)f_\varepsilon(p)$— yields, for any $p \in J_M$,

$$\mathbb{E}\big[\mathrm{GFT}_{\varepsilon,t}(\bar{v}_\varepsilon)\big] - \mathbb{E}\big[\mathrm{GFT}_{\varepsilon,t}(p)\big] = M|\bar{v}_\varepsilon - p|^2\,. \tag{3}$$

Note also that for all $\varepsilon \in [-\varepsilon_M, \varepsilon_M]$, $t \in \mathbb{N}$, and $p \in [0, 1] \setminus J_M$,

$$\mathbb{E}\big[\mathrm{GFT}_{\varepsilon,t}(p)\big] \le \mathbb{E}\left[\mathrm{GFT}_{\varepsilon,t}\left(1/2\right)\right]\,. \tag{4}$$

Fix any arbitrary deterministic algorithm for the full feedback setting $(\tilde{\alpha}_t)_{t\in\mathbb{N}}$, i.e., a sequence of functions $\tilde{\alpha}_t : \big([0,1]\times[0,1]\big)^{t-1} \to [0,1]$ mapping past feedback into prices (with the convention that $\tilde{\alpha}_1$ is just a number in $[0,1]$). For each $t \in \mathbb{N}$, define $\alpha_t : \big([0,1]\times[0,1]\big)^{t-1} \to J_M$ equal to $\tilde{\alpha}_t$ whenever $\tilde{\alpha}_t$ takes values in J_M, and equal to $1/2$ otherwise. Defining $Z \coloneqq \frac{1+E}{2}$, and R_T^ν as the regret of the algorithm $(\tilde{\alpha}_t)_{t\in\mathbb{N}}$ at time T when the underlying sequence of traders' valuations follows the distribution ν, we have that the worst-case regret $\sup_{\nu\in\mathcal{D}_M} R_T^\nu$ is lower bounded by

$$\begin{aligned}
&\sup_{\varepsilon\in[-\varepsilon_M,\varepsilon_M]} \sum_{t=1}^T \mathbb{E}\Big[\mathrm{GFT}_{\varepsilon,t}(\bar{v}_\varepsilon) - \mathrm{GFT}_{\varepsilon,t}\big(\tilde{\alpha}_t(V_{\varepsilon,1},\dots,V_{\varepsilon,2(t-1)})\big)\Big] \\
&\overset{(4)}{\ge} \sup_{\varepsilon\in[-\varepsilon_M,\varepsilon_M]} \sum_{t=1}^T \mathbb{E}\Big[\mathrm{GFT}_{\varepsilon,t}(\bar{v}_\varepsilon) - \mathrm{GFT}_{\varepsilon,t}\big(\alpha_t(V_{\varepsilon,1},\dots,V_{\varepsilon,2(t-1)})\big)\Big] \\
&\overset{\spadesuit}{=} M \sup_{\varepsilon\in[-\varepsilon_M,\varepsilon_M]} \sum_{t=1}^T \mathbb{E}\Big[\big|\bar{v}_\varepsilon - \alpha_t(V_{\varepsilon,1},\dots,V_{\varepsilon,2(t-1)})\big|^2\Big] \\
&\ge M \sum_{t=1}^T \mathbb{E}\Big[\big|\bar{v}_E - \alpha_t(V_{E,1},\dots,V_{E,2(t-1)})\big|^2\Big] \\
&\overset{\heartsuit}{\ge} M \sum_{t=1}^T \mathbb{E}\Big[\big|\bar{v}_E - \mathbb{E}[\bar{v}_E \mid V_{E,1},\dots,V_{E,2(t-1)}]\big|^2\Big] \\
&= \frac{M}{196} \sum_{t=1}^T \mathbb{E}\Big[\big|E - \mathbb{E}[E \mid V_{E,1},\dots,V_{E,2(t-1)}]\big|^2\Big] \\
&\overset{\blacklozenge}{\ge} \frac{M}{196} \sum_{t=1}^T \mathbb{E}\Big[\big|E - \mathbb{E}[E \mid B_{\frac{1+E}{2},1},\dots,B_{\frac{1+E}{2},2(t-1)}]\big|^2\Big] = \\
&= \frac{M}{98} \sum_{t=1}^T \mathbb{E}\Big[\big|Z - \mathbb{E}[Z \mid B_{Z,1},\dots,B_{Z,2(t-1)}]\big|^2\Big]
\end{aligned}$$

where ♠ follows from (3) and the fact that α_t takes values in J_M; ♥ from the fact that the minimizer of the $L^2(\mathbb{P})$-distance from $\bar{v}_E$ in $\sigma(V_{E,1},\dots,V_{E,2(t-1)})$ is $\mathbb{E}[\bar{v}_E \mid V_{E,1},\dots,V_{E,2(t-1)}]$ (see, e.g., [25, Section 9.4]); ♦ follows from the fact that, by Eq. (2) and the independence of E from $\Big((B_{q,t})_{t\in\mathbb{N},q\in[0,1]}, (\tilde{B}_t)_{t\in\mathbb{N}}, (U_t)_{t\in\mathbb{N}}\Big)$, the conditional expectation $\mathbb{E}[E \mid V_{E,1},\dots,V_{E,2(t-1)}]$ is a measurable function of $B_{\frac{1+E}{2},1},\dots,B_{\frac{1+E}{2},2(t-1)}$, together with the same observation made in ♥ about the minimization of $L^2(\mathbb{P})$ distance.

Finally, the general term of this last sum is the expected squared distance between the random parameter (drawn uniformly over $[(1-\varepsilon_M)/2, (1+\varepsilon_M)/2]$) of an i.i.d. sequence of Bernoulli random variables and the conditional expectation of this random parameter given $2(t-1)$ independent realizations of these Bernoullis. A probabilistic argument shows that there exist two universal constants $\tilde{c}, c_2 > 0$ such that, for all $t \ge c_2 M^4$,

$$\mathbb{E}\Big[\big|Z - \mathbb{E}[Z \mid B_{Z,1},\dots,B_{Z,2(t-1)}]\big|^2\Big] \ge \tilde{c}\frac{1}{t-1}\,. \tag{5}$$

At a high level, this is because, in an event of probability $\Omega(1)$, if t is large enough, the conditional expectation $\mathbb{E}[Z \mid B_{Z,1}, \dots, B_{Z,2(t-1)}]$ is very close to the empirical average $\frac{1}{2(t-1)} \sum_{s=1}^{2(t-1)} B_{Z,s}$, whose expected squared distance from Z is $\Omega\big(1/(t-1)\big)$. For a formal proof (5) with explicit constants, see the supplementary material. Summing over t and putting everything together gives the result. □

4 TWO-BIT FEEDBACK

We now study the two-bit feedback case (corresponding to posted price mechanisms) under the bounded density assumption.

4.1 Upper Bound

Motivated once more by the intuition provided by Theorem 2.3, we begin this section by giving a way to approximate the expected value of traders' valuations on the basis of the two-bit feedback and quantify the approximation power of this strategy.

Lemma 4.1. *For any random variable X on $[0,1]$ and any $T_0 \in \mathbb{N}$,*

$$0 \le \mathbb{E}[X] - \frac{1}{T_0} \sum_{t=1}^{T_0} \mathbb{P}\left[\frac{t}{T_0} \le X\right] \le \frac{1}{T_0}$$

Proof. Notice that

$$\begin{aligned}\frac{1}{T_0} \sum_{t=1}^{T_0} \mathbb{P}\left[\frac{t}{T_0} \le X\right] &= \sum_{t=1}^{T_0} \int_{\frac{t-1}{T_0}}^{\frac{t}{T_0}} \mathbb{P}\left[\frac{t}{T_0} \le X\right] \mathrm{d}\lambda \\ &\le \sum_{t=1}^{T_0} \int_{\frac{t-1}{T_0}}^{\frac{t}{T_0}} \mathbb{P}\left[\lambda \le X\right] \mathrm{d}\lambda \\ &\le \sum_{t=1}^{T_0} \int_{\frac{t-1}{T_0}}^{\frac{t}{T_0}} \mathbb{P}\left[\frac{t-1}{T_0} \le X\right] = \frac{1}{T_0} \sum_{t=1}^{T_0} \mathbb{P}\left[\frac{t-1}{T_0} \le X\right] \mathrm{d}\lambda\,.\end{aligned}$$

Since by Fubini's Theorem,

$$\mathbb{E}[X] = \int_0^1 \mathbb{P}[\lambda \le X]\,\mathrm{d}\lambda = \sum_{t=1}^{T_0} \int_{\frac{t-1}{T_0}}^{\frac{t}{T_0}} \mathbb{P}[\lambda \le X]\,\mathrm{d}\lambda\,,$$

we obtain

$$\begin{aligned}0 \le T_0\mathbb{E}[X] - \sum_{t=1}^{T_0} \mathbb{P}\left[\frac{t}{T_0} \le X\right] &\le \sum_{t=1}^{T_0} \left(\mathbb{P}\left[\frac{t-1}{T_0} \le X\right] - \mathbb{P}\left[\frac{t}{T_0} \le X\right]\right) \\ = \sum_{t=1}^{T_0} \mathbb{P}\left[\frac{t-1}{T_0} \le X < \frac{t}{T_0}\right] &= \mathbb{P}[0 \le X < 1] \le 1\,.\end{aligned}$$

□

The previous lemma suggests the design of a simple Explore-then-Commit (ETC) strategy (Algorithm 2), where the learner spends an initial phase of length T_0 trying to estimate $\mathbb{E}[V]$ and then posts this estimate every round up to the time horizon T. ETC algorithms are of great practical importance due to their easy implementability and interpretability. This usually comes at a cost of performance. As the following result (together with Theorem 4.3, in the next section) will show, our ECT algorithm is free of this flaw.

Theorem 4.2. *If ν has density bounded by some constant $M > 0$, then the regret of ETC satisfies, for all time horizons T,*

$$R_T \le T_0 - \frac{1}{2} + M(T - T_0)\left(\frac{2}{T_0^2} + \frac{1}{T_0}\right)$$

Algorithm 2: Explore-then-Commit (ETC)

Input: Exploration time $T_0 \in \mathbb{N}$;
for *time* $t = 1, 2, \dots, T_0$ **do**
  Post $P_t := t/T_0$;
  Receive feedback $\mathbb{I}\{P_t \le V_{2t-1}\}, \mathbb{I}\{P_t \le V_{2t}\}$;
for *time* $t = T_0 + 1, T_0 + 2, \dots$ **do**
  Post $P_t := \frac{1}{2T_0} \sum_{s=1}^{T_0} \big(\mathbb{I}\{P_s \le V_{2s-1}\} + \mathbb{I}\{P_s \le V_{2s}\}\big)$;

Tuning the parameter $T_0 := \lceil \sqrt{MT} \rceil$ yields

$$R_T \le 2.5 + 2\sqrt{MT}\,.$$

Proof. Fix any $T_0 \in \mathbb{N}$ and let $p_0 := \frac{1}{T_0} \sum_{s=1}^{T_0} \mathbb{P}\left[\frac{s}{T_0} \le V\right]$. By Hoeffding's inequality and Fubini's theorem, we get

$$\begin{aligned}\mathbb{E}\left[\left|p_0 - P_{T_0+1}\right|^2\right] &= \int_0^{+\infty} \mathbb{P}\left[\left|p_0 - P_{T_0+1}\right|^2 \ge \varepsilon\right] \mathrm{d}\varepsilon \\ &\le \int_0^{+\infty} 2\exp(-4\varepsilon T_0)\,\mathrm{d}\varepsilon = \frac{1}{2T_0}\,,\end{aligned}$$

from which, leveraging also Lemma 4.1, it follows that

$$\mathbb{E}\left[\left|\mathbb{E}[V] - P_{T_0+1}\right|^2\right] \le 2\left|\mathbb{E}[V] - p_0\right|^2 + 2\mathbb{E}\left[\left|p_0 - P_{T_0+1}\right|^2\right] \le \frac{2}{T_0^2} + \frac{1}{T_0}\,.$$

Proceeding as in the proof of Theorem 3.1, we obtain, for all $t \in \mathbb{N}$,

$$\mathbb{E}\Big[\mathrm{GFT}_t\big(\mathbb{E}[V]\big) - \mathrm{GFT}_t(P_t)\Big] \le M\mathbb{E}\left[\left|\mathbb{E}[V] - P_t\right|^2\right]\,.$$

Putting everything together, we get, for all $T \ge T_0 + 1$

$$\begin{aligned}R_T - T_0 + \frac{1}{2} &\le \sum_{t=T_0+1}^{T} \mathbb{E}\Big[\mathrm{GFT}_t\big(\mathbb{E}[V]\big) - \mathrm{GFT}_t(P_t)\Big] \\ &\le M \sum_{t=T_0+1}^{T} \mathbb{E}\left[\left|\mathbb{E}[V] - P_t\right|^2\right] = M \sum_{t=T_0+1}^{T} \mathbb{E}\left[\left|\mathbb{E}[V] - P_{T_0+1}\right|^2\right] \\ &\le M(T - T_0)\left(\frac{2}{T_0^2} + \frac{1}{T_0}\right)\,.\end{aligned}$$

Substituting the selected parameters in the final expression yields the second part of the result. □

4.2 Lower Bound

In this section, we prove the optimality of our ETC algorithm by showing a matching $\sqrt{MT}$ lower bound. For a high-level overview of its proof, we refer the reader back to Section 1.3.

Theorem 4.3. *There exist two numerical constants $c_1, c_2 > 0$ such that, for any $M \ge 2$ and any time horizon $T \ge c_2 M^3$, the worst-case regret of any algorithm satisfies*

$$\sup_{\nu} R_T^{\nu} \ge c_1\sqrt{MT}\,,$$

where R_T^{ν} is the regret at time T of the algorithm when the underlying i.i.d. sequence of traders' valuations follows the distribution ν, and $\mathcal{D}_M$ is the set of all distributions ν with density bounded by M.

Proof sketch. Fix $M \ge 2$ and $T \in \mathbb{N}$. We will use the same random variables, distributions, densities, and notation as in the proof of Theorem 3.2. We will show that for each algorithm for the

2-bit feedback setting and each time horizon T, if R_T^ν is the regret of the algorithm at time horizon T when the underlying distribution of the traders' valuations is ν, then $\max\left(R_T^{\nu_{-\varepsilon}}, R_T^{\nu_{+\varepsilon}}\right) = \Omega\left(\sqrt{MT}\right)$ if $T = \Omega(M^3)$.

Note that for all $\varepsilon > 0$, $t \in \mathbb{N}$, and $p < \frac{1}{2}$

$$\mathbb{E}\left[\mathrm{GFT}_{\varepsilon,t}\left(1/2\right)\right] \geq \mathbb{E}\left[\mathrm{GFT}_{\varepsilon,t}(p)\right] . \tag{6}$$

Similarly, for all $\varepsilon < 0$, $t \in \mathbb{N}$, and $p > \frac{1}{2}$,

$$\mathbb{E}\left[\mathrm{GFT}_{\varepsilon,t}\left(1/2\right)\right] \geq \mathbb{E}\left[\mathrm{GFT}_{\varepsilon,t}(p)\right] . \tag{7}$$

Furthermore, a direct verification shows that, for each $\varepsilon \in [-1,1]$ and $t \in \mathbb{N}$,

$$\max_{p\in[0,1]} \mathbb{E}\left[\mathrm{GFT}_{\varepsilon,t}(p)\right] - \max_{p\in[\frac{1}{7},\frac{2}{7}]} \mathbb{E}\left[\mathrm{GFT}_{\varepsilon,t}(p)\right] \geq \frac{1}{50} = \Omega(1) . \tag{8}$$

Now, assume that $T \geq M^3/14^4$ so that, defining $\varepsilon := (MT)^{-1/4}$, we have that the maximizer of the expected gain from trade $\frac{1}{2} + \frac{\varepsilon}{196}$ belongs to the spike region J_M. In the $+\varepsilon$ (resp., $-\varepsilon$) case, the optimal price belongs to the region $\left(\frac{1}{2}, \frac{1}{2} + \frac{1}{14M}\right]$ (resp., $\left[\frac{1}{2} - \frac{1}{14M}, \frac{1}{2}\right)$). By posting prices in the wrong region $\left[0, \frac{1}{2}\right]$ (resp., $\left[\frac{1}{2}, 1\right]$) in the $+\varepsilon$ (resp., $-\varepsilon$) case, the learner incurs a $\Omega(M\varepsilon^2) = \Omega\left(\sqrt{M/T}\right)$ instantaneous regret by (3) and (6) (resp., (3) and (7)). Then, in order to attempt suffering less than $\Omega\left(\sqrt{M/T}\cdot T\right) = \Omega\left(\sqrt{MT}\right)$ regret, the algorithm would have to detect the sign of $\pm\varepsilon$ and play accordingly. We will show now that even this strategy will not improve the regret of the algorithm (by more than a constant) because of the cost of determining the sign of $\pm\varepsilon$ with the available feedback. Since the feedback received from the two traders at time t by posting a price p is $\mathbb{I}\{p \leq V_{\pm\varepsilon,2t-1}\}$ and $\mathbb{I}\{p \leq V_{\pm\varepsilon,2t}\}$, the only way to obtain information about (the sign of) $\pm\varepsilon$ is to post in the costly ($\Omega(1)$-instantaneous regret by Eq. (8)) sub-optimal region $[\frac{1}{7}, \frac{2}{7}]$ (see Fig. 1). However, posting prices in the region $[\frac{1}{7}, \frac{2}{7}]$ at time t can't give more information about $\pm\varepsilon$ than the information carried by $V_{\pm\varepsilon,2t-1}$ and $V_{\pm\varepsilon,2t}$, which, in turn, can't give more information about $\pm\varepsilon$ than the information carried by the two Bernoullis $B_{\frac{1\pm\varepsilon}{2},2t-1}$ and $B_{\frac{1\pm\varepsilon}{2},2t}$. Since (via an information-theoretic argument) in order to distinguish the sign of $\pm\varepsilon$ having access to i.i.d. Bernoulli random variables of parameter $\frac{1\pm\varepsilon}{2}$ requires $\Omega(1/\varepsilon^2) = \Omega\left(\sqrt{MT}\right)$ samples, we are forced to post at least $\Omega\left(\sqrt{MT}\right)$ prices in the costly region $\left[\frac{1}{7}, \frac{2}{7}\right]$ suffering a regret of $\Omega\left(\sqrt{MT}\right)\cdot\Omega(1) = \Omega\left(\sqrt{MT}\right)$. Putting everything together, no matter what the strategy, each algorithm will pay at least $\Omega\left(\sqrt{MT}\right)$ regret. □

5 BEYOND BOUNDED DENSITIES

In this section, we investigate how the problem changes when the bounded density assumption is no longer guaranteed to hold.

5.1 The Full Feedback Case

Our Representation Lemma shows that our goal is to maximize $\rho(\nu)$, where ρ is known (by Lemma 2.2) but ν is not (by assumption). A natural strategy is then to approximate ν through an empirical distribution $\hat{\nu}_t$ and then maximize $\rho(\hat{\nu}_t)$. This is precisely what we do in Algorithm 3. Note that maximizing $\rho(\hat{\nu}_t)$ can be done efficiently by an exhaustive search of $\Theta(t)$ candidate points. The next result gives a regret guarantee of $\sqrt{T}$ for FTρ.

Algorithm 3: Follow-the-ρ (FTρ)

Post $P_1 := 1/2$, then receive feedback V_1, V_2;
for *time* $t = 2, 3, \ldots$ **do**
  Let $\hat{\nu}_t := \frac{1}{2(t-1)} \sum_{s=1}^{2(t-1)} \delta_{V_s}$;
  Post $P_t \in \mathrm{argmax}_{p\in[0,1]} := \rho(\hat{\nu}_t)(p)$;
  Receive feedback V_{2t-1}, V_{2t};

Theorem 5.1. *The regret of FTρ satisfies, for all time horizons T,*

$$R_T \leq 1/2 + 4\left(3\sqrt{\pi} + \sqrt{2}\right)\sqrt{T-1} .$$

Proof sketch. For any price $p^\star \in [0,1]$, and time step $t \geq 2$,

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{GFT}_t(p^\star)\right] - \mathbb{E}\left[\mathrm{GFT}_t(P_t)\right] \overset{\spadesuit}{=} \rho(\nu)(p^\star) - \mathbb{E}\left[\rho(\nu)(P_t)\right] \\
&= \mathbb{E}\left[\rho(\nu)(p^\star) - \rho(\hat{\nu}_t)(p^\star)\right] + \mathbb{E}\left[\rho(\hat{\nu}_t)(p^\star) - \rho(\hat{\nu}_t)(P_t)\right] \\
&\quad + \mathbb{E}\left[\rho(\hat{\nu}_t)(P_t) - \rho(\nu)(P_t)\right] \overset{\clubsuit}{\leq} 2\mathbb{E}\left[\sup_{p\in[0,1]} \left|\rho(\hat{\nu}_t)(p) - \rho(\nu)(p)\right|\right] \\
&\overset{\blacklozenge}{\leq} 2\mathbb{E}\left[6 \sup_{\lambda\in[0,1]} \left|(\nu - \nu_t)[0,\lambda]\right|\right] + 2\mathbb{E}\left[\left|\mathbb{E}[V] - \frac{\sum_{s=1}^{2(t-1)} V_s}{2(t-1)}\right|\right] \\
&\overset{\heartsuit}{\leq} 12\int_0^{+\infty} \mathbb{P}\left[\sup_{\lambda\in[0,1]} \left|(\nu - \nu_t)[0,\lambda]\right| \geq \varepsilon\right] \mathrm{d}\varepsilon + \frac{4}{\sqrt{2(t-1)}} \\
&\overset{\bullet}{\leq} 24\int_0^{+\infty} e^{4\varepsilon^2(t-1)}\,\mathrm{d}\varepsilon + \frac{4}{\sqrt{2(t-1)}} \leq 2\frac{3\sqrt{\pi}+\sqrt{2}}{\sqrt{t-1}} ,
\end{aligned}$$

where ♠ follows by Lemma 2.2 and the Freezing Lemma; ♣ by definition of P_t; ♦ by elementary computations; ♥ by Fubini's Theorem and upper bounding $\mathbb{E}\left[|\ldots|\right]$ with the variance of $\frac{\sum_{s=1}^{2(t-1)} V_s}{2(t-1)}$; and • by the DKW inequality (see, e.g., [8, Theorem J.1]). Summing over t from 2 to T, yields the result. □

Next, we show that running FTM until evidence is observed that ν does not have a bounded density (i.e., until we observe the same sample twice), then switching to FTρ (FTMTρ, Algorithm 4), keeps

Algorithm 4: Follow-the-Mean-then-ρ (FTMTρ)

for *time* $t = 1, 2, \ldots$ **do**
  Post P_t according to FTM (Algorithm 1);
  if $\{V_{2t-1}, V_t\} \cap \{V_1, \ldots, V_{2(t-1)}\} \neq \varnothing$ **then**
    $\tau := t$ and **break**;
Run FTρ (Algorithm 3) up to time τ without posting prices;
for *time* $t = \tau+1, \tau+2, \ldots$ **do**
  Post P_t according to FTρ (Algorithm 3);

the guarantees of FTM if ν has a bounded density and those of FTρ if ν does not, without requiring this *a priori* knowledge.

Theorem 5.2. *For all time horizons T, the regret of FTMTρ satisfies*

(1) If ν has density bounded by some constant $M > 0$ and $T \geq 2$,

$$R_T \leq \frac{1}{2} + \frac{M}{4}\left(1 + \ln(T-1)\right) .$$

(2) Otherwise, $R_T \leq 7.5 + 6\left(2\sqrt{\pi} + \sqrt{2}\right)\sqrt{T-1}$.

Proof. If ν has density bounded by some constant $M > 0$, then the condition $\{V_{2t-1}, V_t\} \cap \{V_1, \dots, V_{2(t-1)}\} \neq \varnothing$ never occurs with probability 1; therefore, the expected regret of FTMTρ coincides with that of FTM, and Item 1 follows by Theorem 3.1.

For Item 2, define the following: $\varepsilon := \sup_{p\in[0,1]} \nu\{p\}$; $p_\varepsilon \in [0,1]$ such that $\nu\{p\} \geq \varepsilon/2$; $\tau' := \tau \wedge T$; $\tau'' := \inf\{t \in \mathbb{N} \mid \sum_{s=1}^{t} \mathbb{I}\{V_{2s} = p_\varepsilon\} \geq 2\}$, and note that $\tau' \leq \tau \leq \tau''$; $P'_1, P'_2, \dots$ (resp., $P''_1, P''_2, \dots$) are the prices posted by FTM (resp., FTρ) run with feedback $V_1, V_2, \dots$; $p^\star \in \operatorname{argmax}_{p\in[0,1]} \rho(\nu)(p)$; $X_t(p) := \mathrm{GFT}_t(p^\star) - \mathrm{GFT}_t(p)$ for all $t \in \mathbb{N}$ and $p \in [0,1]$; $x(p) := \rho(\nu)(p^\star) - \rho(\nu)(p)$ for all $p \in [0,1]$, and note that $x \geq 0$. Then:

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} x(P_t)\right] = \mathbb{E}\left[\sum_{t=1}^{\tau'} x(P'_t)\right] + \mathbb{E}\left[\sum_{t=\tau'+1}^{T} x(P''_t)\right] \leq \mathbb{E}\left[\sum_{t=1}^{\tau'} x(P'_t)\right] + \mathbb{E}\left[\sum_{t=1}^{T} x(P''_t)\right],$$

where the first equality follows from the Freezing Lemma, the inequality from $x \geq 0$, and note that the second expectation in the last formula is the regret of FTρ (by the Freezing Lemma), and can be controlled applying Theorem 5.1. For the first term, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{\tau'} x(P'_t)\right] - \varepsilon\mathbb{E}[\tau'] &\leq \mathbb{E}\left[\sum_{t=1}^{\tau'} \widetilde{\rho}(\nu)\big(\mathbb{E}[V]\big) - \sum_{t=1}^{\tau'} \widetilde{\rho}(\nu)(P'_t)\right] \\ &\leq 2\sum_{t=1}^{T} \mathbb{E}\Big[\big|\mathbb{E}[V] - P'_t\big|\Big] \leq 1 + 2\sum_{t=2}^{T} \sqrt{\mathrm{Var}\left(\frac{\sum_{s=1}^{2(t-1)} V_s}{2(t-1)}\right)} \\ &= 1 + \sqrt{2}\sum_{t=1}^{T-1} t^{-1/2} \leq 1 + 2\sqrt{2(T-1)}\,,\end{aligned}$$

where the first inequality follows from the definition of ρ and the second from Lemma 2.1. Noting that τ'' has a negative binomial distribution with parameters 2 and $\varepsilon/2$, we get

$$\varepsilon\mathbb{E}[\tau'] \leq \varepsilon\mathbb{E}[\tau''] = \varepsilon \cdot 2\frac{1-\varepsilon/2}{\varepsilon/2} \leq 4\,.$$

Putting everything together gives the result. □

We now show that when ν does not have a bounded density, the $\sqrt{T}$ guarantee of FTρ and FTMTρ is optimal.

Theorem 5.3. *There exists a numerical constant $c > 0$ such that, for any time horizon T, the worst-case regret of any algorithm satisfies*

$$\sup_{\nu} R_T \geq c\sqrt{T}\,,$$

where the sup *is over all distributions ν.*

Proof sketch. For each $\varepsilon \in \left[-\frac{1}{8}, \frac{1}{8}\right]$, consider the distribution $\nu_\varepsilon := \frac{1}{4}\delta_0 + \left(\frac{1}{4}+\varepsilon\right)\delta_{1/3} + \left(\frac{1}{4}-\varepsilon\right)\delta_{2/3} + \frac{1}{4}\delta_1$. Then, prices $p \in \left\{\frac{1}{3}, \frac{2}{3}\right\}$ are either $\Theta(\varepsilon)$-suboptimal or optimal (depending on the sign of ε), and the remaining prices p in $[0,1]$ are $\Omega(1)$-suboptimal. Noting that $\Omega(\varepsilon T)$ is the regret paid by posting suboptimal prices for T time steps and, via an information-theoretic argument, that $\Omega(1/\varepsilon^2)$ rounds are needed to determine the sign of ε, one can conclude that the regret of any algorithm is $\Omega\big(\min\big(\varepsilon T, \varepsilon\frac{1}{\varepsilon^2}\big)\big) = \Omega\big(\sqrt{T}\big)$, if $\varepsilon = T^{-1/2}$. See the supplementary material for additional details. □

5.2 The Two-Bit Feedback Case

Finally, learning becomes impossible if the bounded density assumption on ν is lifted in the two-bit feedback case. We show this by leveraging a needle-in-a-haystack phenomenon. The idea is that any deterministic algorithm is capable of posting just a finite number of prices since all the possible sequences of 2-bit feedback, up to a certain time horizon T, are 4^T. The goal is then to design a hard instance whose maximizer is never played by the algorithm and whose value is $\Omega(1)$ higher than the other values.

Theorem 5.4. *For any time horizon T, the worst-case regret of any algorithm satisfies*

$$\sup_{\nu} R_T \geq \frac{T}{9}\,,$$

where the sup *is over all distributions ν.*

Proof. Given that we are in a stochastic i.i.d. setting, it is enough to consider deterministic algorithms. Let $(\alpha_t)_{t\in\mathbb{N}}$ be a deterministic algorithm for the two-bit feedback setting, i.e., a sequence of functions $\alpha_t : \big(\{0,1\}\times\{0,1\}\big)^{t-1} \to [0,1]$ mapping past feedback into prices (with the convention that α_1 is just a number in $[0,1]$). Fix a time horizon T. Note that there are 4^T different sequences of pairs of zeroes and ones representing the feedback the algorithm could receive up to time T. Hence, the algorithm $(\alpha_t)_{t\in\mathbb{N}}$ selects the prices it posts in a set A which has no more than 4^T different prices. For each $\eta \in \left(0, \frac{1}{2}\right)$, select any $x \in \left(\frac{1}{2}-\eta, \frac{1}{2}\right) \setminus A$ and define

$$\nu_x := \frac{1}{3}\delta_0 + \frac{1}{3}\delta_x + \frac{1}{3}\delta_1\,.$$

Consider an i.i.d. sequence of traders' valuations $V_1, V_2, \dots$ with common distribution ν_x, and for each $t \in \mathbb{N}$ and $p \in [0,1]$, let $\mathrm{GFT}_t(p) := \mathrm{g}(p, V_{2t-1}, V_{2t})$. For each $t \in \mathbb{N}$, notice that

$$\mathbb{E}\big[\mathrm{GFT}_t(x)\big] = \max_{p\in[0,1]} \mathbb{E}\big[\mathrm{GFT}_t(p)\big]$$

$$\mathbb{E}\big[\mathrm{GFT}_t(x)\big] - \max_{p\in[0,1]\setminus\{x\}} \mathbb{E}\big[\mathrm{GFT}_t(p)\big] = \frac{4}{9} - \frac{4}{9} + \frac{2}{9}x \geq \frac{1-2\eta}{9}\,.$$

Given that the algorithm never posts the price x up to time T, it follows that the regret at time T of the algorithm is lower bounded by $\frac{1-2\eta}{9}T$. Given that η was arbitrarily chosen, the worst-case regret of the algorithm at time T is lower bounded by $\frac{T}{9}$. □

6 FUTURE RESEARCH DIRECTIONS

Our paper motivates further study of brokerage in a few different directions. It would be interesting to study the "contextual" version of this problem, where some information about the traders (and, therefore, about their valuations for the item being traded) is available to the broker *before* making a decision. Another intriguing topic is discovering intermediate cases for which something meaningful could be said between the stationary (that we fully fleshed out) and adversarial (that, as discussed in the Related Work section, is unlearnable) cases. Finally, a natural question in bilateral trade is learning to optimize performance measures focused more on the broker's earnings than the traders'; for instance, rather than the gain from trade, a natural objective is to maximize the total number of trades, as it directly increases the broker's business volume and, consequently, their profits.

REFERENCES

[1] Thomas Archbold, Bart de Keijzer, and Carmine Ventre. 2023. Non-Obvious Manipulability for Single-Parameter Agents and Bilateral Trade. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, USA, 2107–2115.

[2] Yossi Azar, Amos Fiat, and Federico Fusco. 2022. An alpha-regret analysis of Adversarial Bilateral Trade. *Advances in Neural Information Processing Systems* 35 (2022), 1685–1697.

[3] Moshe Babaioff, Kira Goldner, and Yannai A. Gonczarowski. 2020. Bulow-Klemperer-Style Results for Welfare Maximization in Two-Sided Markets. In *Proceedings of the Thirty-First Annual ACM-SIAM Symposium on Discrete Algorithms* (Salt Lake City, Utah) *(SODA '20)*. Society for Industrial and Applied Mathematics, USA, 2452–2471.

[4] Richard F Bass. 2013. *Real analysis for graduate students*. Createspace Ind Pub, USA.

[5] Liad Blumrosen and Yehonatan Mizrahi. 2016. Approximating Gains-from-Trade in Bilateral Trading. In *Web and Internet Economics, WINE'16 (Lecture Notes in Computer Science, Vol. 10123)*. Springer, Germany, 400–413.

[6] Nataša Bolić, Tommaso Cesari, and Roberto Colomboni. 2023. *An Online Learning Theory of Brokerage*. Technical Report. arXiv preprint arXiv:2310.12107.

[7] Johannes Brustle, Yang Cai, Fa Wu, and Mingfei Zhao. 2017. Approximating Gains from Trade in Two-Sided Markets via Simple Mechanisms. In *Proceedings of the 2017 ACM Conference on Economics and Computation* (Cambridge, Massachusetts, USA) *(EC '17)*. Association for Computing Machinery, New York, NY, USA, 589–590.

[8] Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. 2023. Bilateral trade: A regret minimization perspective. *Mathematics of Operations Research* 0, 0 (2023), null.

[9] Nicolò Cesa-Bianchi, Tommaso Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. 2024. *Regret Analysis of Bilateral Trade with a Smoothed Adversary*. Technical Report. hal preprint hal-04383576.

[10] Nicolò Cesa-Bianchi, Tommaso R Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. 2021. A regret analysis of bilateral trade. In *Proceedings of the 22nd ACM Conference on Economics and Computation*. Association for Computing Machinery, USA, 289–309.

[11] Nicolò Cesa-Bianchi, Tommaso R Cesari, Roberto Colomboni, Federico Fusco, and Stefano Leonardi. 2023. Repeated Bilateral Trade Against a Smoothed Adversary. In *The Thirty Sixth Annual Conference on Learning Theory*. PMLR, PMLR, USA, 1095–1130.

[12] Nicolo Cesa-Bianchi and Gábor Lugosi. 2006. *Prediction, learning, and games*. Cambridge University Press, UK.

[13] Tommaso R Cesari and Roberto Colomboni. 2021. A nearest neighbor characterization of Lebesgue points in metric measure spaces. *Mathematical Statistics and Learning* 3, 1 (2021), 71–112.

[14] Riccardo Colini-Baldeschi, Bart de Keijzer, Stefano Leonardi, and Stefano Turchetta. 2016. Approximately Efficient Double Auctions with Strong Budget Balance. In *ACM-SIAM Symposium on Discrete Algorithms, SODA'16*. SIAM, USA, 1424–1443.

[15] Riccardo Colini-Baldeschi, Paul W. Goldberg, Bart de Keijzer, Stefano Leonardi, and Stefano Turchetta. 2017. Fixed Price Approximability of the Optimal Gain from Trade. In *Web and Internet Economics, WINE'17 (Lecture Notes in Computer Science, Vol. 10660)*. Springer, Germany, 146–160.

[16] Riccardo Colini-Baldeschi, Paul W Goldberg, Bart de Keijzer, Stefano Leonardi, Tim Roughgarden, and Stefano Turchetta. 2020. Approximately efficient two-sided combinatorial auctions. *ACM Transactions on Economics and Computation (TEAC)* 8, 1 (2020), 1–29.

[17] Yuan Deng, Jieming Mao, Balasubramanian Sivan, and Kangning Wang. 2022. Approximately efficient bilateral trade. In *STOC*. ACM, Italy, 718–721.

[18] Paul Dütting, Federico Fusco, Philip Lazos, Stefano Leonardi, and Rebecca Reiffenhäuser. 2021. Efficient Two-Sided Markets with Limited Information. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (Virtual, Italy) *(STOC 2021)*. Association for Computing Machinery, New York, NY, USA, 1452–1465.

[19] Bank for International Settlements. 2022. OTC derivatives statistics at end-June 2022. https://www.bis.org/publ/otc_hy2211.pdf

[20] Zi Yang Kang, Francisco Pernice, and Jan Vondrák. 2022. Fixed-price approximations in bilateral trade. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, Society for Industrial and Applied Mathematics, Alexandria, VA, USA, 2964–2985.

[21] Robert E Lucas Jr. 1989. The effects of monetary shocks when prices are set in advance.

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

[23] Katerina Sherstyuk, Krit Phankitnirundorn, and Michael J Roberts. 2020. Randomized double auctions: gains from trade, trader roles, and price discovery. *Experimental Economics* 24, 4 (2020), 1–40.

[24] Pierre-Olivier Weill. 2020. The search theory of over-the-counter markets. *Annual Review of Economics* 12 (2020), 747–773.

[25] David Williams. 1991. *Probability with martingales*. Cambridge university press, UK.