A Regret Analysis of Bilateral Trade

Nicolò Cesa-Bianchi, Tommaso R Cesari, Roberto Colomboni, Federico Fusco, Stefano Leonardi

A Regret Analysis of Bilateral Trade*

Nicolò Cesa-Bianchi[1], Tommaso R. Cesari[2,3], Roberto Colomboni[1,4], Federico Fusco[5], and Stefano Leonardi[5]

[1]Università degli Studi di Milano, Milano, Italy
[2]Toulouse School of Economics (TSE), Toulouse, France
[3]Artificial and Natural Intelligence Toulouse Institute (ANITI), Toulouse, France
[4]Istituto Italiano di Tecnologia, Genova, Italy
[5]Sapienza Università di Roma, Roma, Italy

February 15, 2021

Abstract

Bilateral trade, a fundamental topic in economics, models the problem of intermediating between two strategic agents, a seller and a buyer, willing to trade a good for which they hold private valuations. Despite the simplicity of this problem, a classical result by Myerson and Satterthwaite (1983) affirms the impossibility of designing a mechanism which is simultaneously efficient, incentive compatible, individually rational, and budget balanced.

This impossibility result fostered an intense investigation of meaningful trade-offs between these desired properties. Much work has focused on approximately efficient fixed-price mechanisms, i.e., Blumrosen and Dobzinski (2014; 2016), Colini-Baldeschi et al. (2016), which have been shown to fully characterize strong budget balanced and ex-post individually rational direct revelation mechanisms. All these results, however, either assume some knowledge on the priors of the seller/buyer valuations, or a black box access to some samples of the distributions, as in Dütting et al. (2021).

In this paper, we cast for the first time the bilateral trade problem in a regret minimization framework over T rounds of seller/buyer interactions, with no prior knowledge on the private seller/buyer valuations. Our main contribution is a complete characterization of the regret regimes for fixed-price mechanisms with different models of feedback and private valuations, using as benchmark the best fixed price in hindsight. More precisely, we prove the following bounds on the regret:

- $\widetilde{\Theta}(\sqrt{T})$ for full-feedback (i.e., direct revelation mechanisms);
- $\widetilde{\Theta}(T^{2/3})$ for realistic feedback (i.e., posted-price mechanisms) and independent seller/buyer valuations with bounded densities;
- $\Theta(T)$ for realistic feedback and seller/buyer valuations with bounded densities;
- $\Theta(T)$ for realistic feedback and independent seller/buyer valuations;
- $\Theta(T)$ for the adversarial setting.

*Partially supported by ERC Advanced Grant 788893 AMDROMA "Algorithmic and Mechanism Design Research in Online Markets" and MIUR PRIN project ALGADIMAR "Algorithms, Games, and Digital Markets". This work has also benefited from the AI Interdisciplinary Institute ANITI. ANITI is funded by the French "Investing for the Future – PIA3" program under the Grant agreement n. ANR-19-PI3A-0004.

Contents

1 Introduction

In the bilateral trade problem, two strategic agents —a seller and a buyer— wish to trade some good. They both privately hold a personal valuation for it, and strive to maximize their own quasi-linear utility. An ideal mechanism for this problem would optimize the efficiency, i.e., the social welfare resulting by trading the item, while enforcing incentive compatibility (IC) and individual rationality (IR). The assumption that makes two-sided mechanism design more complex than the one-sided counterpart is budget balance (BB): the mechanism cannot subsidize or make a profit from the market. Unfortunately, as Vickrey observed in his seminal work Vickrey (1961), the optimal incentive compatible mechanism maximizing social welfare for bilateral trade may not be budget balanced.

A more general result due to Myerson and Satterthwaite Myerson and Satterthwaite (1983) shows that a fully efficient mechanism for bilateral trade that satisfies IC, IR, and BB may not exist at all. This impossibility result holds even if prior information on the buyer and seller's valuations is available, the truthful notion is relaxed to Bayesian incentive compatibility (BIC), and the exact budget balance constraint is loosened to weak budget balance (WBB). To circumvent this obstacle, a long line of research has focused on the design of approximating mechanisms that satisfy the above requirements while being nearly efficient.

These approximation results build on a Bayesian assumption: seller and buyer valuations are drawn from two distributions, which are both known to the mechanism designer. Although in some sense necessary —without any information on the priors there is no way to extract any meaningful approximation result Dütting et al. (2021)— this assumption is unrealistic. Following a recent line of research Cesa-Bianchi et al. (2015); Lykouris et al. (2016); Daskalakis and Syrgkanis (2016), in this work we study this basic mechanism design problem in a regret minimization setting. Our goal is bounding the total loss in efficiency experienced by the mechanism in the long period by learning the salient features of the prior distributions.

At each time t, a new seller/buyer pair arrives. The seller has a private valuation $s_t \in [0,1]$ representing the smallest price she is willing to accept in order to trade. Similarly, the buyer has a private value $b_t \in [0,1]$ representing the highest price that she will pay for the item. The mechanism sets a price p_t which results in a trade if and only if $s_t \le p_t \le b_t$.

There are two common utility functions that reflect the performance of the mechanism at each time step: the social welfare, which sums the utilities of the two players after the trade (and remains equal to the seller's valuation if no trade occurs), and the gain from trade, consisting in the net gain in the utilities. In formulae,

- **Social Welfare**: $\mathrm{SW}(p_t, s_t, b_t) = \mathrm{SW}_t(p_t) = s_t + (b_t - s_t)\mathbb{I}\{s_t \le p \le b_t\}$;
- **Gain from Trade**: $\mathrm{GFT}(p_t, s_t, b_t) = \mathrm{GFT}_t(p_t) = (b_t - s_t)\mathbb{I}\{s_t \le p \le b_t\}$.

We begin by investigating the standard assumption in which s_t and b_t are realizations of S_t and B_t, where $(S_1, B_1), (S_2, B_2), \dots$ are i.i.d. random variables, supported in $[0,1]^2$, representing the valuations of seller and buyer respectively (stochastic i.i.d. setting). We also consider the case where $(s_1, b_1), (s_2, b_2), \dots$ is an arbitrary deterministic process (adversarial setting).

In our online learning framework, we aim at minimizing the *regret* over a time horizon T:

$$\max_{p \in [0,1]} \mathbb{E}\left[\sum_{t=1}^{T} \mathrm{GFT}_t(p) - \sum_{t=1}^{T} \mathrm{GFT}_t(p_t)\right].$$

Note that since $\mathrm{GFT}_t(p_t) = \mathrm{SW}_t(p_t) - s_t$ and s_t does not depend on the choice of p, gain from trade and social welfare lead to the same notion of regret.

The regret is hence the difference between the expected total performance of our algorithm, which can only *sequentially learn* the distribution, and our reference benchmark, corresponding to the the best fixed-price strategy assuming *full knowledge* of the distribution. Our main goal is to design strategies with asymptotically vanishing time-averaged regret with respect to the best fixed-price strategy or, equivalently, regret sublinear in the time horizon T.

The class of fixed price mechanisms is of particular importance in bilateral trade as they are simple to implement, clearly truthful, individually rational, budget balanced, and enjoy the desirable property of

	Stochastic (iid)				Adversarial
	iid	**+iv**	**+bd**	**+iv+bd**	**adv**
Full	$T^{1/2}$ (thm 3.1)	$T^{1/2}$	$T^{1/2}$	$T^{1/2}$ (thm 3.2)	T (thm 5.1)
Real	T	T (thm 4.4)	T (thm 4.3)	$T^{2/3}$ (thms 4.1+4.2)	T

Table 1: Our main results for fixed price mechanisms. The rates are both upper and lower bounds, up to a $\log T$ factor. The slots without references are immediate consequences of the others.

asking the agents very little information. Moreover, it can be shown that fixed prices are the *only* direct revelation mechanisms which enjoy strong budget balance, dominant strategy incentive compatibility, and ex-post individual rationality Colini-Baldeschi et al. (2016).

To complete the description of the problem, we need to specify the feedback obtained by the mechanism after each sequential round. We propose two main feedback models:

- *Full feedback.* In the full feedback model, the pair (s_t, b_t) is revealed to the mechanism after the t-th trading round. The information collected by this feedback model corresponds to *direct revelation mechanisms*, where the agents communicate their valuations before each round, and the price proposed by the mechanism at time t only depends on past bids.
- *Realistic feedback.* In the harder realistic feedback model, only the relative orderings between s_t and p_t and between b_t and p_t are revealed after the t-th round. This model corresponds to *posted price mechanisms*, where seller and buyer separately accept or refuse the posted price. The price computed at time t only depends on past bids, and the values s_t and b_t are never revealed to the mechanisms.

1.1 Overview of our Results

We investigate the stochastic setting (under various assumptions), the adversarial setting, and how regret bounds change depending on the quality of the received feedback. In all cases, we provide matching upper and lower bounds. In particular, our positive result are constructive: explicit algorithms are given for each of them. More precisely, we show (see Table 1 for a summary):

- Algorithm 2 (Follow the Best Price) for the full-feedback model achieving a $\widetilde{O}(T^{1/2})$ regret in the stochastic (iid) setting (Theorem 3.1); this rate cannot be improved by more than a $\log T$ factor, not even under some additional natural assumptions (Theorem 3.2);
- Algorithm 3 (Scouting Bandits) for the harder realistic-feedback model achieving a $\widetilde{O}(T^{2/3})$ regret in a stochastic (iid) setting in which the valuations of the seller and the buyer are independent of each other (iv) and have bounded densities (bd) (Theorem 4.1); this rate cannot be improved by more than a $\log T$ factor (Theorem 4.2);
- impossibility results:
  - for the realistic-feedback model, if either the (iv) or the (bd) assumptions are dropped from the previous stochastic setting, no strategy can achieve sublinear worst-case regret (Theorems 4.2 and 4.4);
  - in an adversarial setting, no strategy can achieve sublinear worst-case regret, not even in the simpler full-feedback model (Theorem 3.2).

1.2 Technical Challenges

The two feedback models we consider are both challenging for different reasons.

Full feedback. The full feedback model fits nicely in the learning with expert advice framework Cesa-Bianchi and Lugosi (2006). Each price $p \in [0,1]$ can be viewed as an expert, and the revelation of s_t and b_t allows the mechanism to compute $\mathrm{GFT}_t(p)$ for all p, including the mechanism's own reward $\mathrm{GFT}_t(p_t)$. A common approach to reduce the cardinality of a continuous expert space is to assume some regularity (e.g., Lipschitzness) of the reward function, so that a finite grid of *representative* prices can be used. This approach yields a $\widetilde{O}(\sqrt{T})$ bound under density boundedness assumptions on the joint distribution of the seller and the buyer. By exploiting the structure of the reward function $\mathbb{E}\big[\mathrm{GFT}_t(\cdot)\big]$, we obtain the same regret bound without any assumptions on the distribution (other than iid). In Theorem 3.1, we show how to decompose the expression of the expected gain from trade in pieces that can be quickly learned via sampling. The full feedback received in each new round is used to refine the estimate of the actual gain from trade as a function of the price, while the posted prices are chosen so to maximize it. Our *Follow the Leader* strategy is shown to achieve a $\widetilde{O}(\sqrt{T})$ bound in the stochastic (iid) setting. This holds for arbitrary joint distributions of the seller and the buyer. In particular, even when the buyer and seller have a correlated behavior. The main challenge for the lower bounds is how to embed a hard instance in a setting where we cannot control the gain from trade, but only the distributions of seller and buyer. We solve this problem by designing a reduction from a 2-action partial monitoring game to our setting, and then using known lower bounds for partial monitoring.

Realistic Feedback. Here, at each time t, only $\mathbb{I}\{s_t \le p_t\}$ and $\mathbb{I}\{p_t \le b_t\}$ are revealed to the mechanism. Hence, we face the two competing goals of estimating the underlying distributions while optimizing the estimated gain from trade. The realistic feedback model does not fit the expert prediction framework, nor the harder bandits model Cesa-Bianchi and Lugosi (2006), because the observations of $\mathbb{I}\{s_t \le p_t\}$ and $\mathbb{I}\{p_t \le b_t\}$ are not enough to reconstruct the gain from trade at time t. On the one hand, if the trade occurs, there is no way to directly infer the difference $b_t - s_t$. On the other hand, if the trade does not occur, little can be done to argue which prices would have resulted in a trade. We show how to decompose the expected gain from trade at a posted price p into a *global* part that can be be quickly estimated by uniform sampling on the whole $[0,1]$ interval, and a *local* part that can be only learned by posting selected prices. Theorem 4.1 shows a general technique (Scouting Bandits, Algorithm 3) which takes advantage of this decomposition, and relies on any bandit algorithm to learn the local part of the expected gain from trade. We derive a sublinear regret of $\widetilde{O}(T^{2/3})$ in a stochastic (iid) setting in which the valuations of the seller and the buyer are independent of each other (iv) and have bounded densities (bd). Dropping the (iv) assumption leads to a pathological *lack of observability* phenomenon in which it is impossible to distinguish between two scenarios, and the optimal price in one of them is highly suboptimal in the other (Theorem 4.3). Dropping the (bd) assumption leads to a pathological *needle in a haystack* phenomenon in which all but one prices suffer a high regret and it is essentially impossible to find the optimal one among the continuum amount of suboptimal ones (Theorem 4.4). Similarly to the full feedback lower bound, the realistic feedback lower bound is based on reducing a partial monitoring game to our setting. However, additional challenges arise in this case due to the specific nature of the realistic feedback, see Theorems 4.3 and 4.4.

Adversarial setting. Finally, we investigate the adversarial setting in which the valuations of the buyer and the seller form an arbitrary deterministic process generated by an oblivious adversary. This setting is significantly more challenging than the stochastic (iid) case. Indeed, using a construction inspired by Cantor ternary set, we show that even under a full-feedback model, no strategy can lead to a sublinear worst-case regret

Lower Bound Techniques Due to space constraints, the proofs of the stochastic lower bounds, i.e., Theorems 3.2 and 4.2 to 4.4 are only sketched in the main text and completed in the Appendix. In particular, the formal reductions from various instances of partial monitoring rely on a very general notion of sequential games subsuming both partial monitoring and our problems. These reductions are shown through two key lemmas (B.1, B.2): an Embedding Lemma and e Simulation Lemma, which may be of independent interest.

1.3 Further Related Work

The study of the bilateral trade problem dates back to the already mentioned seminal works of Vickrey Vickrey (1961) and Myerson and Satterthwaite Myerson and Satterthwaite (1983). A more recent line of research focused on Bayesian mechanisms that achieve the IC, BB, and IR requirements while approximating the optimal social welfare or the gain form trade. Blumrosen and Dobzinski Blumrosen and Dobzinski (2014) proposed the *median mechanism* that sets a posted price equal to the median of the seller distribution and shows that this mechanism obtains an approximation factor of 2 to the optimal social welfare. Subsequent work by the same authors Blumrosen and Dobzinski (2016) improved the approximation guarantee to $e/(e-1)$ through a randomized mechanism whose prices depend on the seller distribution in a more intricate way. In Colini-Baldeschi et al. (2016) it is demonstrated that all DSIC mechanisms that are BB and IR must post a fixed price to the buyer and to the seller. In a different research direction aimed to characterize the information theoretical requirements of two-sided markets mechanisms, (Dütting et al., 2021) shows that setting the price equal to a single sample from the seller distribution gives a 2-approximation to the optimal social welfare. In a parallel line of work it has been considered the harder objective of approximating the *gain from trade*. An asymptotically tight fixed-price $O\big(\log \frac{1}{r}\big)$ approximation bound is also achieved in Colini-Baldeschi et al. (2017), with r being the probability that a trade happens (i.e., the value of the buyer is higher than the value of the seller). A BIC 2-approximation of the second best with a simple mechanism is obtained in Brustle et al. (2017).

In the following we discuss the relationship between the approximation results mentioned above and the regret analysis we develop in this work that compares online learning mechanisms against the best ex-ante fixed price mechanism. First of all, in the realistic feedback setting, the approximation mechanisms for bilateral trade cannot be easily implemented. For example, the single sample 2-approximation to the optimal social welfare Dütting et al. (2021) requires multiple rounds of interaction in order to obtain, approximately, a random sample from the distribution. The median mechanism of Blumrosen and Dobzinski (2014) requires an even larger number of rounds in order to estimate the median of the seller distribution. It is also interesting to relate the guarantee of our online algorithms with the one provided by the approximation mechanisms. Here we notice that the two approaches cannot be directly compared as there exist simple examples[1] showing for the median and the sample mechanisms, respectively, a factor of 2 and 4/3 away from the optimum fixed price ex-ante, whereas our online learning approach provides a strictly better sublinear regret.

There is a vast body of literature on regret analysis in (one-sided) dynamic pricing and online posted price auctions —see, e.g., the excellent survey published by den Boer (2015) and the tutorial slides by Slivkins and Zeevi (2015). In their seminal paper, Kleinberg and Leighton prove a $O(T^{2/3})$ upper bound (ignoring logarithmic factors) on the regret in the adversarial setting Kleinberg and Leighton (2003). Later works show simultaneous multiplicative and additive bounds on the regret when prices have range $[1, h]$ Blum et al. (2004); Blum and Hartline (2005). These bounds have the form $\varepsilon\, G_T^\star + O\big((h \ln h)/\varepsilon^2\big)$ ignoring $\ln \ln h$ factors, where $G_T^\star$ is the total revenue of the optimal price $p^\star$. Recent improvements on these results prove that the additive term can be made $O(p^\star(\ln h)/\varepsilon^2)$, where the linear scaling is now with respect to the optimal price rather than the maximum price h Bubeck et al. (2017). Other variants consider settings in which the number of copies of the item to sell is limited Agrawal and Devanur (2014); Babaioff et al. (2015); Badanidiyuru et al. (2018), buyers act strategically in order to maximize their utility in future rounds Amin et al. (2013); Devanur et al. (2019); Mohri and Medina (2014); Drutsa (2018), or there are features associated with the goods on sale Cohen et al. (2020). In the stochastic setting, previous works typically assume parametric Broder and Rusmevichientong (2012), locally smooth Kleinberg and Leighton (2003), or piecewise constant demand curves Cesa-Bianchi et al. (2019); den Boer and Keskin (2020).

[1] Consider a seller with value $\varepsilon > 0$ or 0 with equal probability and a buyer with value 1. The best fixed price has welfare of 1. For small ε, the median and the sample mechanism, respectively, obtains a welfare close to 1/2 and 3/4.

2 The Bilateral Trade learning protocol

In this section, we present the learning protocol for the sequential problem of bilateral trade (see learning protocol 1). We recall that the reward collected from a trade is the gain from trade, defined for all $p, s, b \in [0,1]$, by $\mathrm{GFT}(p,s,b) := (b-s)\mathbb{I}\{s \le p \le b\}$.

LEARNING PROTOCOL 1: Bilateral Trade

for *time* $t = 1, 2, \dots$ **do**
- a new seller/buyer pair arrives with (hidden) valuations $(S_t, B_t) \in [0,1]^2$;
- the learner posts a price $P_t \in [0,1]$;
- the learner receives a (hidden) reward $\mathrm{GFT}_t(P_t) := \mathrm{GFT}(P_t, S_t, B_t) \in [0,1]$;
- a feedback Z_t is revealed;

At each time step t, a seller and a buyer arrive, each with a privately held valuation $S_t, B_t \in [0,1]$. The learner then posts a price $P_t \in [0,1]$ and a trade occurs if and only if $S_t \le P_t \le B_t$. When this happens, the learner gains a reward $\mathrm{GFT}(P_t, S_t, B_t)$, which is not revealed. Some feedback Z_t is revealed instead. The nature of the sequence of valuations $(S_1, B_1), (S_2, B_2), \dots$ and feedbacks $Z_1, Z_2, \dots$ depends on the specific instance of the problem and is described below.

The goal of the learner is to determine a strategy α generating the prices $P_1, P_2, \dots$ (as in Learning Model 1) achieving sublinear *regret*

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

where the expectation is taken with respect to the sequence of buyers and sellers and (possibly) the internal randomization of α. To lighten the notation, we denote by $p^\star$ (one of) the $p \in [0,1]$ maximizing the previous expectation.

We now introduce several instances of bilateral trade, depending on the type of the received feedback and the nature of the environment.

2.1 Feedback

Full feedback: the feedback Z_t received at time t is the entire seller/buyer pair (S_t, B_t); in this setting, the seller and the buyer reveal their valuations at the end of a trade.

Realistic feedback: the feedback Z_t received at time t is the pair $\big(\mathbb{I}\{S_t \le P_t\}, \mathbb{I}\{P_t \le B_t\}\big)$; in this setting, the seller and the buyer only reveal whether or not they accept the trade at price P_t.

2.2 Environment

Stochastic (iid): $(S_1, B_1), (S_2, B_2), \dots$ is an i.i.d. sequence of seller/buyer pairs, where S_t and B_t could be (arbitrarily) correlated.

We will also investigate the (iid) setting under the following further assumptions.

Independent valuations (iv): S_t and B_t are independent of each other.

Bounded density (bd): (S_t, B_t) admits a joint density bounded by some $M \ge 1$.

Adversarial (adv): $(S_t, B_t)_{t \in \mathbb{N}}$ is an arbitrary deterministic sequence $(s_t, b_t)_{t \in \mathbb{N}} \subset [0,1]^2$.

3 Full-Feedback Stochastic (iid) Setting

We begin by considering the full-feedback model (corresponding to revelation mechanisms) under the assumption that the seller/buyer pairs $(S_1, B_1), (S_2, B_2), \dots$ are $[0,1]^2$-valued i.i.d. random variables, without any further assumptions on their common distribution (S, B) (in particular, S and B could be arbitrarily correlated). Here, sellers and buyers declare their actual valuations to the mechanism. The incentive compatibility is guaranteed by the fact that the posted prices does not depend on the declared valuations at each specific round, but only on past ones, so that there is no point in misreporting.

In Section 3.1, we show that a *Follow the Leader* approach, which we call Follow the Best Price (FBP), whose pseudocode is given in Algorithm 2, achieves a $O(\sqrt{T \log T})$ upper bound. In Section 3.2, we provide a lower bound that matches this rate, up to a $\sqrt{\log T}$ factor.

3.1 Follow the Best Price (FBP)

We begin by presenting our Follow the Best Price (FBP) algorithm. It consists in posting the best price with respect to the samples that have been observed so far. Notably, it does not need preliminary knowledge of the time horizon T.

ALGORITHM 2: Follow the Best Price (FBP)

init: select $P_1 \in [0,1]$ arbitrarily
for $t = 1, 2, \dots$ **do**
  post price P_t;
  receive feedback (S_t, B_t);
  compute $P_{t+1} \leftarrow \operatorname{argmax}_{p \in \{S_1, B_1, \dots, S_t, B_t\}} \sum_{i=1}^{t} \mathrm{GFT}(p, S_i, B_i) \in [0,1]$; ▷ `Ties broken arbitrarily`

For each time $t \geq 2$, let O_t be the sequence containing all the $t-1$ pairs of valuations observed so far, i.e., $O_t := \big((S_1, B_1), \dots, (S_{t-1}, B_{t-1})\big)$. Given O_t, one can reconstruct the actual $\mathrm{GFT}_i(\cdot) := \mathrm{GFT}(\cdot, S_i, B_i)$ function at each past time step $i \leq t-1$ and compute (one of) the best price(s)

$$P_t \in \operatorname*{argmax}_{p \in [0,1]} \sum_{i=1}^{t-1} \mathrm{GFT}_i(p) \,. \tag{1}$$

Note that at least one of the elements in the argmax belongs to the set of past valuations, given the structure of the gain from trade, so even a naive enumeration approach is computationally efficient.

Before moving on, we describe a property on P_t which will be useful in the analysis. Given the samples in O_t, it is possible to build an estimate of the random pair (S, B) of which they are i.i.d. samples. More precisely, one can consider the random pair (S'_t, B'_t) which follows the empirical distribution of the data, i.e., for all $(s, b) \in [0,1]^2$

$$\mathbb{P}\big[(S'_t, B'_t) = (s, b) \mid O_t\big] = \frac{1}{t-1} \sum_{i=1}^{t-1} \mathbb{I}\big\{(S_i, B_i) = (s, b)\big\} \,.$$

The formulation in Equation (1) is then equivalent to finding a price P_t that maximizes the expected gain from trade for the seller and buyer's valuations with respect to (S'_t, B'_t). Indeed, let GFT'_t be the gain from trade associated to (S'_t, B'_t), i.e., $\mathrm{GFT}'_t(\cdot) := \mathrm{GFT}(\cdot, S'_t, B'_t)$, then, for all $p \in [0,1]$,

$$\mathbb{E}\big[\mathrm{GFT}'_t(p) \mid O_t\big] = \frac{1}{t-1} \sum_{i=1}^{t-1} \mathrm{GFT}_i(p) \,, \tag{2}$$

where the expectation conditioned to O_t is with respect to a random sample of (S'_t, B'_t).

We show now that if at a certain time $t \geq 2$ the distribution of (S'_t, B'_t) given O_t is close to the distribution of (S, B), then our strategy performs well, on expectation, if compared to the optimal price $p^\star$. In order to do so, we first write the following decomposition:

$$\begin{aligned}\mathbb{E}\big[\mathrm{GFT}_t(p^\star)\big] - \mathbb{E}\big[\mathrm{GFT}_t(P_t) \mid O_t\big] = \mathbb{E}\big[\mathrm{GFT}_t(p^\star)\big] - \mathbb{E}\big[\mathrm{GFT}'_t(p^\star) \mid O_t\big] \\ + \mathbb{E}\big[\mathrm{GFT}'_t(p^\star) \mid O_t\big] - \mathbb{E}\big[\mathrm{GFT}'_t(p) \mid O_t\big]\big|_{p=P_t} \\ + \mathbb{E}\big[\mathrm{GFT}'_t(p) \mid O_t\big]\big|_{p=P_t} - \mathbb{E}\big[\mathrm{GFT}_t(p)\big]\big|_{p=P_t} .\end{aligned}$$

where, for a function g, we denoted $g(a) := g(p)\big|_{p=a}$. Note that the middle term is always non-positive, since P_t maximizes the expected gain from trade of (S'_t, B'_t) given O_t by (2). Hence, we have

$$\mathbb{E}\big[\mathrm{GFT}_t(p^\star)\big] - \mathbb{E}\big[\mathrm{GFT}_t(P_t) \mid O_t\big] \leq 2 \max_{q\in[0,1]} \Big|\mathbb{E}\big[\mathrm{GFT}_t(q)\big] - \mathbb{E}\big[\mathrm{GFT}'_t(q) \mid O_t\big]\Big| . \tag{3}$$

If at each time step $t \geq 2$ the mechanism inherits a good estimate of the distribution of (S_t, B_t), then we show that the best price given the past performs almost as well as the optimal price $p^\star$. This is a consequence of the following lemma, which reduces the problem to accurately estimating the distribution of (S_t, B_t) on the rectangles $\mathcal{R} = \big\{[a, b] \times [c, d] \mid a, b, c, d \in [0, 1]\big\}$.

Lemma 3.1 (First Decomposition Lemma)**.** *Let μ be any probability measure on $[0,1]^2$ and $p \in [0,1]$, then*

$$\mathbb{E}_{(s,b)\sim\mu}\big[\mathrm{GFT}(p, s, b)\big] = \int_0^1 \mu\Big[[0, p] \times \big[\max\{\lambda, p\}, 1\big]\Big]\,\mathrm{d}\lambda - \int_0^1 \mu\big[[\lambda, p] \times [p, 1]\big]\,\mathrm{d}\lambda.$$

As a consequence, for any $\varepsilon_t > 0$, in the event $\big\{\forall R \in \mathcal{R},\ \big|\mathbb{P}_{(S,B)}[R] - \mathbb{P}_{(S'_t,B'_t)|O_t}[R]\big| \leq \varepsilon_t\big\}$ (i.e., if the distribution of (S'_t, B'_t) given O_t is ε_t-close to that of (S_t, B_t), uniformly over rectangles), we have that

$$\max_{q\in[0,1]} \Big|\mathbb{E}\big[\mathrm{GFT}_t(q)\big] - \mathbb{E}\big[\mathrm{GFT}'_t(q)\big]\Big| \leq 2\varepsilon_t . \tag{4}$$

Proof. Consider μ and $p \in [0,1]$ as in the statement, we have

$$\begin{aligned}\mathbb{E}_{(s,b)\sim\mu}\big[\mathrm{GFT}(p, s, b)\big] &= \int_{[0,1]^2} (b - s)\mathbb{I}\{s \leq p \leq b\}\,\mathrm{d}\mu(s, b) = \int_{[0,p]\times[p,1]} (b - s)\,\mathrm{d}\mu(s, b) \\ &= \int_{[0,p]\times[p,1]} \left(\int_0^b \mathrm{d}\lambda - \int_0^s \mathrm{d}\lambda\right) \mathrm{d}\mu(s, b) \\ &= \int_0^1 \int_{[0,p]\times[p,1]} \mathbb{I}\{\lambda \leq b\}\,\mathrm{d}\mu(s, b)\,\mathrm{d}\lambda - \int_0^1 \int_{[0,p]\times[p,1]} \mathbb{I}\{\lambda \leq s\}\,\mathrm{d}\mu(s, b)\,\mathrm{d}\lambda \\ &= \int_0^1 \mu\big[[0, p] \times [\max\{\lambda, p\}, 1]\big]\,\mathrm{d}\lambda - \int_0^1 \mu\big[[\lambda, p] \times [p, 1]\big]\,\mathrm{d}\lambda .\end{aligned}$$

The consequence follows immediately from the decomposition and the fact that the subsets considered, i.e., $[0, p] \times \big[\max\{\lambda, p\}, 1\big]$ and $[\lambda, p] \times [p, 1]$ are indeed rectangles for all choices of p and λ, over which the two measures coincide up to an ε_t additive factor. □

We can now prove the regret guarantees of FBP that we claimed at the beginning of the section.

Theorem 3.1. *In the full-feedback stochastic (iid) setting, the regret of Follow the Best Price satisfies, for all $T \in \mathbb{N}$*

$$R_T(\mathrm{FBP}) \leq C\sqrt{T \log T}\,, \qquad \text{where } C \leq 90 .$$

Proof. For any time t, let $\varepsilon_t := 17\sqrt{\ln(16T)/t}$, and G_t be the event that the distribution of (S'_t, B'_t) given O_t is ε_t-close to that of (S_t, B_t), uniformly over rectangles, i.e.,

$$G_t := \left\{\forall R \in \mathcal{R},\ \left|\mathbb{P}_{(S,B)}[R] - \mathbb{P}_{(S'_t,B'_t)|O_t}[R]\right| \le \varepsilon_t\right\}.$$

Since the VC-dimension of $\mathcal{R}$ is 4, we have that $\mathbb{P}[G_t^c] \le \varepsilon_t$. This is an immediate consequence of VC-theory (see, e.g., (Mitzenmacher and Upfal, 2017, Theorem 14.15)). Putting this together with Eqs. (3) and (4), we conclude that for all $T \in \mathbb{N}$, the regret of the Follow the Best Price algorithm satisfies

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\mathrm{GFT}_t(p^\star) - \sum_{t=1}^{T}\mathrm{GFT}_t(P_t)\right] &= \sum_{t=1}^{T}\mathbb{E}\big[\mathbb{E}[\mathrm{GFT}_t(p^\star)] - \mathbb{E}[\mathrm{GFT}_t(P_t) \mid O_t]\big]\\ &\le \sum_{t=1}^{T}\mathbb{E}\big[\big(\mathbb{E}[\mathrm{GFT}_t(p^\star)] - \mathbb{E}[\mathrm{GFT}_t(P_t) \mid O_t]\big)\mathbb{I}_{G_t}\big] + \sum_{t=1}^{T}\mathbb{P}[G_t^c] \le \sum_{t=1}^{T} 2\cdot(2\varepsilon_t) + \sum_{t=1}^{T}\varepsilon_t \le 90\sqrt{T\ln T}.\end{aligned}$$

This concludes the proof. □

3.2 $\sqrt{T}$ Lower Bound (iv+bd)

In this section, we show that the upper bound on the minimax regret we proved in Section 3.1 is tight, up to logarithmic factors. No strategy can beat the $T^{1/2}$ rate when the seller/buyer pair (S_t, B_t) is drawn i.i.d. from an unknown fixed distribution, even under the further assumptions that the valuations of the seller and buyer are independent of each other and have bounded densities. For a full proof of the following theorem, see Appendix C.

Theorem 3.2. *In the full-feedback model, for all horizons T, the minimax regret $R_T^\star$ satisfies*

$$R_T^\star := \inf_{\alpha}\sup_{(S,B)\sim\mathcal{D}} R_T(\alpha) \ge c\sqrt{T},$$

where $c \ge 1/160$, the infimum is over all of the learner's strategies α, and the supremum is over all distributions $\mathcal{D}$ of the seller S and buyer B such that:

(iid) *$(S_1, B_1), (S_2, B_2), \ldots \sim (S, B)$ is an i.i.d. sequence;*

(iv) *S and B are independent of each other;*

(bd) *S and B admit densities bounded by $M \ge 4$.*

Prook sketch. We build a family of distributions $\mathcal{D}_{\pm\varepsilon}$ of the seller and buyer (S, B) parameterized by $\varepsilon \in [0, 1]$. For the seller, for any $\varepsilon \in [0, 1]$, we define the density

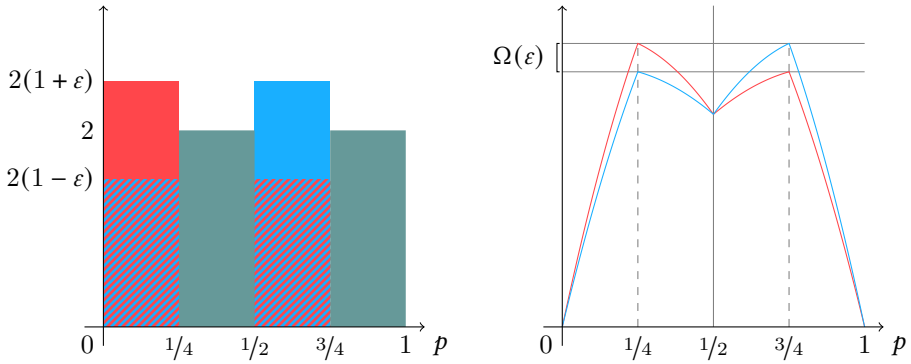
$$f_{S,\pm\varepsilon} := 2(1 \pm \varepsilon)\mathbb{I}_{[0,\frac{1}{4}]} + 2(1 \mp \varepsilon)\mathbb{I}_{[\frac{1}{2},\frac{3}{4}]}. \qquad \text{(Fig. 1(a), in red/blue)}$$

For the buyer, we define a single density (independently of ε)

$$f_B := 2\mathbb{I}_{[\frac{1}{4},\frac{1}{2}]\cup[\frac{3}{4},1]}. \qquad \text{(Fig. 1(a), in green)}$$

In the $+\varepsilon$ (resp., $-\varepsilon$) case, the optimal price belongs to the region $[0, 1/2]$ (resp., $(1/2, 1]$, see Fig. 1(b)). By posting prices in the wrong region $(1/2, 1]$ (resp., $[0, 1/2]$) in the $+\varepsilon$ (resp., $-\varepsilon$) case, the learner incurs a $\Omega(\varepsilon)$ regret. Thus, if ε is bounded-away from zero, the only way to avoid suffering linear regret is to identify the sign of $\pm\varepsilon$ and play accordingly.

This closely resembles the learning dilemma present in two-armed bandits. In fact, a technical proof (see Appendix C), shows that our setting is harder (i.e., it has a higher minimax regret) than an instance of a stochastic two-armed bandit problem, which has a known lower bound on its minimax regret of $\frac{1}{8}\left(\frac{1}{20}\sqrt{T}\right)$ Cesa-Bianchi and Lugosi (2006); Bubeck and Cesa-Bianchi (2012). □



(a) Distributions $f_{S_{\pm\varepsilon}}$ (red/blue) and f_B (green) (b) Expected GFT relative to $+\varepsilon$ (red) and $-\varepsilon$ (blue)

Figure 1: The best posted price is $1/4$ (resp., $3/4$) in the $+\varepsilon$ (resp., $-\varepsilon$) case. By posting $1/4$, the player suffers a $\Omega(\varepsilon)$ regret in the $-\varepsilon$ case, and the same is true posting $3/4$ if in $+\varepsilon$ case.

4 Realistic-Feedback Stochastic (iid) Setting

In this section, we tackle the problem in the more challenging realistic-feedback model, again under the assumption that the seller/buyer pairs $(S_1, B_1), (S_2, B_2), \dots$ are $[0,1]^2$-valued i.i.d. random variables, all distributed as a common (S, B). We will first study the case in which S and B are independent (iv) and have bounded densities (bd), then discuss what happens if either of the two assumptions is lifted.

We recall that in the realistic-feedback model, the only information collected by the mechanism at the end of each round t consists of $\mathbb{I}\{S_t \le P_t\}$ and $\mathbb{I}\{P_t \le B_t\}$. The main tool we use to leverage the structure of the objective function is the following decomposition Lemma. The first part of this result follows directly from specialising Equation (4) to independent distributions; alternatively, it can be directly derived, as in Myerson and Satterthwaite (1983).

Lemma 4.1 (Second decomposition Lemma)**.** *Let S and B be independent random variables in $[0,1]$, then for all prices $p \in [0,1]$ it holds*

$$\mathbb{E}\big[\mathrm{GFT}(p, S, B)\big] = \mathbb{P}[S \le p] \int_p^1 \mathbb{P}[B \ge \lambda]\, \mathrm{d}\lambda + \mathbb{P}[B \ge p] \int_0^p \mathbb{P}[S \le \lambda]\, \mathrm{d}\lambda\,. \tag{5}$$

Moreover, if S and B admit densities bounded above by $M \ge 1$, then $\mathbb{E}[\mathrm{GFT}(\cdot)]$ is Lipschitz in the prices, with constant $4M$.

Proof. We start form Equation (4) and use the independence of the distributions:

$$\begin{aligned}
\mathbb{E}\big[\mathrm{GFT}(p, S, B)\big] &= \int_0^1 \mathbb{P}_{(S,B)}\big[[0, p] \times [\max\{\lambda, p\}, 1]\big]\, \mathrm{d}\lambda - \int_0^1 \mathbb{P}_{(S,B)}\big[[\lambda, p] \times [p, 1]\big]\, \mathrm{d}\lambda \\
&= \mathbb{P}[S \le p] \int_0^1 \mathbb{P}\big[B \ge \max\{\lambda, p\}\big]\, \mathrm{d}\lambda - \mathbb{P}[B \ge p] \int_0^1 \mathbb{P}[\lambda \le S \le p]\, \mathrm{d}\lambda \\
&= \mathbb{P}[S \le p] \int_p^1 \mathbb{P}[B \ge \lambda]\, \mathrm{d}\lambda + \mathbb{P}[B \ge p] \int_0^p \mathbb{P}[S \le \lambda]\, \mathrm{d}\lambda\,.
\end{aligned}$$

Now, we just need to address the Lipschitzness given that S and B admit densities bounded from above by some $M \ge 1$. Note that assumption implies that the cumulative distribution functions (CDFs) of S and B,

denoted by F_S and F_B, are M-Lipschitz. Let $0 \le p < q \le 1$, then

$$\begin{aligned}\left|\mathrm{GFT}(q,S,B) - \mathrm{GFT}(p,S,B)\right| = \Bigg| & F_S(p)\int_p^q \big(1-F_B(\lambda)\big)\,\mathrm{d}\lambda - \big(1-F_B(q)\big)\int_p^q F_S(\lambda)\,\mathrm{d}\lambda + \\ & \big(F_B(q)-F_B(p)\big)\int_0^p \big(1-F_B(\lambda)\big)\,\mathrm{d}\lambda - \big(F_S(q)-F_S(p)\big)\int_q^1 F_S(\lambda)\,\mathrm{d}\lambda\Bigg| \\ & \le M\left[2\int_p^q \mathrm{d}\lambda + 2(q-p)\right] \le 4M|q-p|\,,\end{aligned}$$

where in the inequality we used that F_S and F_B are upper bounded by 1 and that they are both M-Lipschitz. □

The previous result is also important since it relates the regularity of the distributions, specifically the boundedness of the densities, to the regularity of the objective function, i.e., its Lipschitzness.

4.1 Scouting Bandits (SB): from Realistic Feedback to Multi-Armed Bandits

The main challenge in designing a low-regret algorithm for this problem lies in the fact that posting a price does not reveal the corresponding gain from trade. This uncertainty then translates to a slow learning of the function $\mathbb{E}\big[\mathrm{GFT}(\cdot,S,B)\big]$ over the interval of possible prices.

This can be overcome by sampling. The structure of the gain from trade, however, is such that this sampling needs to be spread over the entire unit interval, i.e., to estimate $\mathbb{E}\big[\mathrm{GFT}(p,S,B)\big]$ for a given price p it is *not* sufficient to simply post p repeatedly. Eq. (5) helps visualizing this phenomenon. While the local terms, i.e., $\mathbb{P}[S \le p]$ and $\mathbb{P}[B \ge p]$, can be reconstructed by multiple posting of p, the integral terms exhibit a global behaviour: they depend on what happens in $(p,1]$ or $[0,p)$, and hence need prices to be posted in those regions to be estimated accurately. This rules out direct applications of well established algorithms, like action elimination or UCB Slivkins (2019), which crucially depend on the locality of the exploration. Similarly, estimating naively the CDFs on a grid of prices and using this information to reconstruct both the global and the local terms falls short of yielding the desired $T^{2/3}$ regret bound.

ALGORITHM 3: Scouting Bandits

input: bandit algorithm α, upper bound on the densities M, time horizon T, and precision parameter ε;
$\ell \leftarrow M^{2/3}$, $\delta \leftarrow \varepsilon M^{1/3}$, $K \leftarrow \lceil \ell/\varepsilon \rceil$, $T_0 \leftarrow \lceil \ln(4K/\delta)/(2\varepsilon^2) \rceil$;
$q_i \leftarrow i(\varepsilon/\ell)$, for all $i = 0,1,2,\dots,K-1$;
for $t = 1,2,\dots,T_0$ **do**
  draw P_t from $[0,1]$ uniformly at random;
  post price P_t and observe feedback $\big(\mathbb{I}\{S_t \le P_t\}, \mathbb{I}\{P_t \le B_t\}\big)$;
  let $\widehat{I}_i^t \leftarrow \mathbb{I}\{S_t \le P_t \le q_i\}$, and $\widehat{J}_i^t \leftarrow \widehat{\mathbb{I}}\{q_i \le P_t \le B_t\}$, for all $i = 0,1,\dots,K-1$
let $\widehat{I}_i \leftarrow \frac{1}{T_0}\sum_{t=1}^{T_0} \widehat{I}_i^t$ and $\widehat{J}_i \leftarrow \frac{1}{T_0}\sum_{t=1}^{T_0} \widehat{J}_i^t$, for all $i = 0,1,\dots,K-1$;
initialize the bandit algorithm α with the number of arms K and the horizon $T - T_0$;
for $t = T_0+1,\dots,T$ **do**
  receive an arm i_t from α;
  post price $P_t \leftarrow q_{i_t}$ and observe $(\mathbb{I}\{S_t \le P_t\}, \mathbb{I}\{P_t \le B_t\})$;
  feed to α reward $r_t = \mathbb{I}\{S_t \le P_t\}\widehat{J}_{i_t} + \mathbb{I}\{P_t \le B_t\}\widehat{I}_{i_t}$

Our Algorithm 3 (Scouting Bandits) consists of exploiting the decomposition in Eq. (5) and learns separately the global and local part of the gain from trade. First, a global exploration phase is run (Scouting), in which prices uniformly sampled in $[0,1]$ are posted and used to simultaneously estimate all the integral terms on a suitable grid. Once this is done, we can run any bandit algorithm (Bandits) on the prices of

the grid, complementing the realistic feedback received with the estimated integrals. We use the assumption on the independence of S and B (iv) to apply Lemma 4.1 and bounded densities (bd) to have Lipschitzness of the expected gain from trade. Later, we show how dropping either of these assumptions leads to linear regret (Theorems 4.3 and 4.4).

Theorem 4.1. *In the realistic-feedback stochastic (iid) setting where the distributions of the seller and buyer are independent (iv) and have densities bounded by $M \geq 1$, the regret of Scouting Bandits (SB) run with a bandit algorithm α, upper bound on the densities $M \geq 1$, time horizon T, and parameter $\varepsilon > 0$ satisfies*

$$R_T(\mathrm{SB}) = O\left(\frac{1}{\varepsilon^2}\ln\frac{M}{\varepsilon} + \varepsilon M^{1/3}T + \mathcal{R}\left(\left\lceil\frac{M^{2/3}}{\varepsilon}\right\rceil, T\right)\right),$$

where $\mathcal{R}(\kappa,\tau)$ is any monotone bound (in τ) on the regret of α on κ arms for time horizon τ. In particular, if α is either the Action Elimination or UCB algorithm Slivkins (2019), the resulting regret in $O\big(M^{1/3}T^{2/3}\ln(MT)\big)$.

We prove the result in two steps, first we show how the global exploration phase is indeed enough to build good estimates of the integrals for a suitable grid of prices, then we analyze the rest of the algorithm, conditioning on the event that such estimates are accurate.

Lemma 4.2. *Fix any precision $\varepsilon > 0$, probability $\delta > 0$, regularity term $\ell > 0$. Let $K = \lceil \ell/\varepsilon \rceil$ and fix the grid of prices $q_i = i(\varepsilon/\ell)$, for all $i = 0, 1, \ldots, K-1$. Moreover, for all $i = 0, 1, \ldots, K-1$,*

$$I_i = \int_0^{q_i} \mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda\,, \qquad J_i = \int_{q_i}^{1} \mathbb{P}[B \geq \lambda]\,\mathrm{d}\lambda\,.$$

Consider the estimators $\widehat{I}_i, \widehat{J}_i$ determined at the end of the exploration phase of Algorithm 3, i.e., $\widehat{I}_i = \frac{1}{T_0}\sum_{t=1}^{T_0}\widehat{I}_i^t$, and $\widehat{J}_i = \frac{1}{T_0}\sum_{t=1}^{T_0}\widehat{J}_i^t$. If $T_0 \geq \frac{1}{2\varepsilon^2}\ln\frac{4K}{\delta}$, it holds that

$$\max_{i=0,1,\ldots,K-1}\left\{|I_i - \widehat{I}_i|,\, |J_i - \widehat{J}_i|\right\} < \varepsilon,$$

with probability at least $1-\delta$, where the probability is with respect to $(S_1, B_1), \ldots, (S_{T_0}, B_{T_0})$.

Proof. We first show that the $\widehat{I}_i^t, \widehat{J}_i^t$ are unbiased estimators of I_i, J_i. For all $t \leq T_0$, conditioning on $P_t \leq q_i$, P_t follows a uniform distribution in $[0, q_i]$, hence, calling U_i a random variable uniformly distributed in that interval, one has:

$$\begin{aligned}\mathbb{E}\left[\widehat{I}_i^t\right] &= \mathbb{P}[S_t \leq P_t \leq q_i] = \mathbb{P}[P_t \leq q_i]\mathbb{P}[S_t \leq P_t \leq q_i | P_t \leq q_i] \\ &= \mathbb{P}[P_t \leq q_i]\mathbb{P}[S \leq U_i] = \int_0^{q_i}\mathbb{P}[S \leq \lambda]\,\mathrm{d}\lambda = I_i\,.\end{aligned}$$

A similar argument, conditioning on $P_t \geq q_i$, gives that $\mathbb{E}\left[\widehat{J}_i\right] = J_i$.

For all i, let $E_i = \left\{|\widehat{I}_i - I_i| > \varepsilon\right\}$ and $F_i = \left\{|\widehat{J}_i - J_i| > \varepsilon\right\}$ be the events in which there is an error greater that ε in the estimates. By the Chernoff-Hoeffding inequality one has that the probabilities of each event is upper bounded by $2e^{-2\varepsilon^2 T_0}$. Let $\mathcal{E}$ be the good event corresponding to all the integrals being estimated within an ε accuracy. Clearly $\mathcal{E}$ is the complement of $\bigcup_{i=0}^{K-1}(E_i \cup F_i)$. Hence, we have that:

$$\mathbb{P}[\mathcal{E}^c] \leq \sum_{i=0}^{K-1}\mathbb{P}[E_i] + \sum_{i=0}^{K-1}\mathbb{P}[F_i] \leq 4Ke^{-2\varepsilon^2 T_0} \leq \delta\,,$$

where the last inequality has been obtained by simply plugging in $T_0 \geq \frac{1}{2\varepsilon^2}\ln\frac{4K}{\delta}$. □

We are now ready to prove the main result of this section. After an initial global exploration phase, any optimal multi-armed bandit algorithm gives a regret for our bilateral trade problem that is optimal, up to logarithmic terms (by Theorem 4.2).

Proof of Theorem 4.1. Let $\varepsilon > 0$ be a precision parameter we set later, then consider the result of Lemma 4.2 on an initial exploration phase with $\delta \le \varepsilon M^{1/3}$ and $\ell = M^{2/3}$. Recall that, for all t, $\mathrm{GFT}_t(p) := \mathrm{GFT}(p, S_t, B_t)$. By the Lipschitzness of the gain from trade (with constant $4M$, as shown in Lemma 4.1) and the fact that the grid is spaced by $\frac{\varepsilon}{M^{2/3}}$, we get a discretization error at each time step which can be bounded as follows

$$\max_{p\in[0,1]} \mathbb{E}\big[\mathrm{GFT}(p,S,B)\big] - \max_{i=0,1,\dots,K-1} \mathbb{E}\big[\mathrm{GFT}(q_i,S,B)\big]$$
$$\le \mathbb{E}\big[\mathrm{GFT}(p^\star,S,B)\big] - \max_{i=i(p^\star),i(p^\star)+1} \mathbb{E}\big[\mathrm{GFT}(q_i,S,B)\big] \le 4M \min_{i=i(p^\star),i(p^\star)+1} |q_i - p^\star| \le 2\varepsilon M^{1/3}$$

where $q_{i(p^\star)}$ is the largest element in the grid smaller or equal to $p^\star$ and $q_{i(p^\star)+1}$ is set to 1 if $q_i(p^\star) = K-1$. Now, we have

$$R_T(\mathrm{SB}) = \max_{p\in[0,1]} \sum_{t=1}^{T} \mathbb{E}\big[\mathrm{GFT}_t(p) - \mathrm{GFT}_t(P_t)\big] \le T_0 + \sum_{t=T_0+1}^{T} \mathbb{E}\big[\mathrm{GFT}_t(p^\star) - \mathrm{GFT}_t(P_t)\big]$$
$$\le T_0 + 2\varepsilon M^{1/3} T + \max_{i=0,\dots,K-1} \mathbb{E}\left[\sum_{t=T_0+1}^{T} \mathrm{GFT}_t(q_i) - \sum_{t=T_0+1}^{T} \mathrm{GFT}_t(P_t)\right]$$

Let $i^\star \in \mathrm{argmax}_{i=0,\dots,K-1} \mathbb{E}\big[\sum_{t=T_0+1}^{T} \mathrm{GFT}_t(q_i)\big]$. Let also $\mathcal{E}$ be the same "good" event as in the proof of Lemma 4.2 for our choice of parameters. Conditioning with respect to it, we get

$$R_T(\mathrm{SB}) \le T_0 + 2\varepsilon M^{1/3} T + T\mathbb{P}\big[\mathcal{E}^c\big] + \sum_{t=T_0+1}^{T} \mathbb{E}\big[\mathrm{GFT}_t(q_{i^\star}) - \mathrm{GFT}_t(P_t)|\mathcal{E}\big]$$
$$\le T_0 + 2\varepsilon M^{1/3} T + T\delta + \sum_{t=T_0+1}^{T} \mathbb{E}\big[\mathrm{GFT}_t(q_{i^\star}) - \mathrm{GFT}_t(P_t)|\mathcal{E}\big] \,. \tag{6}$$

We now focus on the last term. Conditioning on $\mathcal{E}$, for all $t > T_0$ and all $i \in \{0,1,\dots,K-1\}$ we have that the expected gain of posting price q_i is ε-near to the expected reward for the multi-armed bandit instance associated. In fact on the one hand, by Lemma 4.1 and the fact that what (S_t, B_t) is independent from $\mathcal{E}$

$$\mathbb{E}[\mathrm{GFT}_t(q_i)|\mathcal{E}] = \mathbb{E}[\mathrm{GFT}_t(q_i)] = \mathbb{P}[S \le q_i] J_i + \mathbb{P}[q_i \le B] I_i.$$

On the other hand, defining $r_t(i) := \mathbb{I}\{S_t \le q_i\}\widehat{J}_i + \mathbb{I}\{q_i \le B_t\}\widehat{I}_i$ and conditioning on $\mathcal{E}$, we get

$$\mathbb{E}[r_t(i)|\mathcal{E}] = \mathbb{P}[S_t \le q_i]\mathbb{E}\big[\widehat{J}_i|\mathcal{E}\big] + \mathbb{P}[q_i \le B_t]\mathbb{E}\big[\widehat{I}_i|\mathcal{E}\big].$$

Putting those two formulae together, we have the claimed inequality

$$\left|\mathbb{E}\big[\mathrm{GFT}_t(q_i) - r_t(i)|\mathcal{E}\big]\right| \le \mathbb{P}[S_t \le q_i]\left|J_i - \mathbb{E}\big[\widehat{J}_i|\mathcal{E}\big]\right| + \mathbb{P}[q_i \le B_t]\left|I_i - \mathbb{E}\big[\widehat{I}_i|\mathcal{E}\big]\right| \le 2\varepsilon \,.$$

Plugging in this result, we get

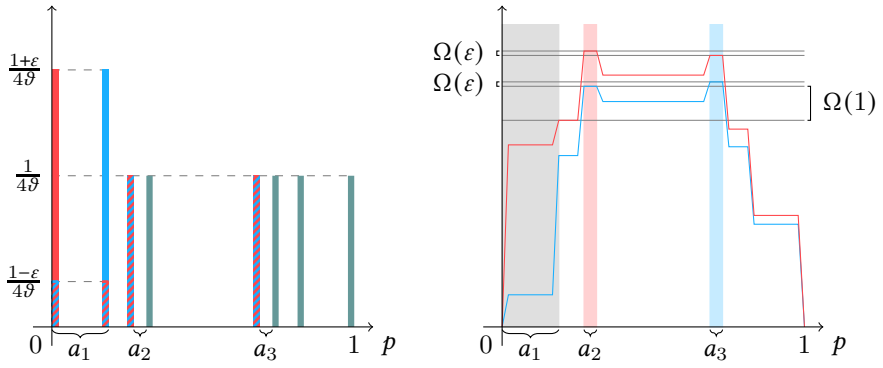
$$R_T(\mathrm{SB}) \le T_0 + T\delta + 2\varepsilon M^{1/3} T + 4\varepsilon T + \sum_{t=T_0+1}^{T} \mathbb{E}\big[r_t(i^\star) - r_t(i_t) \mid \mathcal{E}\big] \,.$$

We now focus on the last term. Note that, for any arm i, the sequence $r_{T_0+1}(i), r_{T_0+2}(i), \dots \in [0,2]$ is a $\mathbb{P}[\cdot \mid \mathcal{E}]$-i.i.d. sequence of random variables. Thus, we can exploit the worst-case regret guarantees of α and the monotonicity of the regret, obtaining

$$\sum_{t=T_0+1}^{T} \mathbb{E}\big[r_t(i^\star) - r_t(i_t) \mid \mathcal{E}\big] \le \mathcal{R}(K, T - T_0) \,.$$

Putting everything together, gives the first part of the result.

For the second part, pick $\varepsilon = T^{-1/3}$ and consider any algorithm (e.g, Action Elimination or UCB) with $\mathcal{R}(\kappa,\tau) = O\big(\sqrt{\kappa\tau\log\tau}\big)$. □



(a) Distributions $f_{S,\pm\varepsilon}$ (red/blue) and f_B (green)

(b) Expected GFT relative to $+\varepsilon$ (red) and $-\varepsilon$ (blue)

Figure 2: The only three regions where it makes sense for the learner to post prices are a_1, a_2, a_3. Prices in a_1 reveal information about the sign of $\pm\varepsilon$ suffering a $\Omega(1)$ regret; prices in a_2 are optimal if the distribution of the seller is the red one ($+\varepsilon$) but incur $\Omega(\varepsilon)$ regret if it is the blue one ($-\varepsilon$); the converse happens in a_3.

Note that Scouting Bandits needs to know in advance the time horizon T to set the length of the initial exploration phase and also to pass that information to the multi-armed bandit algorithm embedded, if needed. This dependence, however, can be lifted with a standard doubling trick Cesa-Bianchi and Lugosi (2006).

4.2 $T^{2/3}$ Lower Bound Under Realistic Feedback (iv+bd)

In this section, we show that the upper bound on the minimax regret we proved in Section 4.1 is tight, up to logarithmic factors. No strategy can beat the $T^{2/3}$ rate when the seller/buyer pair (S_t, B_t) is drawn i.i.d. from an unknown fixed distribution, even under the further assumptions that the valuations of the seller and buyer are independent of each other and have bounded densities. For a full proof of the following theorem, see Appendix D.

Theorem 4.2. *In the realistic-feedback model, for all horizons T, the minimax regret $R_T^\star$ satisfies*

$$R_T^\star := \inf_\alpha \sup_{(S,B)\sim\mathcal{D}} R_T(\alpha) \geq cT^{2/3}\,,$$

where $c \geq 11/672$, the infimum is over all of the learner's strategies α, and the supremum is over all distributions $\mathcal{D}$ of the seller S and buyer B such that:

(iid) *$(S_1, B_1), (S_2, B_2), \ldots \sim (S, B)$ is an i.i.d. sequence;*

(iv) *S and B are independent of each other;*

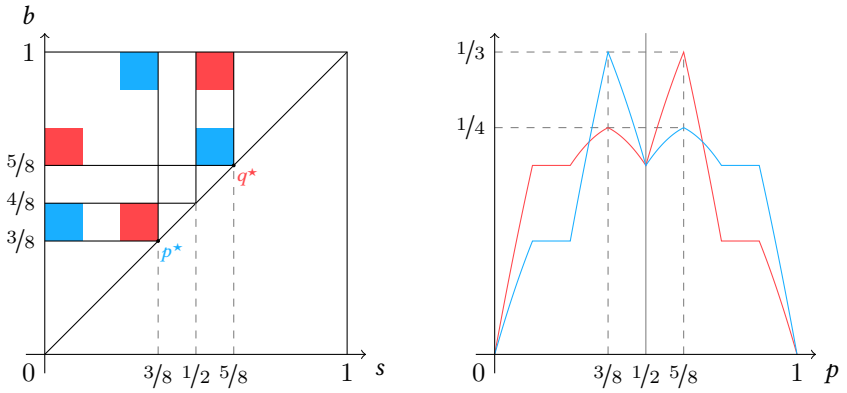
(bd) *S and B admit densities bounded by $M \geq 24$.*

Proof sketch. We build a family of distributions $\mathcal{D}_{\pm\varepsilon}$ of the seller and buyer (S, B) parameterized by $\varepsilon \in [0, 1]$. For the seller, for any $\varepsilon \in [0, 1]$, we define the density

$$f_{S,\pm\varepsilon} := \frac{1}{4\vartheta}\Big((1\pm\varepsilon)\mathbb{I}_{[0,\vartheta]} + (1\mp\varepsilon)\mathbb{I}_{\left[\frac{1}{6},\frac{1}{6}+\vartheta\right]} + \mathbb{I}_{\left[\frac{1}{4},\frac{1}{4}+\vartheta\right]} + \mathbb{I}_{\left[\frac{2}{3},\frac{2}{3}+\vartheta\right]}\Big)\,, \qquad \text{(Fig. 2(a), in red/blue)}$$

where $\vartheta := 1/48$ is a normalization constant. For the buyer, we define a single density (independently of ε)

$$f_B := \frac{1}{4\vartheta}\Big(\mathbb{I}_{\left[\frac{1}{3}-\vartheta,\frac{1}{3}\right]} + \mathbb{I}_{\left[\frac{3}{4}-\vartheta,\frac{3}{4}\right]} + \mathbb{I}_{\left[\frac{5}{6}-\vartheta,\frac{5}{6}\right]} + \mathbb{I}_{[1-\vartheta,1]}\Big)\,. \qquad \text{(Fig. 2(a), in green)}$$



(a) Supports of distributions f (blue) and g (red)

(b) Expected GFT relative to f (blue) and g (red)

Figure 3: Under realistic feedback, the two densities f and g are indistinguishable. The optimal price $p^\star$ for f gives constant regret under g and $q^\star$ does the converse.

In the $+\varepsilon$ (resp., $-\varepsilon$) case, the optimal price belongs to a region a_2 (resp., a_3, see Fig. 2(b)). By posting prices in the wrong region a_3 (resp., a_2) in the $+\varepsilon$ (resp., $-\varepsilon$) case, the learner incurs $\Omega(\varepsilon)$ regret. Thus, if ε is bounded-away from zero, the only way to avoid suffering linear regret is to identify the sign of $\pm\varepsilon$ and play accordingly. Clearly, the feedback received from the buyer gives no information on $\pm\varepsilon$. Since the feedback received from the seller at time t by posting a price p is $\mathbb{I}\{S_t \le p\}$, one can obtain information about (the sign of) $\pm\varepsilon$ only by posting prices in the costly ($\Omega(1)$-regret) sub-optimal region a_1.

This closely resembles the learning dilemma present in the so-called *revealing action* partial monitoring game Cesa-Bianchi and Lugosi (2006). In fact, a technical proof (see Appendix D), shows that our setting is harder (i.e., it has a higher minimax regret) than an instance of a revealing action problem, which has a known lower bound on its minimax regret of $\frac{11}{96}\left(\frac{1}{7}T^{2/3}\right)$ Cesa-Bianchi et al. (2006). □

4.3 Linear Lower Bound Under Realistic Feedback (bd)

In this section, we show that no strategy that can achieve worst-case sublinear regret when the seller/buyer pair (S_t, B_t) is drawn i.i.d. from an unknown fixed distribution, even under the further assumption that the valuations of the seller and buyer have bounded densities. This is due to a lack of observability. For a full proof of the following theorem, see Appendix E.

Theorem 4.3. *In the realistic-feedback model, for all horizons T, the minimax regret $R_T^\star$ satisfies*

$$R_T^\star := \inf_{\alpha} \sup_{(S,B)\sim\mathcal{D}} R_T(\alpha) \ge cT\,,$$

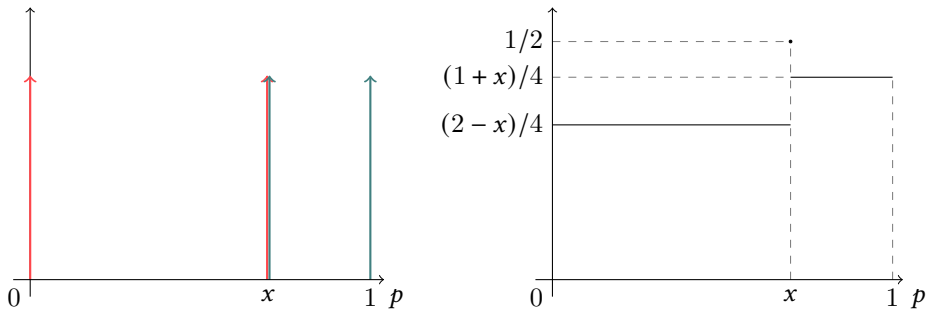
where $c \ge 1/24$, the infimum is over all of the learner's strategies α, and the supremum is over all distributions $\mathcal{D}$ of the seller S and buyer B such that:

(iid) $(S_1, B_1), (S_2, B_2), \ldots \sim (S, B)$ *is an i.i.d. sequence;*

(bd) *S and B admit densities bounded by $M \ge 24$.*

Proof sketch. Consider the two joint densities f and g of the seller and buyer (S, B) as the normalized indicator functions of the red and blue squares in Fig. 3(a). Formally

$$f = \frac{64}{3}\left(\mathbb{I}_{[0/8,\,1/8]\times[3/8,\,4/8]} + \mathbb{I}_{[2/8,\,3/8]\times[7/8,\,8/8]} + \mathbb{I}_{[4/8,\,5/8]\times[5/8,\,6/8]}\right)$$



(a) Distribution of S^x (red) and B^x (green) (b) Expected gain from trade relative to S^x and B^x

Figure 4: All prices, except for x, have high regret. However, under realistic feedback, finding x in a finite time is as harder than finding a needle in a haystack.

and $g(s,b) = f(1-b, 1-s)$. In the f (resp., g) case, the optimal price belongs to the region $[0, 1/2]$ (resp., $(1/2, 1]$, see Fig. 3(b)). By posting prices in the wrong region $(1/2, 1]$ (resp., $[0, 1/2]$) in the f (resp., g) case, the learner incurs at least a $1/3 - 1/4 = 1/12$ regret. Thus, the only way to avoid suffering linear regret is to determine if the valuations of the seller and buyer are generated by f or g. For each price $p \in [0,1]$, consider the four rectangles with opposite vertices (p,p) and (u_i, v_i), where $\{(u_i, v_i)\}_{i=1,\dots,4}$ are the four vertices of the unit square. Note that the only information on the distribution of (S,B) that the learner can gather from the realistic feedback $\big(\mathbb{I}\{S_t \le p\}, \mathbb{I}\{p \le B_t\}\big)$ received after posting a price p is (an estimate of) the area of the portion of the support of the distribution included in each of these four rectangles. However, these areas coincide in the cases f and g. Hence, in under realistic feedback, f and g are completely indistinguishable. Therefore, given that the optimal price in the f (resp., g) case is $3/8$ (resp., $5/8$), the best that the learner can do is to sample prices uniformly at random in the set $\{3/8, 5/8\}$, incurring a regret of $T/24$. For a formalization of this argument that leverages the techniques we described in the introduction, see Appendix E. □

4.4 Linear Lower Bound Under Realistic Feedback (iv)

In this section, we prove that in the realistic-feedback case, no strategy can achieve sublinear regret without any limitations on how concentrated the distributions of the valuations of the seller and buyer are, not even if they are independent of each other (iv).

At a high level, if the two distributions are of the seller and the buyer are very concentrated in a small region, finding an optimal price is like finding a needle in a haystack. For a full proof of the following theorem, see Appendix F.

Theorem 4.4. *In the realistic-feedback model, for all horizons T, the minimax regret $R_T^\star$ satisfies*

$$R_T^\star := \inf_{\alpha} \sup_{(S,B)\sim\mathcal{D}} R_T(\alpha) \ge cT\,,$$

where $c \ge 1/8$, the infimum is over all of the learner's strategies α, and the supremum is over all distributions $\mathcal{D}$ of the seller S and buyer B such that:

(iid) $(S_1,B_1),(S_2,B_2),\dots \sim (S,B)$ *is an i.i.d. sequence;*

(iv) *S and B are independent of each other.*

Proof sketch. Consider a family of seller/buyer distributions (S^x, B^x), parameterized by $x \in I$, where I is a small interval centered in $1/2$, S^x and B^x are independent of each other, and they satisfy

$$S^x = \begin{cases} x & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2} \end{cases}, \qquad B^x = \begin{cases} x & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2} \end{cases}.$$

The distributions and the corresponding gain from trade are represented in Fig. 4(a) and Fig. 4(b), respectively. A direct verification shows that the best fixed price with respect to (S^x, B^x) is $p = x$. Furthermore, by posting any other prices, the learner incurs a regret of approximately $1/2$ with probability $1/4$. Since it is intuitively clear that no strategy can locate (exactly!) each possible $x \in I$ in a finite number of steps. This results, for any strategy, in regret of at least (approximately) $T/8$. For the interested reader, a more detailed analysis is carried over in the Appendix F. □

5 Adversarial Setting: Linear Lower Bound Under Full Feedback

In this section, we prove that even in the simpler full-feedback case, no strategy can achieve worst-case sublinear regret in an adversarial setting. Lower bounds for the adversarial setting have a slightly different structure that the stochastic ones. The idea of the proof is to build, for any strategy, an *hard* sequence of sellers and buyers' valuations $(s_1, b_1), (s_2, b_2), \dots$ which causes the algorithm to suffer linear regret for any horizon T.

Theorem 5.1. *In the full-feedback adversarial (adv) setting, for all horizons $T \in \mathbb{N}$, the minimax regret $R_T^\star$ satisfies*

$$R_T^\star := \inf_\alpha \sup_{(s_1,b_1),(s_2,b_2),\dots} R_T(\alpha) \geq cT \,,$$

where $c \geq 1/4$, the infimum is over all of the learner's strategies α, and the supremum is over all deterministic sequences $(s_1, b_1), (s_2, b_2), \dots \in [0,1]^2$ of the seller and buyer's valuations.

Proof. Let $\varepsilon \in (0, \frac{1}{18})$. We begin by fixing any strategy α of the learner. This is a sequence of functions α_t mapping the past feedback $(s_1, b_1), \dots, (s_{t-1}, b_{t-1})$, together with some internal randomization, to the price P_t to be posted by the learner at time t. In other words, the strategy maintains a distribution ν_t over the prices that is updated after observing each new pair (s_t, b_t) and used to draw each new price P_t. We will show how to constructively determine a sequence of seller/buyer valuations that is hard for α to learn. This sequence is oblivious to the prices $P_1, P_2, \dots$ posted by α, in the sense it does not have access to the realizations of its internal randomization. The idea is, at any time t, to determine a seller/buyer pair (s_t, b_t) either of the form $(c_t, 1)$ or $(0, d_t)$, with $c_t \approx \frac{1}{2} \approx d_t$, such that the probability ν_t that the strategy picks a price $P_t \in [s_t, b_t]$ (i.e., that there is a trade) is at most $1/2$ and, at the same time, there is common price $p^\star$ which belongs to $[s_t, b_t]$ for all times t. This way, since $b_t - s_t \approx \frac{1}{2}$ for all t, the regret of α with respect to $(s_1, b_1), (s_2, b_2), \dots$ is at least (approximately) greater than or equal to $T/4$.

The formal construction proceeds inductively as follows. Let

$$\begin{cases} c_1 := \frac{1}{2} - \frac{3}{2}\varepsilon,\ d_1 := \frac{1}{2} - \frac{1}{2}\varepsilon,\ s_1 := 0,\ b_1 := d_1, & \text{if } \nu_1\big[\big[0, \frac{1}{2} - \frac{1}{2}\varepsilon\big]\big] \leq \frac{1}{2}\,, \\ c_1 := \frac{1}{2} + \frac{1}{2}\varepsilon,\ d_1 := \frac{1}{2} + \frac{3}{2}\varepsilon,\ s_1 := c_1,\ b_1 := 1, & \text{otherwise.} \end{cases}$$

Then, for any time t, given that c_i, d_i, s_i, b_i are defined for all $i \leq t$ and recalling that ν_{t+1} is the distribution over the prices at time $t+1$ (of the strategy α after observing the feedback $(s_1, b_1), \dots, (s_t, b_t)$), let

$$\begin{cases} c_{t+1} := c_t,\ d_{t+1} := d_t - \frac{2\varepsilon}{3^t},\ s_{t+1} := 0,\ b_{t+1} := d_{t+1}, & \text{if } \nu_{t+1}\big[\big[0, c_t + \frac{\varepsilon}{3^t}\big]\big] \leq \frac{1}{2}\,, \\ c_{t+1} := c_t + \frac{2\varepsilon}{3^t},\ d_{t+1} := d_t,\ s_{t+1} := c_{t+1},\ b_{t+1} := 1, & \text{otherwise.} \end{cases}$$

Then the sequence of seller/buyer valuations $(s_1, b_1), (s_2, b_2), \dots$ defined above by induction satisfies:

- $\nu_t\big[[s_t, b_t]\big] \leq \frac{1}{2}$, for each time t;
- there exists $p^\star \in [0,1]$ such that $p^\star \in [s_t, b_t]$, for each time t (e.g. $p^\star := \lim_{t\to\infty} c_t$);
- $b_t - s_t \geq \frac{1-3\varepsilon}{2}$, for each time t.

This implies, for any horizon T,

$$R_T(\alpha) = \sum_{t=1}^{T} \mathrm{GFT}(p^\star, s_t, b_t) - \sum_{t=1}^{T} \mathbb{E}[\mathrm{GFT}(P_t, s_t, b_t)] \geq \sum_{t=1}^{T} (b_t - s_t)\big(1 - \nu_t\big[[s_t, b_t]\big]\big) \geq \frac{1-3\varepsilon}{4} T \ .$$

Since ε and α are arbitrary, this yields immediately $R^\star \geq T/4$. □

6 Conclusions

This work initiates the study of the bilateral trade problem in a regret minimization framework. We prove tight bounds on the regret rates that can be achieved under various feedback and private valuation models.

Our work opens several possibilities of future investigation. One first and natural research direction is related to the more general settings of two-sided markets with multiple buyers and sellers, different prior distributions, and complex valuation functions.

A second direction is related to the tight characterization of the regret rates for weak budget balance mechanisms. These can be proved to be strictly better than strong budget balance mechanisms, at least for the realistic feedback setting with correlated distributions (details will appear in the full version of this work).

Finally, we believe other classes of markets, which assume prior knowledge of the agent's preferences, could be fruitfully studied in a regret minimization framework.

References

Shipra Agrawal and Nikhil R. Devanur. 2014. Bandits with concave rewards and convex knapsacks. In *ACM Conference on Economics and Computation, EC'14*. ACM, Association for Computing Machinery, New York, NY, USA, 989–1006.

Kareem Amin, Afshin Rostamizadeh, and Umar Syed. 2013. Learning prices for repeated auctions with strategic buyers. In *Advances in Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, 1169–1177.

Moshe Babaioff, Shaddin Dughmi, Robert Kleinberg, and Aleksandrs Slivkins. 2015. Dynamic pricing with limited supply. *ACM Transactions on Economics and Computation (TEAC)* 3, 1 (2015), 4.

Ashwinkumar Badanidiyuru, Robert Kleinberg, and Aleksandrs Slivkins. 2018. Bandits with Knapsacks. *J. ACM* 65, 3 (2018), 13:1–13:55.

Gábor Bartók, Dean P Foster, Dávid Pál, Alexander Rakhlin, and Csaba Szepesvári. 2014. Partial monitoring—classification, regret bounds, and algorithms. *Mathematics of Operations Research* 39, 4 (2014), 967–997.

Avrim Blum and Jason D. Hartline. 2005. Near-optimal online auctions. In *ACM-SIAM Symposium on Discrete Algorithms, SODA'05*. Society for Industrial and Applied Mathematics, Society for Industrial and Applied Mathematics, USA, 1156–1163.

Avrim Blum, Vijay Kumar, Atri Rudra, and Felix Wu. 2004. Online learning in online auctions. *Theoretical Computer Science* 324, 2-3 (2004), 137–146.

Liad Blumrosen and Shahar Dobzinski. 2014. Reallocation Mechanisms. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation (EC '14)*. Association for Computing Machinery, New York, NY, USA, 617.

Liad Blumrosen and Shahar Dobzinski. 2016. (Almost) Efficient Mechanisms for Bilateral Trading. *CoRR* abs/1604.04876 (2016), 1–16.

Josef Broder and Paat Rusmevichientong. 2012. Dynamic pricing under a general parametric choice model. *Operations Research* 60, 4 (2012), 965–980.

Johannes Brustle, Yang Cai, Fa Wu, and Mingfei Zhao. 2017. Approximating Gains from Trade in Two-sided Markets via Simple Mechanisms. In *ACM Conference on Economics and Computation, EC'17*. ACM, New York, NY, USA, 589–590.

Sébastien Bubeck and Nicolò Cesa-Bianchi. 2012. Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *Found. Trends Mach. Learn.* 5, 1 (2012), 1–122.

Sebastien Bubeck, Nikhil R. Devanur, Zhiyi Huang, and Rad Niazadeh. 2017. Online Auctions and Multi-Scale Online Learning. In *Proceedings of the 2017 ACM Conference on Economics and Computation* (Cambridge, Massachusetts, USA) *(EC '17)*. Association for Computing Machinery, New York, NY, USA, 497–514.

Nicolò Cesa-Bianchi, Tommaso Cesari, and Vianney Perchet. 2019. Dynamic Pricing with Finitely Many Unknown Valuations. In *Proceedings of the 30th International Conference on Algorithmic Learning Theory (Proceedings of Machine Learning Research, Vol. 98)*, Aurélien Garivier and Satyen Kale (Eds.). PMLR, Chicago, Illinois, 247–273.

Nicolò Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. 2015. Regret Minimization for Reserve Prices in Second-Price Auctions. *IEEE Trans. Inf. Theory* 61, 1 (2015), 549–564.

Nicolò Cesa-Bianchi and Gábor Lugosi. 2006. *Prediction, learning, and games*. Cambridge University Press, UK.

Nicolo Cesa-Bianchi, Gábor Lugosi, and Gilles Stoltz. 2006. Regret minimization under partial monitoring. *Mathematics of Operations Research* 31, 3 (2006), 562–580.

Maxime C. Cohen, Ilan Lobel, and Renato Paes Leme. 2020. Feature-Based Dynamic Pricing. *Manag. Sci.* 66, 11 (2020), 4921–4943.

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.

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.

Constantinos Daskalakis and Vasilis Syrgkanis. 2016. Learning in Auctions: Regret is Hard, Envy is Easy. In *IEEE Symposium on Foundations of Computer Science FOCS'16*. IEEE Computer Society, USA, 219–228.

Arnoud V. den Boer. 2015. Dynamic pricing and learning: historical origins, current research, and new directions. *Surveys in operations research and management science* 20, 1 (2015), 1–18.

Arnoud V den Boer and N Bora Keskin. 2020. Discontinuous demand functions: estimation and pricing. *Management Science* 66, 10 (2020), 4516–4534.

Nikhil R. Devanur, Yuval Peres, and Balasubramanian Sivan. 2019. Perfect Bayesian Equilibria in repeated sales. *Games Econ. Behav.* 118 (2019), 570–588.

Alexey Drutsa. 2018. Weakly consistent optimal pricing algorithms in repeated posted-price auctions with strategic buyer. In *International Conference on Machine Learning, ICML'18 (Proceedings of Machine Learning Research, Vol. 80)*. PMLR, USA, 1318–1327.

Paul Dütting, Federico Fusco, Philip Lazos, Stefano Leonardi, and Rebecca Reiffenhäuser. 2021. Efficient Two-Sided Markets with Limited Information. Preprint on authors' webpage https://www.plazos.me/pdfs/2sided_xos.pdf.

Robert D. Kleinberg and Frank Thomson Leighton. 2003. The Value of Knowing a Demand Curve: Bounds on Regret for Online Posted-Price Auctions. In *IEEE Symposium on Foundations of Computer Science FOCS'13*. IEEE Computer Society, USA, 594–605.

Tor Lattimore and Csaba Szepesvári. 2020. *Bandit algorithms*. Cambridge University Press, UK.

Thodoris Lykouris, Vasilis Syrgkanis, and Éva Tardos. 2016. Learning and Efficiency in Games with Dynamic Population. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '16)*. Society for Industrial and Applied Mathematics, USA, 120–129.

Michael Mitzenmacher and Eli Upfal. 2017. *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis, Second Edition*. Cambridge university press, UK.

Mehryar Mohri and Andres Munoz Medina. 2014. Optimal Regret Minimization in Posted-Price Auctions with Strategic Buyers. In *Advances in Neural Information Processing Systems*. NeurIPS, USA, 1871–1879.

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

Aleksandrs Slivkins. 2019. Introduction to Multi-Armed Bandits. *Foundations and Trends® in Machine Learning* 12, 1-2 (2019), 1–286. https://doi.org/10.1561/2200000068

Aleksandrs Slivkins and Assaf Zeevi. 2015. Dynamic Pricing Under Model Uncertainty. Tutorial given at the 16th ACM Conference on Economics and Computation.

William Vickrey. 1961. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance* 16, 1 (1961), 8–37.

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

A Model and Notation

For all $T \in \mathbb{N}$, we denote the set of the first T integers $\{1, \dots, T\}$ by $[T]$. If $\mathbb{P}$ is a probability measure and X is a random variable, we denote by $\mathbb{P}_X$ the probability measure defined for any (measurable) set E, by $\mathbb{P}_X[E] := \mathbb{P}[X \in E]$. We denote the expectation of a random variable X with respect to probability measure $\mathbb{P}$ by $\mathbb{E}_{\mathbb{P}}[X]$. If a measure ν is absolutely continuous with respect another measure μ with density f, we denote ν by $f\mu$, so that for any (measurable) set E, $(f\mu)[E] := \nu[E] = \int_E f(x)\,\mathrm{d}\mu(x)$. We denote the Lebesgue measure on the interval $[0,1]$ by μ_L and the product Lebesgue measure on $[0,1]^{\mathbb{N}}$ by $\boldsymbol{\mu}_L$. For any set E and $x \in E$, we denote the Dirac measure on x by δ_x (the dependence on E will always be clear from context).

A.1 The Learning Model

In this section, we introduce an abstract notion of sequential games which encompasses all the settings we discussed in the main part of the paper, providing a unified perspective. This will be especially useful when proving lower bounds.

Definition 1 (Sequential game)**.** A *(sequential) game* is a tuple $\mathscr{G} := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P})$, where:

- $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ are sets called the *player's action space*, *adversary's action space*, and *feedback space*;
- $\rho\colon \mathcal{X} \times \mathcal{Y} \to [0,1]$ and $\varphi\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ are called the *reward* and *feedback* functions[2];
- $\mathscr{P}$ is a set of probabilities on the set $\mathcal{Y}^{\mathbb{N}}$ of sequences in $\mathcal{Y}$, called the *adversary's behavior*.

This definition generalizes the partial monitoring games of (Lattimore and Szepesvári, 2020; Bartók et al., 2014) to settings with infinitely many arms and is able to model adversarial, i.i.d., and more general stochastic settings all at once. Before proceeding, we introduce another few extra handy definitions that will be used throughout the paper.

Definition 2. If $\mathscr{G} = (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P})$ is a game, then we say the following. *The sample space* is the set $\Omega := \mathcal{Y}^{\mathbb{N}} \times [0,1]^{\mathbb{N}}$. *The adversary's actions* $\left(Y_t\right)_{t\in\mathbb{N}}$ and *the player's randomization* $\left(U_t\right)_{t\in\mathbb{N}}$ are sequences of random variables defined, for all $t \in \mathbb{N}$ and $\omega = \left((y_n)_{n\in\mathbb{N}}, (u_n)_{n\in\mathbb{N}}\right) \in \Omega$, by $Y_t(\omega) := y_t$ and $U_t(\omega) := u_t$. *The set of scenarios* $\mathcal{S}$ is the set of probability measures $\mathbb{P}$ on Ω of the form $\mathbb{P} = \boldsymbol{\mu} \otimes \boldsymbol{\mu}_L$, where $\boldsymbol{\mu} \in \mathscr{P}$.

For the sake of conciseness, whenever we fix a game $\mathscr{G}$, we will assume that all the objects (sets, functions, random variables) presented in Definitions 1–2 are fixed and denoted by the same letters without declaring them explicitly each time, unless strictly needed.

Note that this setting models an *oblivious* adversary since its actions are independent of the player's past randomization, i.e., for all $t \in \mathbb{N}$, $\mathbb{P}_{Y_{t+1}|Y_1,\dots,Y_t,U_1,\dots,U_t} = \mathbb{P}_{Y_{t+1}|Y_1,\dots,Y_t}$. Note also that we are assuming that the randomization of the player's strategy is carried out by drawing numbers in the interval $[0,1]$ independently and uniformly at random. We can restrict ourselves to this case in light of the Skorokhod Representation Theorem (Williams, 1991, Section 17.3) without losing (much) generality. We now introduce formally the strategies of the player, the resulting played actions, and the corresponding feedback.

Definition 3 (Player's strategies, actions, and feedback)**.** Given a game $\mathscr{G}$, we define a *player's strategy* as a sequence of functions $\alpha = (\alpha_t)_{t\in\mathbb{N}}$ such that, for each $t \in \mathbb{N}$, $\alpha_t\colon [0,1]^t \times \mathcal{Z}^{t-1} \to \mathcal{X}$.[3] Given a player's strategy α, we define inductively (on t) the corresponding sequences of *player's actions* $(X_t)_{t\in\mathbb{N}}$ and *player's feedback* $(Z_t)_{t\in\mathbb{N}}$ by $X_t := \alpha_t(U_1, \dots, U_t, Z_1, \dots, Z_{t-1})$, $Z_t := \varphi(X_t, Y_t)$. In the sequel, we will denote the set of all strategies for a game $\mathscr{G}$ by $\mathscr{A}(\mathscr{G})$.

To lighten the notation, we will write $\mathscr{A}$ instead of $\mathscr{A}(\mathscr{G})$ if it is clear from context. We can now extend the standard notions of regret, worst-case regret, and minimax regret to our general setting.

[2] More precisely, we need $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ to be non-empty measurable spaces and ρ, φ to be measurable functions. To avoid clutter, in the following we will never mention explicitly these types of standard measurability assumptions unless strictly needed.

[3] When $t = 1$, $[0,1]^t \times \mathcal{Z}^{t-1} := [0,1]$. In the following, we will always adopt this type of convention without mention it.

Definition 4 (Regret)**.** Given a game $\mathscr{G}$ and a horizon $T \in \mathbb{N}$, we define the *regret* (of $\alpha \in \mathscr{A}$ in scenario $\mathbb{P} \in \mathcal{S}$), the *worst-case regret* (of $\alpha \in \mathscr{A}$), and the *minimax regret* (of $\mathscr{G}$), respectively, by

$$R_T^{\mathbb{P}}(\alpha) := \sup_{x \in \mathcal{X}} \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{T} \rho(x, Y_t) - \sum_{t=1}^{T} \rho(X_t, Y_t)\right], \quad R_T^{\mathcal{S}}(\alpha) := \sup_{\mathbb{P} \in \mathcal{S}} R_T^{\mathbb{P}}(\alpha), \quad R_T^{\star}(\mathscr{G}) := \inf_{\alpha \in \mathscr{A}(\mathscr{G})} R_T(\alpha).$$

Informally, if $\mathscr{G}$ and $\widetilde{\mathscr{G}}$ are two games and $R_T^{\star}(\mathscr{G}) \geq R_T^{\star}(\widetilde{\mathscr{G}})$, we say that $\widetilde{\mathscr{G}}$ is *easier* than $\mathscr{G}$ (or equivalently, that $\mathscr{G}$ is *harder* than $\widetilde{\mathscr{G}}$). When it is clear from the context, we will omit the dependence on $\mathscr{G}$ in $R_T^{\star}(\mathscr{G})$.

A.2 Bilateral Trade as a Game

We now formally cast the various instances of bilateral trade we introduced in Section 2 into our sequential game setting. In this context, we think of the learner as the *player* and the environment as the *adversary*.

A.2.1 Player's Actions, Adversary's Actions, and Reward

The player's action space $\mathcal{X}$ is the unit interval $[0,1]$. This corresponds to the player posting the same price price to both the seller and the buyer (strong budget balance). The adversary's action space $\mathcal{Y}$ is $[0,1]^2$. They are the pairs of valuations of the seller and buyer. The reward function ρ is the gain from trade $\mathrm{GFT}\colon [0,1] \times [0,1]^2 \to [0,1]$, $\big(p, (s,b)\big) \mapsto (b-s)\mathbb{I}\{s \leq p \leq b\}$.

A.2.2 Available Feedback

Realistic the feedback space $\mathcal{Z}$ is the boolean square $\{0,1\}^2$ and the feedback function is $\varphi\colon [0,1] \times [0,1]^2 \to \{0,1\}^2$, $\big(p, (s,b)\big) \mapsto \big(\mathbb{I}\{s \leq p\}, \mathbb{I}\{p \leq b\}\big)$. This corresponds to the seller and the buyer accepting or rejecting a trade at a price p.

Full the feedback space $\mathcal{Z}$ is the unit square $[0,1]^2$ and the feedback function is $\varphi\colon [0,1] \times [0,1]^2 \to [0,1]^2$, $\big(p, (s,b)\big) \mapsto (s,b)$. This corresponds to the seller and the buyer revealing their valuations at the end of a trade.

A.2.3 Adversary's Behavior

Stochastic (iid): the adversary's behavior $\mathscr{P} = \mathscr{P}_{\mathrm{iid}}$ consists of products of a single probability on $\mathcal{Y} = [0,1]^2$, i.e., $\boldsymbol{\mu} \in \mathscr{P}_{\mathrm{iid}}$ if and only if there exists a probability measure μ on $[0,1]^2$ such that $\boldsymbol{\mu} = \otimes_{t \in \mathbb{N}} \mu$. This corresponds to a stochastic i.i.d. environment, where however the valuations of the seller and the buyer could be correlated.

We will also investigate the following stronger assumptions.

- **Independent valuations (iv)** the adversary's behavior $\mathscr{P} = \mathscr{P}_{\mathrm{iv}}$ is the subset of $\mathscr{P}_{\mathrm{iid}}$ in which the valuations of the seller and the buyer are independent, i.e., $\boldsymbol{\mu} \in \mathscr{P}_{\mathrm{iv}}$ if and only if there exist two probability measures μ_S, μ_B on $[0,1]$ such that $\boldsymbol{\mu} = \otimes_{t \in \mathbb{N}} (\mu_S \otimes \mu_B)$.
- **Bounded density (bd)** for a fixed $M \geq 1$, the adversary's behavior $\mathscr{P} = \mathscr{P}_{\mathrm{bd}}^M$ is the subset of $\mathscr{P}_{\mathrm{iid}}$ in which the joint distribution of the valuations of buyer and seller has a density bounded by M, i.e., $\boldsymbol{\mu} \in \mathscr{P}_{\mathrm{bd}}^M$ if and only if there exists a density $f\colon [0,1]^2 \to [0,M]$ such that $\boldsymbol{\mu} = \otimes_{t \in \mathbb{N}} (f\mu)$, where $\mu = \mu_L \otimes \mu_L$.
- **Independent valuations with bounded density (iv+bd)** for a fixed $M \geq 1$, the adversary's behavior $\mathscr{P} = \mathscr{P}_{\mathrm{iv+bd}}^M$ is the subset $\mathscr{P}_{\mathrm{iv}} \cap \mathscr{P}_{\mathrm{bd}}^M$ of $\mathscr{P}_{\mathrm{iid}}$.

Adversarial (adv): the adversary's behavior $\mathscr{P} = \mathscr{P}_{\mathrm{adv}}$ consists of products of Dirac measures on $\mathcal{Y} = [0,1]^2$, i.e., $\boldsymbol{\mu} \in \mathscr{P}_{\mathrm{adv}}$ if and only if there exists a sequence $(s_t, b_t)_{t \in \mathbb{N}} \subset [0,1]^2$ such that $\boldsymbol{\mu} = \otimes_{t \in \mathbb{N}} \delta_{(s_t, b_t)}$. This corresponds to a deterministic, oblivious, and adversarial environment (Section 5).

B Two Key Lemmas on Simplifying Sequential Games

In this section we introduce some useful techniques that could be of independent interest for proving lower bounds in sequential games. The idea is to give sufficient conditions for given game to be harder than another, where the second one has a known lower bound on its minimax regret.

At a high level, the first lemma shows that if the adversary's actions are independent of each other, a game $\tilde{\mathscr{G}}$ is easier than game $\mathscr{G}$ if $\tilde{\mathscr{G}}$ can be embedded in $\mathscr{G}$ in such a way that the optimal player's actions of $\tilde{\mathscr{G}}$ are no better than the ones in $\mathscr{G}$, the suboptimal player's actions of $\tilde{\mathscr{G}}$ no worse than the ones in $\mathscr{G}$, and at distributional level, the quality of the feedback does not decrease in the second game. The proof is deferred to Appendix B.1.

Lemma B.1 (Embedding)**.** *Let $\mathscr{G} := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P})$ and $\tilde{\mathscr{G}} := (\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}, \tilde{\mathcal{Z}}, \tilde{\rho}, \tilde{\varphi}, \tilde{\mathscr{P}})$ be two games, $\mathcal{S}, \tilde{\mathcal{S}}$ their respective sets of scenarios, $(Y_t)_{t\in\mathbb{N}}, (\tilde{Y}_t)_{t\in\mathbb{N}}$ their adversaries' actions, and $T \in \mathbb{N}$ a horizon. Assume that $Y_1, \dots, Y_T$ are $\mathbb{P}$-independent for any scenario $\mathbb{P} \in \mathcal{S}$, $\tilde{Y}_1, \dots, \tilde{Y}_T$ are $\tilde{\mathbb{P}}$-independent for any scenario $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$, and that there exist $\tilde{f}\colon \mathcal{X} \to \tilde{\mathcal{X}}$, $g\colon \tilde{\mathcal{Z}} \to \mathcal{Z}$, and $\hbar\colon \tilde{\mathcal{S}} \to \mathcal{S}$ satisfying:*

1. $\sup_{\tilde{x}\in\tilde{\mathcal{X}}} \sum_{t=1}^{T} \mathbb{E}_{\tilde{\mathbb{P}}}\big[\tilde{\rho}(\tilde{x}, \tilde{Y}_t)\big] \le \sup_{x\in\mathcal{X}} \sum_{t=1}^{T} \mathbb{E}_{\hbar(\tilde{\mathbb{P}})}\big[\rho(x, Y_t)\big]$ *for any scenario $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$;*
2. $\mathbb{E}_{\tilde{\mathbb{P}}}\big[\tilde{\rho}\big(\tilde{f}(x), \tilde{Y}_t\big)\big] \ge \mathbb{E}_{\hbar(\tilde{\mathbb{P}})}\big[\rho(x, Y_t)\big]$ *for any time $t \in [T]$, scenario $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$, and action $x \in \mathcal{X}$;*
3. $\tilde{\mathbb{P}}_{g\left(\tilde{\varphi}\left(\tilde{f}(x), \tilde{Y}_t\right)\right)} = \big(\hbar(\tilde{\mathbb{P}})\big)_{\varphi(x, Y_t)}$ *for any time $t \in [T]$, scenario $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$, and action $x \in \mathcal{X}$.*

Then $R_T^\star(\mathscr{G}) \ge R_T^\star(\tilde{\mathscr{G}})$.

The second lemma addresses feedback with uniformative (i.e., scenario-independent) components. At a high level, if the feedback of some of the player's actions has one or more uninformative components, the game can be simplified by getting rid of the uninformative parts of the feedback. The player can achieve this by simulating the uninformative parts of the feedback using their randomization. The proof is deferred to Appendix B.1.

Lemma B.2 (Simulation)**.** *Let $\mathcal{V}, \mathcal{W}$ be two sets, $\mathscr{G} := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P})$ a game with $\mathcal{Z} = \mathcal{V} \times \mathcal{W}$, $\mathcal{S}$ its set of scenarios, $(Y_t)_{t\in\mathbb{N}}$ its adversary's actions, $\pi\colon \mathcal{Z} \to \mathcal{V}$ the projection on $\mathcal{V}$, and $T \in \mathbb{N}$ a horizon. Assume that $Y_1, \dots, Y_T$ are $\mathbb{P}$-independent for any scenario $\mathbb{P} \in \mathcal{S}$ and that there exist disjoint sets $\mathcal{I}, \mathcal{U} \subset \mathcal{X}$ such that $\mathcal{I} \cup \mathcal{U} = \mathcal{X}$ and*

1. *for any time $t \in [T]$ and action $x \in \mathcal{I}$ there exists $\psi_{t,x}\colon [0,1] \to \mathcal{W}$ such that, for all $\mathbb{P} \in \mathcal{S}$,*

$$\mathbb{P}_{\varphi(x, Y_t)} = \mathbb{P}_{\pi\left(\varphi(x, Y_t)\right)} \otimes (\mu_L)_{\psi_{t,x}} \ ;$$

2. *for any time $t \in [T]$ and action $x \in \mathcal{U}$, there exists $\gamma_{t,x}\colon [0,1] \to \mathcal{Z}$ such that, for all $\mathbb{P} \in \mathcal{S}$,*

$$\mathbb{P}_{\varphi(x, Y_t)} = (\mu_L)_{\gamma_{t,x}} \ .$$

Let $* \in \mathcal{V}$ and define

$$\tilde{\varphi}\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{V}, \ (x, y) \mapsto \begin{cases} \pi\big(\varphi(x, y)\big) \ , & \text{if } x \in \mathcal{I}, \\ * \ , & \text{if } x \in \mathcal{U}. \end{cases}$$

Define the game $\tilde{\mathscr{G}} := (\mathcal{X}, \mathcal{Y}, \mathcal{V}, \rho, \tilde{\varphi}, \mathscr{P})$. Then $R_T^\star(\mathscr{G}) \ge R_T^\star(\tilde{\mathscr{G}})$.

B.1 Proofs of the lemmas

In this section, we will give a full proof of the two important Embedding and Simulation lemmas introduces in Appendix B. To lighten the notation, for any $m, n \in \mathbb{N}$, with $m \leq n$ and a family $(\lambda_k)_{k\in\mathbb{N}}$ we let $\lambda_{m:n} := (\lambda_m, \lambda_{m+1}, \dots, \lambda_n)$ and similarly $\lambda_{n:m} := (\lambda_n, \lambda_{n-1} \dots, \lambda_m)$.

We begin by proving the Embedding lemma.

Lemma B.1 (Embedding)**.** *Let $\mathscr{G} := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P})$ and $\tilde{\mathscr{G}} := (\tilde{\mathcal{X}}, \tilde{\mathcal{Y}}, \tilde{\mathcal{Z}}, \tilde{\rho}, \tilde{\varphi}, \tilde{\mathscr{P}})$ be two games, $\mathcal{S}, \tilde{\mathcal{S}}$ their respective sets of scenarios, $(Y_t)_{t\in\mathbb{N}}, (\tilde{Y}_t)_{t\in\mathbb{N}}$ their adversaries' actions, and $T \in \mathbb{N}$ a horizon. Assume that $Y_1, \dots, Y_T$ are $\mathbb{P}$-independent for any scenario $\mathbb{P} \in \mathcal{S}$, $\tilde{Y}_1, \dots, \tilde{Y}_T$ are $\tilde{\mathbb{P}}$-independent for any scenario $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$, and that there exist $\tilde{f}\colon \mathcal{X} \to \tilde{\mathcal{X}}$, $g\colon \tilde{\mathcal{Z}} \to \mathcal{Z}$, and $\hbar\colon \tilde{\mathcal{S}} \to \mathcal{S}$ satisfying:*

1. $\sup_{\tilde{x}\in\tilde{\mathcal{X}}} \sum_{t=1}^{T} \mathbb{E}_{\tilde{\mathbb{P}}}\big[\tilde{\rho}(\tilde{x}, \tilde{Y}_t)\big] \leq \sup_{x\in\mathcal{X}} \sum_{t=1}^{T} \mathbb{E}_{\hbar(\tilde{\mathbb{P}})}\big[\rho(x, Y_t)\big]$ *for any scenario $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$;*
2. $\mathbb{E}_{\tilde{\mathbb{P}}}\big[\tilde{\rho}\big(\tilde{f}(x), \tilde{Y}_t\big)\big] \geq \mathbb{E}_{\hbar(\tilde{\mathbb{P}})}\big[\rho(x, Y_t)\big]$ *for any time $t \in [T]$, scenario $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$, and action $x \in \mathcal{X}$;*
3. $\tilde{\mathbb{P}}_{g\left(\tilde{\varphi}\left(\tilde{f}(x), \tilde{Y}_t\right)\right)} = \big(\hbar(\tilde{\mathbb{P}})\big)_{\varphi(x, Y_t)}$ *for any time $t \in [T]$, scenario $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$, and action $x \in \mathcal{X}$.*

Then $R_T^\star(\mathscr{G}) \geq R_T^\star(\tilde{\mathscr{G}})$.

Proof. Fix any strategy $\alpha \in \mathscr{A}(\mathscr{G})$. For each time $t \in \mathbb{N}$, define

$$\tilde{\alpha}_t\colon [0,1]^t \times \tilde{\mathcal{Z}}^{t-1} \to \tilde{\mathcal{X}}, (u_1, \dots, u_t, \tilde{z}_1, \dots, \tilde{z}_{t-1}) \mapsto \tilde{f}\Big(\alpha_t\big(u_1, \dots, u_t, g(\tilde{z}_1), \dots, g(\tilde{z}_{t-1})\big)\Big).$$

Then $\tilde{\alpha} := (\tilde{\alpha}_t)_{t\in\mathbb{N}} \in \mathscr{A}(\tilde{\mathscr{G}})$. As usual, let $(Y_t)_{t\in\mathbb{N}}$ and $(U_t)_{t\in\mathbb{N}}$ be the adversary's actions and the player's randomization in game $\mathscr{G}$ and $(X_t)_{t\in\mathbb{N}}$ and $(Z_t)_{t\in\mathbb{N}}$ the player's actions and the feedback according to the strategy α. Let $(\tilde{Y}_t)_{t\in\mathbb{N}}, (\tilde{U}_t)_{t\in\mathbb{N}}, (\tilde{X}_t)_{t\in\mathbb{N}}, (\tilde{Z}_t)_{t\in\mathbb{N}}$ be the corresponding objects for the game $\tilde{\mathscr{G}}$ and the strategy $\tilde{\alpha}$. Furthermore, define

$$\widehat{X}_1 = \alpha_1(\tilde{U}_1), \quad \widehat{Z}_1 = g\big(\tilde{\varphi}(\tilde{X}_1, \tilde{Y}_1)\big), \quad \widehat{X}_2 = \alpha_2(\tilde{U}_1, \tilde{U}_2, \widehat{Z}_1), \quad \widehat{Z}_2 = g\big(\tilde{\varphi}(\tilde{X}_2, \tilde{Y}_2)\big), \dots \ .$$

Fix $\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}$, where $\tilde{\mathcal{S}}$ are the scenarios of the game $\tilde{\mathscr{G}}$. Then $\tilde{\mathbb{P}}_{\tilde{U}_1} = \big(\hbar(\tilde{\mathbb{P}})\big)_{U_1}$. Now, since $X_1 = \alpha_1(U_1)$ and $\widehat{X}_1 = \alpha_1(\tilde{U}_1)$, we also have that $\tilde{\mathbb{P}}_{\widehat{X}_1, \tilde{U}_1} = \big(\hbar(\tilde{\mathbb{P}})\big)_{X_1, U_1} =: \mathbb{Q}_1$. Now, up to a set with $\mathbb{Q}_1$-probability zero, if $x_1 \in \mathcal{X}$ and $u_1 \in [0,1]$, we get, using Item (3):

$$\begin{aligned}\tilde{\mathbb{P}}_{\widehat{Z}_1 | \widehat{X}_1 = x_1, \tilde{U}_1 = u_1} &= \tilde{\mathbb{P}}_{g\left(\tilde{\varphi}\left(\tilde{f}(\widehat{X}_1), \tilde{Y}_1\right)\right) | \widehat{X}_1 = x_1, \tilde{U}_1 = u_1} = \tilde{\mathbb{P}}_{g\left(\tilde{\varphi}\left(\tilde{f}(x_1), \tilde{Y}_1\right)\right)} \\ &= \big(\hbar(\tilde{\mathbb{P}})\big)_{\varphi(x_1, Y_1)} = \big(\hbar(\tilde{\mathbb{P}})\big)_{\varphi(X_1, Y_1) | X_1 = x_1, U_1 = u_1} = \big(\hbar(\tilde{\mathbb{P}})\big)_{Z_1 | X_1 = x_1, U_1 = u_1} \ .\end{aligned}$$

So, if $A_1 \subset \mathcal{Z}$ and $D \subset \mathcal{X} \times [0,1]$, then

$$\begin{aligned}\tilde{\mathbb{P}}_{\widehat{Z}_1, \left(\widehat{X}_1, \tilde{U}_1\right)}(A_1 \times D) &= \int_D \mathbb{P}_{\widehat{Z}_1 | \widehat{X}_1 = x_1, \tilde{U}_1 = u_1}(A_1) \, \mathrm{d}\mathbb{P}_{\widehat{X}_1, \tilde{U}_1}(x_1, u_1) \\ &= \int_D \big(\hbar(\tilde{\mathbb{P}})\big)_{Z_1 | X_1 = x_1, U_1 = u_1}(A_1) \, \mathrm{d}\big(\hbar(\tilde{\mathbb{P}})\big)_{X_1, U_1}(x_1, u_1) = \big(\hbar(\tilde{\mathbb{P}})\big)_{Z_1, (X_1, U_1)}(A_1 \times D),\end{aligned}$$

from which it follows that $\tilde{\mathbb{P}}_{\widehat{Z}_1, \widehat{X}_1, \tilde{U}_1} = \big(\hbar(\tilde{\mathbb{P}})\big)_{Z_1, X_1, U_1}$. By induction, suppose that for time $t \in [T-1]$ we have that

$$\tilde{\mathbb{P}}_{\widehat{Z}_t, \dots, \widehat{Z}_1, \widehat{X}_t, \dots, \widehat{X}_1, \tilde{U}_t, \dots, \tilde{U}_1} = \big(\hbar(\tilde{\mathbb{P}})\big)_{Z_t, \dots, Z_1, X_t, \dots, X_1, U_t, \dots, U_1} \ .$$

Then, using independence we have that

$$\tilde{\mathbb{P}}_{\widehat{Z}_t, \dots, \widehat{Z}_1, \widehat{X}_t, \dots, \widehat{X}_1, \tilde{U}_{t+1}, \tilde{U}_t, \dots, \tilde{U}_1} = \big(\hbar(\tilde{\mathbb{P}})\big)_{Z_t, \dots, Z_1, X_t, \dots, X_1, U_{t+1}, U_t, \dots, U_1} \ .$$

Furthermore, since $X_{t+1} = \alpha_{t+1}(U_1, \dots, U_{t+1}, Z_1, \dots, Z_t)$ and $\widehat{X}_{t+1} = \alpha_{t+1}(\tilde{U}_1, \dots, \tilde{U}_{t+1}, \widehat{Z}_1, \dots, \widehat{Z}_t)$, we have that

$$\tilde{\mathbb{P}}_{\widehat{Z}_t,\dots,\widehat{Z}_1,\widehat{X}_{t+1},\widehat{X}_t,\dots,\widehat{X}_1,\tilde{U}_{t+1},\tilde{U}_t,\dots,\tilde{U}_1} = \big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{Z_t,\dots,Z_1,X_{t+1},X_t,\dots,X_1,U_{t+1},U_t,\dots,U_1} =: \mathbb{Q}_{t+1} \ .$$

Now, up to a set with $\mathbb{Q}_{t+1}$-probability zero, if $x_1, \dots, x_{t+1} \in \mathcal{X}$, $u_1, \dots, u_{t+1} \in [0,1]$, and $z_1, \dots, z_t \in \mathcal{Z}$, by the $\tilde{\mathbb{P}}$-independence of $\tilde{Y}_1, \dots, \tilde{Y}_{t+1}$, Item 3, and the $\mathscr{h}(\tilde{\mathbb{P}})$-independence of $Y_1, \dots, Y_{t+1}$, we have

$$\begin{aligned}
&\tilde{\mathbb{P}}_{\widehat{Z}_{t+1}|\widehat{Z}_t=z_t,\dots,\widehat{Z}_1=z_1,\widehat{X}_{t+1}=x_{t+1},\dots,\widehat{X}_1=x_1,\tilde{U}_{t+1}=u_{t+1},\dots,\tilde{U}_1=u_1} \\
&\qquad = \tilde{\mathbb{P}}_{g\left(\tilde{\varphi}\left(\tilde{f}(\widehat{X}_{t+1}),\tilde{Y}_{t+1}\right)\right)|\widehat{Z}_t=z_t,\dots,\widehat{Z}_1=z_1,\widehat{X}_{t+1}=x_{t+1},\dots,\widehat{X}_1=x_1,\tilde{U}_{t+1}=u_{t+1},\dots,\tilde{U}_1=u_1} = \tilde{\mathbb{P}}_{g\left(\tilde{\varphi}\left(\tilde{f}(x_{t+1}),\tilde{Y}_{t+1}\right)\right)} \\
&\qquad = \big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{\varphi(x_{t+1},Y_{t+1})} = \big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{\varphi(X_{t+1},Y_{t+1})|Z_t=z_t,\dots,Z_1=z_1,X_{t+1}=x_{t+1},\dots,X_1=x_1,U_{t+1}=u_{t+1},\dots,U_1=u_1} \\
&\qquad = \big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{Z_{t+1}|Z_t=z_t,\dots,Z_1=z_1,X_{t+1}=x_{t+1},\dots,X_1=x_1,U_{t+1}=u_{t+1},\dots,U_1=u_1} \ .
\end{aligned}$$

So, if $A_{t+1} \subset \mathcal{Z}, D \subset \mathcal{Z}^t \times \mathcal{X}^{t+1} \times [0,1]^{t+1}$, we have that

$$\begin{aligned}
&\tilde{\mathbb{P}}_{\widehat{Z}_{t+1},\left(\widehat{Z}_{t:1},\widehat{X}_{t+1:1},\tilde{U}_{t+1:1}\right)}(A_{t+1} \times D) \\
&\qquad = \int_D \tilde{\mathbb{P}}_{\widehat{Z}_{t+1}|\widehat{Z}_{t:1}=z_{t:1},\widehat{X}_{t+1:1}=x_{t+1:1},\tilde{U}_{t+1:1}=u_{t+1:1}}(A_{t+1}) \, \mathrm{d}\tilde{\mathbb{P}}_{\widehat{Z}_{t:1},\widehat{X}_{t+1:1},\tilde{U}_{t+1:1}}(z_{t:1}, x_{t+1:1}, u_{t+1:1}) \\
&\qquad = \int_D \big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{Z_{t+1}|Z_{t:1}=z_{t:1},C_{t+1:1}=x_{t+1:1},U_{t+1:1}=u_{t+1:1}}(A_{t+1}) \, \mathrm{d}\big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{Z_{t:1},X_{t+1:1},U_{t+1:1}}(z_{t:1}, x_{t+1:1}, u_{t+1:1}) \\
&\qquad = \big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{Z_{t+1},\left(Z_{t:1},X_{t+1:1},U_{t+1:1}\right)}(A_{t+1} \times D) \ ,
\end{aligned}$$

from which follows that $\tilde{\mathbb{P}}_{\widehat{Z}_{t+1},\dots,\widehat{Z}_1,\widehat{X}_{t+1},\dots,\widehat{X}_1,\tilde{U}_{t+1},\dots,\tilde{U}_1} = \big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{Z_{t+1},\dots,Z_1,X_{t+1},\dots,X_1,U_{t+1},\dots,U_1}$. In particular, for each $t \in [T]$ we have that $\tilde{\mathbb{P}}_{\widehat{X}_t} = \big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{X_t}$. Hence, using the $\mathscr{h}(\tilde{\mathbb{P}})$-independence of $Y_1, \dots, Y_T$, Item (2), and the $\tilde{\mathbb{P}}$-independence of $\tilde{Y}_1, \dots, \tilde{Y}_T$, we get

$$\begin{aligned}
\sum_{t=1}^T \mathbb{E}_{\mathscr{h}(\tilde{\mathbb{P}})}\big[\rho(X_t, Y_t)\big] &= \sum_{t=1}^T \int_{\mathcal{X}} \mathbb{E}_{\mathscr{h}(\tilde{\mathbb{P}})}\big[\rho(x, Y_t)\big] \, \mathrm{d}\big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{X_t}(x) \\
&\le \sum_{t=1}^T \int_{\mathcal{X}} \mathbb{E}_{\tilde{\mathbb{P}}}[\tilde{\rho}(\tilde{f}(x), \tilde{Y}_t)] \, \mathrm{d}\big(\mathscr{h}(\tilde{\mathbb{P}})\big)_{X_t}(x) \\
&= \sum_{t=1}^T \int_{\mathcal{X}} \mathbb{E}_{\tilde{\mathbb{P}}}[\tilde{\rho}(\tilde{f}(x), \tilde{Y}_t)] \, \mathrm{d}\tilde{\mathbb{P}}_{\widehat{X}_t}(x) \\
&= \sum_{t=1}^T \mathbb{E}_{\tilde{\mathbb{P}}}\big[\tilde{\rho}(\tilde{f}(\widehat{X}_t), \tilde{Y}_t)\big] = \sum_{t=1}^T \mathbb{E}_{\tilde{\mathbb{P}}}\big[\tilde{\rho}(\tilde{X}_t, \tilde{Y}_t)\big] \ .
\end{aligned}$$

Then, using Item (1), we have

$$\begin{aligned}
R_T^{\mathscr{h}(\tilde{\mathbb{P}})}(\alpha) &= \sup_{x \in \mathcal{X}} \left(\sum_{t=1}^T \mathbb{E}_{\mathscr{h}(\tilde{\mathbb{P}})}\big[\rho(x, Y_t)\big] - \sum_{t=1}^T \mathbb{E}_{\mathscr{h}(\tilde{\mathbb{P}})}\big[\rho(X_t, Y_t)\big] \right) \\
&\ge \sup_{\tilde{x} \in \tilde{\mathcal{X}}} \left(\sum_{t=1}^T \mathbb{E}_{\tilde{\mathbb{P}}}\big[\tilde{\rho}(\tilde{x}, \tilde{Y}_t)\big] - \sum_{t=1}^T \mathbb{E}_{\tilde{\mathbb{P}}}\big[\tilde{\rho}(\tilde{X}_t, \tilde{Y}_t)\big] \right) = R_T^{\tilde{\mathbb{P}}}(\tilde{\alpha}) \ .
\end{aligned}$$

Since $\tilde{\mathbb{P}}$ was arbitrary, we get

$$R_T^\star(\tilde{\mathscr{G}}) = \inf_{\beta \in \mathscr{A}(\tilde{\mathscr{G}})} R_T^{\tilde{\mathcal{S}}}(\beta) \le R_T^{\tilde{\mathcal{S}}}(\tilde{\alpha}) = \sup_{\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}} R_T^{\tilde{\mathbb{P}}}(\tilde{\alpha}) \le \sup_{\tilde{\mathbb{P}} \in \tilde{\mathcal{S}}} R_T^{\mathscr{h}(\tilde{\mathbb{P}})}(\alpha) \le \sup_{\mathbb{P} \in \mathcal{S}} R_T^{\mathbb{P}}(\alpha) = R_T^{\mathcal{S}}(\alpha) \ ,$$

and since α was arbitrary, we get

$$R_T^\star(\tilde{\mathscr{G}}) \le \inf_{\alpha \in \mathscr{A}(\mathscr{G})} R_T^{\mathcal{S}}(\alpha) = R_T^\star(\mathscr{G}) \ . \qquad \square$$

We now prove the Simulation lemma we introduced in Appendix B showing how to get rid of uninformative feedback.

Lemma B.2 (Simulation)**.** *Let $\mathcal{V}, \mathcal{W}$ be two sets, $\mathscr{G} := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P})$ a game with $\mathcal{Z} = \mathcal{V} \times \mathcal{W}$, $\mathcal{S}$ its set of scenarios, $(Y_t)_{t\in\mathbb{N}}$ its adversary's actions, $\pi\colon \mathcal{Z} \to \mathcal{V}$ the projection on $\mathcal{V}$, and $T \in \mathbb{N}$ a horizon. Assume that $Y_1, \dots, Y_T$ are $\mathbb{P}$-independent for any scenario $\mathbb{P} \in \mathcal{S}$ and that there exist disjoint sets $\mathcal{I}, \mathcal{U} \subset \mathcal{X}$ such that $\mathcal{I} \cup \mathcal{U} = \mathcal{X}$ and*

1. *for any time $t \in [T]$ and action $x \in \mathcal{I}$ there exists $\psi_{t,x}\colon [0,1] \to \mathcal{W}$ such that, for all $\mathbb{P} \in \mathcal{S}$,*

$$\mathbb{P}_{\varphi(x,Y_t)} = \mathbb{P}_{\pi\left(\varphi(x,Y_t)\right)} \otimes (\mu_L)_{\psi_{t,x}} \ ;$$

2. *for any time $t \in [T]$ and action $x \in \mathcal{U}$, there exists $\gamma_{t,x}\colon [0,1] \to \mathcal{Z}$ such that, for all $\mathbb{P} \in \mathcal{S}$,*

$$\mathbb{P}_{\varphi(x,Y_t)} = (\mu_L)_{\gamma_{t,x}} \ .$$

Let $* \in \mathcal{V}$ and define

$$\tilde{\varphi}\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{V},\ (x,y) \mapsto \begin{cases} \pi\big(\varphi(x,y)\big)\,, & \text{if } x \in \mathcal{I}, \\ *\,, & \text{if } x \in \mathcal{U}. \end{cases}$$

Define the game $\tilde{\mathscr{G}} := (\mathcal{X}, \mathcal{Y}, \mathcal{V}, \rho, \tilde{\varphi}, \mathscr{P})$. Then $R_T^\star(\mathscr{G}) \geq R_T^\star(\tilde{\mathscr{G}})$.

Proof. For each number $a \in [0,1]$, fix a binary representation $0.a_1a_2a_3\dots$ of a and define $\xi(a) := 0.a_1a_3a_5\dots$, $\zeta(a) := 0.a_2a_4a_6\dots$. Note that the two resulting functions $\xi, \zeta\colon [0,1] \to [0,1]$ are μ_L-independent with common (uniform) push-forward distribution $(\mu_L)_\xi = \mu_L = (\mu_L)_\zeta$.

Let $(Y_t)_{t\in\mathbb{N}}, (U_t)_{t\in\mathbb{N}}$ be the sequences of adversary's actions and player's randomization for the sequential game $\mathscr{G}$ and note that they are also the same for the sequential game $\tilde{\mathscr{G}}$. For each $t \in \mathbb{N}$ define $\beta_t\colon \mathcal{X} \times \mathcal{V} \times [0,1] \to \mathcal{Z}$ via

$$(x,v,u) \mapsto \begin{cases} \big(v, \psi_{t,x}(u)\big)\,, & \text{if } x \in \mathcal{I}\,, \\ \gamma_{t,x}(u)\,, & \text{if } x \in \mathcal{U}\,, \end{cases}$$

if $t \leq T$ and in an arbitrary manner if $t \geq T+1$. Fix $\alpha = (\alpha_t)_{t\in\mathbb{N}} \in \mathscr{A}(\mathscr{G})$. Let $(X_t)_{t\in\mathbb{N}}, (Z_t)_{t\in\mathbb{N}}$ be the sequences of player's actions and feedback associated to the strategy α.

Fix $(u_t)_{t\in\mathbb{N}} \subset [0,1]$ and $(v_t)_{t\in\mathbb{N}} \subset \mathcal{V}$. Define by induction (on t) the sequences $(x_t)_{t\in\mathbb{N}}$ and $(z_t)_{t\in\mathbb{N}}$ via the relationships

$$x_t = \alpha_t\big(\xi(u_1), \dots, \xi(u_t), z_1, \dots, z_{t-1}\big), \qquad z_t = \beta_t\big(x_t, v_t, \zeta(u_t)\big).$$

Note that for each $t \in \mathbb{N}$, we have that x_t depends only on $u_1, \dots, u_t, v_1, \dots, v_{t-1}$, so we can define

$$\tilde{\alpha}_t(u_1, \dots, u_t, v_1, \dots, v_{t-1}) := x_t.$$

Being $(u_t)_{t\in\mathbb{N}}$ and $(v_t)_{t\in\mathbb{N}}$ arbitrary, this defines a sequence of functions $(\tilde{\alpha}_t)_{t\in\mathbb{N}}$ such that, for all $t \in \mathbb{N}$,

$$\tilde{\alpha}_t\colon [0,1]^t \times \mathcal{V}^{t-1} \to \mathcal{X}$$

i.e., $\tilde{\alpha} := (\tilde{\alpha}_t)_{t\in\mathbb{N}} \in \mathscr{A}(\tilde{\mathscr{G}})$. Let $(\tilde{X}_t)_{t\in\mathbb{N}}$ and $(\tilde{V}_t)_{t\in\mathbb{N}}$ be respectively the sequence of player's actions and the feedback sequence associated with the strategy $\tilde{\alpha}$. For each $t \in \mathbb{N}$, define also $\tilde{Z}_t := \beta_t\big(\tilde{X}_t, \tilde{V}_t, \zeta(U_t)\big)$. Note that for each $t \in \mathbb{N}$ it holds that $\tilde{X}_t = \alpha_t\big(\xi(U_1), \dots, \xi(U_t), \tilde{Z}_1, \dots, \tilde{Z}_{t-1}\big)$.

Fix a scenario $\mathbb{P} \in \mathcal{S}$. Note first that $\mathbb{P}_{\xi(U_1)} = \mathbb{P}_{U_1}$, and since $X_1 = \alpha_1(U_1)$ and $\tilde{X}_1 = \tilde{\alpha}_1(U_1) = \alpha_1\big(\xi(U_1)\big)$, we also have that $\mathbb{P}_{\tilde{X}_1, \xi(U_1)} = \mathbb{P}_{X_1, U_1} =: \mathbb{Q}_1$. Now, up to a set with $\mathbb{Q}_1$-probability zero, if $x_1 \in \mathcal{X}$ and $u_1 \in [0,1]$,

using Items (1) and (2), we have that

$$\begin{aligned}
\mathbb{P}_{\tilde{Z}_1|\tilde{X}_1=x_1,\xi(U_1)=u_1} &= \mathbb{P}_{\beta_1\left(\tilde{X}_1,\tilde{\varphi}(\tilde{X}_1,Y_1),\zeta(U_1)\right)|\tilde{X}_1=x_1,\xi(U_1)=u_1} = \mathbb{P}_{\beta_1\left(x_1,\tilde{\varphi}(x_1,Y_1),\zeta(U_1)\right)} \\
&= \begin{cases} \mathbb{P}_{\beta_1\left(x_1,\pi\left(\varphi(x_1,Y_1)\right),\zeta(U_1)\right)} & \text{if } x_1 \in \mathcal{I} \\ \mathbb{P}_{\beta_1\left(x_1,*,\zeta(U_1)\right)} & \text{if } x_1 \in \mathcal{U} \end{cases} = \begin{cases} \mathbb{P}_{\left(\pi\left(\varphi(x_1,Y_1)\right),\psi_{1,x_1}\left(\zeta(U_1)\right)\right)} & \text{if } x_1 \in \mathcal{I} \\ \mathbb{P}_{\gamma_{1,x_1}\left(\zeta(U_1)\right)} & \text{if } x_1 \in \mathcal{U} \end{cases} \\
&= \begin{cases} \mathbb{P}_{\pi\left(\varphi(x_1,Y_1)\right)} \otimes \mathbb{P}_{\psi_{1,x_1}\left(\zeta(U_1)\right)} & \text{if } x_1 \in \mathcal{I} \\ \mathbb{P}_{\gamma_{1,x_1}\left(\zeta(U_1)\right)} & \text{if } x_1 \in \mathcal{U} \end{cases} = \begin{cases} \mathbb{P}_{\pi\left(\varphi(x_1,Y_1)\right)} \otimes \left(\mathbb{P}_{\zeta(U_1)}\right)_{\psi_{1,x_1}} & \text{if } x_1 \in \mathcal{I} \\ \left(\mathbb{P}_{\zeta(U_1)}\right)_{\gamma_{1,x_1}} & \text{if } x_1 \in \mathcal{U} \end{cases} \\
&= \begin{cases} \mathbb{P}_{\pi\left(\varphi(x_1,Y_1)\right)} \otimes \left(\mu_L\right)_{\psi_{1,x_1}} & \text{if } x_1 \in \mathcal{I} \\ \left(\mu_L\right)_{\gamma_{1,x_1}} & \text{if } x_1 \in \mathcal{U} \end{cases} = \mathbb{P}_{\varphi(x_1,Y_1)} = \mathbb{P}_{\varphi(X_1,Y_1)|X_1=x_1,U_1=u_1} = \mathbb{P}_{Z_1|X_1=x_1,U_1=u_1} \ .
\end{aligned}$$

So, if $A_1 \subset \mathcal{Z}$ and $D \subset \mathcal{X} \times [0,1]$, then

$$\begin{aligned}
\mathbb{P}_{\tilde{Z}_1,\left(\tilde{X}_1,\xi(U_1)\right)}(A_1 \times D) &= \int_D \mathbb{P}_{\tilde{Z}_1|\tilde{X}_1=x_1,\xi(U_1)=u_1}(A_1)\,\mathrm{d}\mathbb{P}_{\tilde{X}_1,\xi(U_1)}(x_1,u_1) \\
&= \int_D \mathbb{P}_{Z_1|X_1=x_1,U_1=u_1}(A_1)\,\mathrm{d}\mathbb{P}_{X_1,U_1}(x_1,u_1) = \mathbb{P}_{Z_1,(X_1,U_1)}(A_1 \times D) \ ,
\end{aligned}$$

from which it follows that $\mathbb{P}_{\tilde{Z}_1,\tilde{X}_1,\xi(U_1)} = \mathbb{P}_{Z_1,X_1,U_1}$. By induction, suppose that for $t \in [T-1]$ we have that

$$\mathbb{P}_{\tilde{Z}_t,\ldots,\tilde{Z}_1,\tilde{X}_t,\ldots,\tilde{X}_1,\xi(U_t),\ldots,\xi(U_1)} = \mathbb{P}_{Z_t,\ldots,Z_1,X_t,\ldots,X_1,U_t,\ldots,U_1} \ .$$

Then, using independence we have that

$$\mathbb{P}_{\tilde{Z}_t,\ldots,\tilde{Z}_1,\tilde{X}_t,\ldots,\tilde{X}_1,\xi(U_{t+1}),\xi(U_t),\ldots,\xi(U_1)} = \mathbb{P}_{Z_t,\ldots,Z_1,X_t,\ldots,X_1,U_{t+1},U_t,\ldots,U_1} \ .$$

Furthermore, since $X_{t+1} = \alpha_{t+1}(U_1,\ldots,U_{t+1},Z_1,\ldots,Z_t)$ and

$$\tilde{X}_{t+1} = \tilde{\alpha}_{t+1}(U_1,\ldots,U_{t+1},\tilde{V}_1,\ldots,\tilde{V}_t) = \alpha_{t+1}(\xi(U_1),\ldots,\xi(U_{t+1}),\tilde{Z}_1,\ldots,\tilde{Z}_t) \ ,$$

we have that

$$\mathbb{P}_{\tilde{Z}_t,\ldots,\tilde{Z}_1,\tilde{X}_{t+1},\tilde{X}_t,\ldots,\tilde{X}_1,\xi(U_{t+1}),\xi(U_t),\ldots,\xi(U_1)} = \mathbb{P}_{Z_t,\ldots,Z_1,X_{t+1},X_t,\ldots,X_1,U_{t+1},U_t,\ldots,U_1} =: \mathbb{Q}_{t+1} \ .$$

Now, up to a set with $\mathbb{Q}_{t+1}$-probability zero, if $x_1,\ldots,x_{t+1} \in \mathcal{X}$, $u_1,\ldots,u_{t+1} \in [0,1]$ and $z_1,\ldots,z_t \in \mathcal{Z}$, using

the $\mathbb{P}$-independence of $Y_1, \dots, Y_{t+1}$ and Items (1)–(2), we have that

$$
\begin{aligned}
&\mathbb{P}_{\tilde{Z}_{t+1}|\tilde{Z}_t=z_t,\dots,\tilde{Z}_1=z_1,\tilde{X}_{t+1}=x_{t+1},\dots,\tilde{X}_1=x_1,\xi(U_{t+1})=u_{t+1},\dots,\xi(U_1)=u_1} \\
&\qquad = \mathbb{P}_{\beta_{t+1}\left(\tilde{X}_{t+1},\tilde{\varphi}(\tilde{X}_{t+1},Y_{t+1}),\zeta(U_{t+1})\right)|\tilde{Z}_t=z_t,\dots,\tilde{Z}_1=z_1,\tilde{X}_{t+1}=x_{t+1},\dots,\tilde{X}_1=x_1,\xi(U_{t+1})=u_{t+1},\dots,\xi(U_1)=u_1} \\
&\qquad = \mathbb{P}_{\beta_{t+1}\left(x_{t+1},\tilde{\varphi}(x_{t+1},Y_{t+1}),\zeta(U_{t+1})\right)} = \begin{cases} \mathbb{P}_{\beta_{t+1}\left(x_{t+1},\pi\left(\varphi(x_{t+1},Y_{t+1})\right),\zeta(U_{t+1})\right)} & \text{if } x_{t+1} \in \mathcal{I} \\ \mathbb{P}_{\beta_{t+1}\left(x_{t+1},*,\zeta(U_{t+1})\right)} & \text{if } x_{t+1} \in \mathcal{U} \end{cases} \\
&\qquad = \begin{cases} \mathbb{P}_{\left(\pi\left(\varphi(x_{t+1},Y_{t+1})\right),\psi_{t+1,x_{t+1}}\left(\zeta(U_{t+1})\right)\right)} & \text{if } x_{t+1} \in \mathcal{I} \\ \mathbb{P}_{\gamma_{t+1,x_{t+1}}\left(\zeta(U_{t+1})\right)} & \text{if } x_{t+1} \in \mathcal{U} \end{cases} \\
&\qquad = \begin{cases} \mathbb{P}_{\pi\left(\varphi(x_{t+1},Y_{t+1})\right)} \otimes \mathbb{P}_{\psi_{t+1,x_{t+1}}\left(\zeta(U_{t+1})\right)} & \text{if } x_{t+1} \in \mathcal{I} \\ \mathbb{P}_{\gamma_{t+1,x_{t+1}}\left(\zeta(U_{t+1})\right)} & \text{if } x_{t+1} \in \mathcal{U} \end{cases} \\
&\qquad = \begin{cases} \mathbb{P}_{\pi\left(\varphi(x_{t+1},Y_{t+1})\right)} \otimes \left(\mathbb{P}_{\zeta(U_{t+1})}\right)_{\psi_{t+1,x_{t+1}}} & \text{if } x_{t+1} \in \mathcal{I} \\ \left(\mathbb{P}_{\zeta(U_{t+1})}\right)_{\gamma_{t+1,x_{t+1}}} & \text{if } x_{t+1} \in \mathcal{U} \end{cases} \\
&\qquad = \begin{cases} \mathbb{P}_{\pi\left(\varphi(x_{t+1},Y_{t+1})\right)} \otimes \left(\mu_L\right)_{\psi_{t+1,x_{t+1}}} & \text{if } x_{t+1} \in \mathcal{I} \\ \left(\mu_L\right)_{\gamma_{t+1,x_{t+1}}} & \text{if } x_{t+1} \in \mathcal{U} \end{cases} \\
&\qquad = \mathbb{P}_{\varphi(x_{t+1},Y_{t+1})} = \mathbb{P}_{\varphi(X_{t+1},Y_{t+1})|Z_t=z_t,\dots,Z_1=z_1,X_{t+1}=x_{t+1},\dots,X_1=x_1,U_{t+1}=u_{t+1},\dots,U_1=u_1} \\
&\qquad = \mathbb{P}_{Z_{t+1}|Z_t=z_t,\dots,Z_1=z_1,X_{t+1}=x_{t+1},\dots,X_1=x_1,U_{t+1}=u_{t+1},\dots,U_1=u_1} \ .
\end{aligned}
$$

So, if $A_{t+1} \subset \mathcal{Z}, D \subset \mathcal{Z}^t \times \mathcal{X}^{t+1} \times [0,1]^{t+1}$, we have that

$$
\begin{aligned}
&\mathbb{P}_{\tilde{Z}_{t+1},\left(\tilde{Z}_t\dots,\tilde{Z}_1,\tilde{X}_{t+1},\dots,\tilde{X}_1,\xi(U_{t+1}),\dots,\xi(U_1)\right)}(A_{t+1} \times D) \\
&\qquad = \int_D \mathbb{P}_{\tilde{Z}_{t+1}|\tilde{Z}_{t:1}=z_{t:1},\tilde{X}_{t+1:1}=x_{t+1:1},\left(\xi(U_{t+1}),\dots,\xi(U_1)\right)=u_{t+1:1}}(A_{t+1})\,\mathrm{d}\mathbb{Q}_{t+1}(z_{t:1},x_{t+1:1},u_{t+1:1}) \\
&\qquad = \int_D \mathbb{P}_{Z_{t+1}|Z_{t:1}=z_{t:1},X_{t+1:1}=x_{t+1:1},U_{t+1:1}=u_{t+1:1}}(A_{t+1})\,\mathrm{d}\mathbb{Q}_{t+1}(z_{t:1},x_{t+1:1},u_{t+1:1}) \\
&\qquad = \mathbb{P}_{Z_{t+1},\left(Z_t,\dots,Z_1,X_{t+1},\dots,X_1,U_{t+1},\dots,U_1\right)}(A_{t+1} \times D)
\end{aligned}
$$

from which it follows that $\mathbb{P}_{\tilde{Z}_{t+1:1},\tilde{X}_{t+1:1},\left(\xi(U_{t+1}),\dots,\xi(U_1)\right)} = \mathbb{P}_{Z_{t+1:1},X_{t+1:1},U_{t+1:1}}$. In particular, for each $t \in [T]$ we have that $\mathbb{P}_{X_t} = \mathbb{P}_{\tilde{X}_t}$. So, for each $t \in [T]$, using the $\mathbb{P}$-independence of $Y_1, \dots, Y_t$, we have that

$$
\mathbb{P}_{X_t,Y_t} = \mathbb{P}_{X_t} \otimes \mathbb{P}_{Y_t} = \mathbb{P}_{\tilde{X}_t} \otimes \mathbb{P}_{Y_t} = \mathbb{P}_{\tilde{X}_t,Y_t} \ ,
$$

and then

$$
\mathbb{E}_{\mathbb{P}}\left[\rho(X_t,Y_t)\right] = \mathbb{E}_{\mathbb{P}_{X_t,Y_t}}\left[\rho\right] = \mathbb{E}_{\mathbb{P}_{\tilde{X}_t,Y_t}}\left[\rho\right] = \mathbb{E}_{\mathbb{P}}\left[\rho(\tilde{X}_t,Y_t)\right] \ .
$$

In conclusion

$$
\begin{aligned}
R_T^{\mathbb{P}}(\alpha) &= \sup_{x \in \mathcal{X}} \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^T \rho(x,Y_t) - \sum_{t=1}^T \rho(X_t,Y_t)\right] = \sup_{x \in \mathcal{X}}\left(\sum_{t=1}^T \mathbb{E}_{\mathbb{P}}[\rho(x,Y_t)] - \sum_{t=1}^T \mathbb{E}_{\mathbb{P}}[\rho(X_t,Y_t)]\right) \\
&= \sup_{x \in \mathcal{X}}\left(\sum_{t=1}^T \mathbb{E}_{\mathbb{P}}[\rho(x,Y_t)] - \sum_{t=1}^T \mathbb{E}_{\mathbb{P}}\left[\rho(\tilde{X}_t,Y_t)\right]\right) = \sup_{x \in \mathcal{X}} \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^T \rho(x,Y_t) - \sum_{t=1}^T \rho(\tilde{X}_t,Y_t)\right] = R_T^{\mathbb{P}}(\tilde{\alpha}) \ .
\end{aligned}
$$

Since $\mathbb{P}$ was arbitrary, it follows that $R_T^{\mathcal{S}}(\alpha) = R_T^{\mathcal{S}}(\tilde{\alpha})$. Since α was arbitrary, it follows that

$$R_T^\star(\mathscr{G}) = \inf_{\alpha \in \mathscr{A}(\mathscr{G})} R_T^{\mathcal{S}}(\alpha) = \inf_{\alpha \in \mathscr{A}(\tilde{\mathscr{G}})} R_T^{\mathcal{S}}(\tilde{\alpha}) \geq \inf_{\alpha' \in \mathscr{A}(\tilde{\mathscr{G}})} R_T^{\mathcal{S}}(\alpha') = R_T^\star(\tilde{\mathscr{G}}) \,.$$

□

C $\sqrt{T}$ Lower Bound Under Full-Feedback (iv+bd)

In this section, we prove that in the full-feedback case, no strategy can beat the $\sqrt{T}$ rate that we proved in Theorem 3.1 when the seller/buyer pair (S_t, B_t) is drawn i.i.d. from an unknown fixed distribution, not even under the further assumptions that the valuations of the seller and buyer are independent of each other and have bounded densities.

The idea of the proof is to build a family of scenarios $\mathbb{P}^{\pm\varepsilon}$ parameterized by $\varepsilon \in [0,1]$, like in Fig. 1. The only way to avoid suffering linear regret in a scenario $\mathbb{P}^{\pm\varepsilon}$ is to identify the sign of $\pm\varepsilon$. Leveraging the Embedding and Simulation lemmas (Lemmas B.1 and B.2), this construction leads to a reduction to a two-armed bandit problem, which has a know lower bound on the regret of order $\sqrt{T}$.

Theorem (Theorem 3.2, restated)**.** *In the full-feedback stochastic (iid) setting with independent valuations (iv) and densities bounded by a constant $M \geq 4$ (bd), for all horizons $T \in \mathbb{N}$, the minimax regret satisfies*

$$R_T^\star \geq \frac{1}{160}\sqrt{T} \,.$$

Proof. Fix any horizon $T \in \mathbb{N}$ and $M \geq 4$. Recalling Appendix A.2, the full-feedback stochastic (iid) setting with independent valuations (iv) and densities bounded (bd) by M is a game $\mathscr{G} := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P})$, where $\mathcal{X} = [0,1]$, $\mathcal{Y} = [0,1]^2$, $\mathcal{Z} = [0,1]^2$, $\rho = \mathrm{GFT}$, $\varphi\colon \big(p, (s,b)\big) \mapsto (s,b)$, and $\mathscr{P} = \mathscr{P}^M_{\text{iv+bd}}$. Define, for each $\varepsilon \in [-1,1]$, the densities $f_{S,\varepsilon} = 2(1+\varepsilon)\mathbb{I}_{[0,\frac{1}{4}]} + 2(1-\varepsilon)\mathbb{I}_{[\frac{1}{2},\frac{3}{4}]}$ and $f_B = 2\mathbb{I}_{[\frac{1}{4},\frac{1}{2}]\cup[\frac{3}{4},1]}$. Fix the adversary's behavior $\mathscr{P}_1$ as the subset of $\mathscr{P}$ whose elements have the form $\boldsymbol{\mu}_\varepsilon := \otimes_{t\in\mathbb{N}}(f_{S,\varepsilon}\mu_L \otimes f_B\mu_L)$, for some $\varepsilon \in [-1,1]$. Since $\mathscr{P}_1 \subset \mathscr{P}$, the game $\mathscr{G}_1 := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P}_1)$ is easier than $\mathscr{G}$ (i.e., $R_T^\star(\mathscr{G}) \geq R_T^\star(\mathscr{G}_1)$) by the Embedding lemma (Lemma B.1) with $\tilde{f}$ and g as the identities, and h as the inclusion. Now, define $\rho_1\colon \mathcal{X}\times\mathcal{Y} \to [0,1]$, $(p,(s,b)) \mapsto (b-s)\mathbb{I}\{s \leq \frac{1}{4} \leq b\}\mathbb{I}\{p \leq \frac{1}{2}\} + (b-s)\mathbb{I}\{s \leq \frac{3}{4} \leq b\}\mathbb{I}\{p > \frac{1}{2}\}$ and note that, defining $\mathscr{G}_2 := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho_1, \varphi, \mathscr{P}_1)$, by the Embedding lemma with $\tilde{f}, g, h$ as the identities, we have that the game $\mathscr{G}_2$ is easier than the game $\mathscr{G}_1$ (i.e., $R_T^\star(\mathscr{G}_1) \geq R_T^\star(\mathscr{G}_2)$). Then, let $\mathcal{Z}_3 := \{0,1\}\times\big[0,\frac{1}{4}\big]\times[0,1]$ and $\varphi_3\colon \mathcal{X}\times\mathcal{Y} \to \mathcal{Z}_3$, $\big(p,(s,b)\big) \mapsto \big(\mathbb{I}\{s \leq 1/4\},\, s\mathbb{I}\{s \leq 1/4\} + (s-1/2)\mathbb{I}\{1/2 \leq s \leq 3/4\},\, b\big)$. Define the game $\mathscr{G}_3 := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}_3, \rho_1, \varphi_3, \mathscr{P}_1)$. By the Embedding lemma with $\tilde{f}, h$ as the identities and $g\colon \mathcal{Z}_3 \to \mathcal{Z}$, $(i, \tilde{s}, b) \mapsto \big(\tilde{s}i + (1/2+\tilde{s})(1-i),\, b\big)$, we have that the game $\mathscr{G}_3$ is easier than the game $\mathscr{G}_2$ (i.e., $R_T^\star(\mathscr{G}_2) \geq R_T^\star(\mathscr{G}_3)$). Next, let $\varphi_4\colon \mathcal{X}\times\mathcal{Y} \to \mathcal{Z}_3$, $\big(p,(s,b)\big) \mapsto \mathbb{I}\{s \leq \frac{1}{4}\}$, and define the game $\mathscr{G}_4 := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}_3, \rho_1, \varphi_4, \mathscr{P}_1)$. Let $(Y_t)_{t\in\mathbb{N}}$ be the adversary's actions in $\mathscr{G}_4$. A tedious computation verifies that for all $t \in \mathbb{N}$, $p \in \mathcal{X}$, and scenarios $\mathbb{P}$ of game $\mathscr{G}_3$, $\mathbb{P}_{\varphi_3(p,Y_t)} = \mathbb{P}_{\pi(\varphi_3(p,Y_t))} \otimes (\nu \otimes f_B\mu_L)$, where $\pi\colon \mathcal{Z}_3 \to \{0,1\}$ is the projection on the first component $\{0,1\}$ of $\mathcal{Z}_3$ and ν is the uniform distribution on $[0, 1/4]$. By the well-known Skorokhod representation (Williams, 1991, Section 17.3), there exists $\psi\colon [0,1] \to [0,1/4]\times[0,1]$ such that $\nu \otimes f_B\mu_L = (\mu_L)_\psi$. Thus, by the Simulation lemma (Lemma B.2) with $\mathcal{I} = \mathcal{X}$ and $\mathcal{U} = \varnothing$, the game $\mathscr{G}_4$ is easier than $\mathscr{G}_3$ (i.e., $R_T^\star(\mathscr{G}_3) \geq R_T^\star(\mathscr{G}_4)$). Finally, consider the game $\mathscr{G}_5 := \big(\{1,2\}, \{1,2\}, \{0,1\}, \rho_5, \varphi_5, \mathscr{P}_5\big)$, where in matrix notation, $\rho_5 = \big[\rho_5(i,j)\big]_{i,j\in\{1,2\}}$ and $\varphi_5 = \big[\varphi_5(i,j)\big]_{i,j\in\{1,2\}}$ are given by

$$\rho_5 := \begin{bmatrix} 1/2 & 3/8 \\ 3/8 & 1/2 \end{bmatrix}, \qquad \varphi_5 := \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

and $\mathscr{P}_5$ is the set of all measures $\tilde{\boldsymbol{\mu}}_\varepsilon$ of the form $\tilde{\boldsymbol{\mu}}_\varepsilon = \otimes_{t=1}^\infty\big(\frac{1+\varepsilon}{2}\delta_1 + \frac{1-\varepsilon}{2}\delta_2\big)$ for some $\varepsilon \in [-1,1]$, where δ_i is the Dirac measure at $i \in \{1,2\}$. Thus, letting $\mathcal{S}_4$ and $\mathcal{S}_5$ be the two sets of scenarios in games $\mathscr{G}_4$ and $\mathscr{G}_5$ respectively (note that $\mathcal{S}_4$ coincides with the set of scenarios of $\mathscr{G}_1$) and using again the Embedding lemma,

this time with $\tilde{f}\colon [0,1] \to \{1,2\}$, $p \mapsto \mathbb{I}\{p \le 1/2\} + 2\mathbb{I}\{p > 1/2\}$, $g\colon \{0,1\} \to \{0,1\}$, $i \mapsto i$, and $h\colon \mathcal{S}_5 \to \mathcal{S}_4$, $\tilde{\boldsymbol{\mu}}_\varepsilon \otimes \boldsymbol{\mu}_L \mapsto \boldsymbol{\mu}_\varepsilon \otimes \boldsymbol{\mu}_L$, we obtain that $\mathcal{G}_5$ is easier than $\mathcal{G}_4$ (i.e., $R_T^\star(\mathcal{G}_4) \ge R_T^\star(\mathcal{G}_5)$). This last game $\mathcal{G}_5$ is a two-armed bandit problem with gap $\Delta = 1/2 - 3/8 = 1/8$, whose minimax regret is known to be lower bounded by $\frac{1}{8}\left(\frac{1}{20}\sqrt{T}\right)$ Cesa-Bianchi and Lugosi (2006); Bubeck and Cesa-Bianchi (2012). In conclusion, we proved that $R_T^\star(\mathcal{G}) \ge R_T^\star(\mathcal{G}_5) \ge \frac{1}{160}\sqrt{T}$. □

D Proof of $T^{2/3}$ Lower Bound Under Realistic Feedback (iv+bd)

In this section we give a detailed proof of our $T^{2/3}$ lower bound of Section 4.2. which hinges in a non-trivial way on our Embedding and Simulation lemmas (Lemmas B.1 and B.2). We denote Bernoulli distributions with parameter λ by Ber_λ.

Theorem (Theorem 4.2)**.** *In the realistic-feedback stochastic (iid) setting with independent valuations (iv) and densities bounded by a constant $M \ge 24$ (bd), for all horizons $T \in \mathbb{N}$, the minimax regret satisfies*

$$R_T^\star \ge \frac{11}{672} T^{2/3} .$$

Proof. Fix an arbitrary horizon $T \in \mathbb{N}$ and any $M \ge 24$. Recalling Appendix A.2, the realistic-feedback stochastic (iid) setting with independent valuations (iv) and densities bounded (bd) by M is a game $\mathcal{G} := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathcal{P})$, where $\mathcal{X} = [0,1]$, $\mathcal{Y} = [0,1]^2$, $\mathcal{Z} = \{0,1\}^2$, $\rho = \mathrm{GFT}$, $\varphi\colon \big(p,(s,b)\big) \mapsto \big(\mathbb{I}\{s \le p\}, \mathbb{I}\{p \le b\}\big)$, and $\mathcal{P} = \mathcal{P}^M_{\text{iv+bd}}$. The idea of the proof is to build a sequence of games, each one easier than the former, the last of which has a known lower bound on its minimax regret. In the first step we limit the adversary's behavior to a parametric family which is easily manageable and well-represents the difficulty of the problem (see Fig. 2). In the second step, we increase the reward of suboptimal actions in order to have only three possible expected-reward values in each scenario. In the third and fifth steps we increase the feedback, presenting it in a way that highlights that only its first component is informative. In step four and six, we simulate-away the uninformative parts of the feedback. Finally, in step 7 we show that the resulting game is harder than a known partial monitoring game with minimax regret of order at least $T^{2/3}$.

Step 1 Let $\vartheta := 1/48$. Define the following densities of the seller and buyer, respectively, by

$$f_{S,\varepsilon} := \frac{1}{4\vartheta}\Big((1+\varepsilon)\mathbb{I}_{[0,\vartheta]} + (1-\varepsilon)\mathbb{I}_{\left[\frac{1}{6},\frac{1}{6}+\vartheta\right]} + \mathbb{I}_{\left[\frac{1}{4},\frac{1}{4}+\vartheta\right]} + \mathbb{I}_{\left[\frac{2}{3},\frac{2}{3}+\vartheta\right]}\Big)\,, \forall \varepsilon \in [-1,1]\,, \qquad \text{(red/blue in Fig. 2)}$$

$$f_B := \frac{1}{4\vartheta}\Big(\mathbb{I}_{\left[\frac{1}{3}-\vartheta,\frac{1}{3}\right]} + \mathbb{I}_{\left[\frac{3}{4}-\vartheta,\frac{3}{4}\right]} + \mathbb{I}_{\left[\frac{5}{6}-\vartheta,\frac{5}{6}\right]} + \mathbb{I}_{[1-\vartheta,1]}\Big)\,. \qquad \text{(green in Fig. 2)}$$

Define $\mathcal{P}_1$ as the subset of $\mathcal{P}$ whose elements have the form $\boldsymbol{\mu}_\varepsilon := \otimes_{t\in\mathbb{N}}(f_{S,\varepsilon}\mu_L \otimes f_B\mu_L)$ for $\varepsilon \in [-1,1]$. Since $\mathcal{P}_1 \subset \mathcal{P}$, the game $\mathcal{G}_1 := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathcal{P}_1)$ is easier than $\mathcal{G}$ (i.e., $R_T^\star(\mathcal{G}) \ge R_T^\star(\mathcal{G}_1)$) by the Embedding lemma (Lemma B.1) with $\tilde{f}$ and g as the identities, and h as the inclusion.

Step 2 Define $\rho_2\colon \mathcal{X}\times\mathcal{Y} \to [0,1]$, $\big(p,(s,b)\big) \mapsto \mathrm{GFT}\big(\frac{1}{6}+\vartheta,(s,b)\big)\mathbb{I}\big\{p < \frac{1}{4}\big\} + \mathrm{GFT}\big(\frac{1}{4}+\vartheta,(s,b)\big)\mathbb{I}\big\{\frac{1}{4} \le p < \frac{1}{3}\big\} + \mathrm{GFT}\big(\frac{2}{3}+\vartheta,(s,b)\big)\mathbb{I}\big\{\frac{1}{3} < p\big\}$. By the Embedding lemma with $\tilde{f}$, g, and h as the identities, we have that the game $\mathcal{G}_2 := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho_2, \varphi, \mathcal{P}_1)$ is easier than $\mathcal{G}_1$ (i.e., $R_T^\star(\mathcal{G}_1) \ge R_T^\star(\mathcal{G}_2)$).

Step 3 Define $\mathcal{Z}_3 := \big\{0, \frac{1}{6}, \frac{1}{4}, \frac{2}{3}\big\} \times [0,\vartheta] \times \{0,1\} \times \{0,1\} \times \mathcal{X}$ and $\varphi_3\colon \mathcal{X}\times\mathcal{Y} \to \mathcal{Z}_3$,

$$\big(p,(s,b)\big) \mapsto \begin{cases} \big(\eta(s), s-\eta(s), 0, \mathbb{I}\{p \le b\}, p\big)\,, & \text{if } p < \frac{1}{4}\,, \\ \big(0,0,\mathbb{I}\{s \le p\}, \mathbb{I}\{p \le b\}, p\big)\,, & \text{if } p \ge \frac{1}{4}\,, \end{cases}$$

where $\eta\colon [0,1] \to \{0, \frac{1}{6}, \frac{1}{4}, \frac{2}{3}\}$, $s \mapsto \frac{1}{6}\mathbb{I}\{\frac{1}{6} \le s \le \frac{1}{6} + \vartheta\} + \frac{1}{4}\mathbb{I}\{\frac{1}{4} \le s \le \frac{1}{4} + \vartheta\} + \frac{2}{3}\mathbb{I}\{\frac{2}{3} \le s \le \frac{2}{3} + \vartheta\}$. Define the game $\mathscr{G}_3 := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}_3, \rho_2, \varphi_3, \mathscr{P}_1)$. By the Embedding lemma with $\tilde{\ell}, \hbar$ as the identities and

$$\mathscr{g}\colon \mathcal{Z}_3 \to \mathcal{Z}\,, \quad (v,u,i,j,p) \mapsto \begin{cases} \big(\mathbb{I}\{v+u \le p\}, j\big) & \text{if } p < \frac{1}{4}\,, \\ (i,j)\,, & \text{if } p \ge \frac{1}{4}\,, \end{cases}$$

we have that the game $\mathscr{G}_3$ is easier than $\mathscr{G}_2$ (i.e., $R_T^\star(\mathscr{G}_2) \ge R_T^\star(\mathscr{G}_3)$).

Step 4 Let $\mathcal{Z}_4 := \{0, \frac{1}{6}, \frac{1}{4}, \frac{2}{3}\}$ and $\varphi_4\colon \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}_4$, $\big(p, (s,b)\big) \mapsto \eta(s)\mathbb{I}\{p < \frac{1}{4}\}$. Define the game $\mathscr{G}_4 := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}_4, \rho_2, \varphi_4, \mathscr{P}_1)$. Let $(Y_t)_{t\in\mathbb{N}} = (S_t, B_t)_{t\in\mathbb{N}}$ be the adversary's actions in $\mathscr{G}_4$, $E := \big[0,\vartheta\big] \cup \big[\frac{1}{6}, \frac{1}{6}+\vartheta\big] \cup \big[\frac{1}{4}, \frac{1}{4}+\vartheta\big] \cup \big[\frac{2}{3}, \frac{2}{3}+\vartheta\big]$ and $F := \big[\frac{1}{3}-\vartheta, \frac{1}{3}\big] \cup \big[\frac{3}{4}-\vartheta, \frac{3}{4}\big] \cup \big[\frac{5}{6}-\vartheta, \frac{5}{6}\big] \cup \big[1-\vartheta, 1\big]$. A long and tedious computation verifies that for all $t \in \mathbb{N}$,

- for each $p \in [0, 1/4)$ and any scenario $\mathbb{P}$ of game $\mathscr{G}_3$, $\mathbb{P}_{\varphi_3(p,Y_t)} = \mathbb{P}_{\eta(S_t)} \otimes (\nu \otimes \delta_0 \otimes \mathrm{Ber}_{\lambda_{F,p}} \otimes \delta_p)$, where ν is the uniform distribution on $[0,\vartheta]$ and $\lambda_{F,p} := \frac{1}{4\vartheta}\mu_L\big[[p,1] \cap F\big]$. By the well-known Skorokhod representation (Williams, 1991, Section 17.3), there exists $\psi_p\colon [0,1] \to [0,\vartheta] \times \{0,1\} \times \{0,1\} \times \mathcal{X}$ such that $\nu \otimes \delta_0 \otimes \mathrm{Ber}_{\lambda_{F,p}} \otimes \delta_p = (\mu_L)_{\psi_p}$;
- for each $p \in [1/4, 1]$ and any scenario $\mathbb{P}$ of game $\mathscr{G}_3$, $\mathbb{P}_{\varphi_3(p,Y_t)} = \delta_0 \otimes \delta_0 \otimes \mathrm{Ber}_{\lambda_{E,p}} \otimes \mathrm{Ber}_{\lambda_{F,p}} \otimes \delta_p$, where $\lambda_{E,p} := \frac{1}{4\vartheta}\mu_L\big[[0,p] \cap E\big]$ and $\lambda_{F,p} := \frac{1}{4\vartheta}\mu_L\big[[p,1] \cap F\big]$. By the Skorokhod representation, there exists $\gamma_p\colon [0,1] \to \mathcal{Z}_3$ such that $\delta_0 \otimes \delta_0 \otimes \mathrm{Ber}_{\lambda_{E,p}} \otimes \mathrm{Ber}_{\lambda_{F,p}} \otimes \delta_p = (\mu_L)_{\gamma_p}$.

Thus, by the Simulation lemma (Lemma B.2) with $\mathcal{I} = [0, 1/4)$ and $\mathcal{U} = [1/4, 1]$, the game $\mathscr{G}_4$ is easier than $\mathscr{G}_3$ (i.e., $R_T^\star(\mathscr{G}_3) \ge R_T^\star(\mathscr{G}_4)$).

Step 5 Let $\mathcal{Y}_5 := \mathcal{Y}^{\mathbb{N}}$, $\mathcal{Z}_5 := \{0,1\} \times \big(\mathbb{N} \cup \{\infty\}\big) \times \{0,1\} \times \mathcal{X}$, $\rho_5\colon \mathcal{X} \times \mathcal{Y}_5 \to [0,1]$, $\big(p, (s_k, b_k)_{k\in\mathbb{N}}\big) \mapsto \rho_2(p, s_1, b_1)$,

$$\varphi_5\colon \mathcal{X} \times \mathcal{Y}_5 \to \mathcal{Z}_5\,, \quad \big(p, (s_k,b_k)_{k\in\mathbb{N}}\big) \mapsto \begin{cases} \Big(\mathbb{I}\big\{\eta(s_\tau) = 0\big\}, \tau, \mathbb{I}\big\{\eta(s_1) = \frac{1}{4}\big\}, p\Big)\,, & \text{if } p \in \big[0, \frac{1}{4}\big)\,, \\ (0,1,0,p)\,, & \text{if } p \in \big[\frac{1}{4}, 1\big]\,, \end{cases}$$

where η is defined in game $\mathscr{G}_3$, $\tau := \inf\big\{k \in \mathbb{N} \mid \eta(s_k) \in \{0, 1/6\}\big\} \in \mathbb{N} \cup \{\infty\}$, and $s_\infty := 0$. Let $\mathscr{P}_5$ be the set of measures on $\mathcal{Y}_5^{\mathbb{N}}$ of the form $\tilde{\boldsymbol{\mu}}_\varepsilon := \otimes_{t\in\mathbb{N}}\big(\otimes_{k\in\mathbb{N}}(f_{S,\varepsilon}\mu_L \otimes f_B\mu_L)\big)$ for $\varepsilon \in [-1,1]$, and define the game $\mathscr{G}_5 := (\mathcal{X}, \mathcal{Y}_5, \mathcal{Z}_5, \rho_5, \varphi_5, \mathscr{P}_5)$. By the Embedding lemma with $\tilde{\ell}$ as the identity,

$$\mathscr{g}\colon \mathcal{Z}_5 \to \mathcal{Z}_4\,, \quad (z,k,j,p) \mapsto \frac{1}{6}(1-z)\mathbb{I}\Big\{p < \frac{1}{4}, k = 1\Big\} + \Big(\frac{1}{4}j + \frac{2}{3}(1-j)\Big)\mathbb{I}\Big\{p < \frac{1}{4}, k > 1\Big\}\,,$$

and $\hbar\colon \tilde{\boldsymbol{\mu}}_\varepsilon \otimes \boldsymbol{\mu}_L \mapsto \boldsymbol{\mu}_\varepsilon \otimes \boldsymbol{\mu}_L$, we have that the game $\mathscr{G}_5$ is easier than $\mathscr{G}_4$ (i.e., $R_T^\star(\mathscr{G}_4) \ge R_T^\star(\mathscr{G}_5)$).

Step 6 Now, define $\pi\colon \mathcal{Z}_5 \to \{0,1\}$ as the projection on the first component $\{0,1\}$ of $\mathcal{Z}_5$, $\mathcal{Z}_6 := \{0,1\}$, $\varphi_6 := \pi \circ \varphi_5$, and the game $\mathscr{G}_6 := (\mathcal{X}, \mathcal{Y}_5, \mathcal{Z}_6, \rho_5, \varphi_6, \mathscr{P}_5)$. Let $(\tilde{Y}_t)_{t\in\mathbb{N}}$ be the adversary's actions in $\mathscr{G}_5$. A straightforward verification shows that for all $t \in \mathbb{N}$,

- for each $p \in [0, 1/4)$ and any scenario $\mathbb{P}$ of game $\mathscr{G}_5$, $\mathbb{P}_{\varphi_5(p,\tilde{Y}_t)} = \mathbb{P}_{\pi\left(\varphi_5(p,\tilde{Y}_t)\right)} \otimes (\nu \otimes \delta_p)$, where ν is the unique distribution on $\big(\mathbb{N} \cup \{\infty\}\big) \times \{0,1\}$ such that, for all $k \in \mathbb{N} \cup \{\infty\}$, $j \in \{0,1\}$, $\nu\big[\{(k,j)\}\big] = \frac{1}{2}\mathbb{I}\{k = 1, j = 0\} + \frac{1}{2^{k+1}}\mathbb{I}\{1 < k < \infty\}$. Using again the Skorokhod representation, there exists $\psi_p\colon [0,1] \to \big(\mathbb{N} \cup \{\infty\}\big) \times \{0,1\} \times [0,1]$ such that $\nu \otimes \delta_p = (\mu_L)_{\psi_p}$;
- for each $p \in [1/4, 1]$ and any scenario $\mathbb{P}$ of game $\mathscr{G}_5$, $\mathbb{P}_{\varphi_5(p,\tilde{Y}_t)} = \delta_{(0,1,0,p)} = (\mu_L)_{\gamma_p}$, where $\gamma_p\colon [0,1] \to \mathcal{Z}_5$, $\lambda \mapsto (0,1,0,p)$.

Thus, by the Simulation lemma with $\mathcal{I} = [0, 1/4)$ and $\mathcal{U} = [1/4, 1]$, the game $\mathscr{G}_6$ is easier than $\mathscr{G}_5$ (i.e., $R_T^\star(\mathscr{G}_5) \ge R_T^\star(\mathscr{G}_6)$).

Step 7 Finally, consider the game $\mathscr{G}_7 := \big(\{1,2,3\},\{1,2\},\{0,1\},\rho_7,\varphi_7,\mathscr{P}_7\big)$, where in matrix notation, $\rho_7 = \big[\rho(i,j)\big]_{i\in\{1,2,3\},j\in\{1,2\}}$ and $\varphi_7 = \big[\varphi(i,j)\big]_{i\in\{1,2,3\},j\in\{1,2\}}$ are given by

$$\rho_7 := \frac{1}{96}\begin{bmatrix}34 & 34\\ 45 & 37\\ 38 & 44\end{bmatrix}, \qquad \varphi_7 := \begin{bmatrix}1 & 0\\ 0 & 0\\ 0 & 0\end{bmatrix},$$

and $\mathscr{P}_7$ is the set of all measures of the form $\otimes_{t\in\mathbb{N}}\big(\frac{1+\varepsilon}{2}\delta_1 + \frac{1-\varepsilon}{2}\delta_2\big)$, for $\varepsilon \in [-1,1]$. Thus, using again the Embedding lemma, this time with $\tilde{\mathscr{f}}\colon [0,1] \to \{1,2,3\}$, $p \mapsto \mathbb{I}\{p < 1/4\} + 2\mathbb{I}\{1/4 \le p \le 1/3\} + 3\mathbb{I}\{1/3 < p\}$, $\mathscr{g}\colon \{0,1\} \to \{0,1\}$, $i \mapsto i$, and $\hbar\colon \otimes_{t\in\mathbb{N}}\big(\frac{1+\varepsilon}{2}\delta_1 + \frac{1-\varepsilon}{2}\delta_2\big) \otimes \boldsymbol{\mu}_L \mapsto \tilde{\boldsymbol{\mu}}_\varepsilon \otimes \boldsymbol{\mu}_L$, we obtain that $\mathscr{G}_7$ is easier than $\mathscr{G}_6$ (i.e., $R_T^\star(\mathscr{G}_6) \ge R_T^\star(\mathscr{G}_7)$). This last game is an instance of the so-called revealing action partial monitoring game, whose minimax regret is known to be lower bounded by $\frac{11}{96}\big(\frac{1}{7}T^{2/3}\big)$ Cesa-Bianchi et al. (2006). In conclusion, we proved that $R_T^\star(\mathscr{G}) \ge R_T^\star(\mathscr{G}_7) \ge \frac{11}{672}T^{2/3}$. □

E Linear Lower Bound Under Realistic Feedback (bd)

In this section, we prove that in the realistic-feedback case, no strategy can achieve sublinear regret in the worst case if the valuations of the buyer and the seller may be dependent, not even if they have a bounded density.

The idea of the proof is to exploit the lack of observability in this setting, building a family of scenarios $\mathbb{P}^\lambda$ (parameterized by $\lambda \in [0,1]$) as convex combinations of the two measures in Fig. 3. If $\lambda < 1/2$, the optimal action is $3/8$, while if $\lambda > 1/2$, the optimal action becomes $5/8$. This family is built is such a way that the feedback gives no information on λ, making it impossible to distinguish between the two cases. Leveraging the Embedding and Simulation lemmas (Lemmas B.1 and B.2), this construction leads to a reduction to an instance of a non-observable partial monitoring game, whose regret is trivially lower bounded by $T/24$.

Theorem (Theorem 4.3)**.** *In the realistic-feedback stochastic (iid) setting with joint density bounded by a constant $M \ge 64/3$ (bd), for all horizons $T \in \mathbb{N}$, the minimax regret satisfies*

$$R_T^\star \ge \frac{1}{24}T\,.$$

Proof. Fix any horizon $T \in \mathbb{N}$ and $M \ge 64/3$. Recalling Appendix A.2, the realistic-feedback stochastic (iid) setting with joint density bounded by M (bd) is a game $\mathscr{G} := (\mathcal{X},\mathcal{Y},\mathcal{Z},\rho,\varphi,\mathscr{P})$, where $\mathcal{X} = [0,1]$, $\mathcal{Y} = [0,1]^2$, $\mathcal{Z} = \{0,1\}^2$, $\rho = \mathrm{GFT}$, $\varphi\colon \big(p,(s,b)\big) \mapsto \big(\mathbb{I}\{s \le p\}, \mathbb{I}\{p \le b\}\big)$, and $\mathscr{P} = \mathscr{P}_{\mathrm{bd}}^M$. Define the two joint densities $f = \frac{64}{3}\big(\mathbb{I}_{[0/8,1/8]\times[3/8,4/8]} + \mathbb{I}_{[2/8,3/8]\times[7/8,8/8]} + \mathbb{I}_{[4/8,5/8]\times[5/8,6/8]}\big)$ and $g\colon [0,1]^2 \to [0,M]$, $(s,b) \mapsto f(1-b,1-s)$ (see Fig. 3, left). Let $\mathscr{P}_1$ be the subset of $\mathscr{P}_{\mathrm{bd}}^M$ whose elements have the form $\boldsymbol{\mu}_\lambda := \otimes_{t\in\mathbb{N}}\big(\big((1-\lambda)f+\lambda g\big)(\mu_L\otimes\mu_L)\big)$ for $\lambda \in [0,1]$. Since $\mathscr{P}_1 \subset \mathscr{P}$ the game $\mathscr{G}_1 := (\mathcal{X},\mathcal{Y},\mathcal{Z},\rho,\varphi,\mathscr{P}_1)$ is easier than $\mathscr{G}$ (i.e., $R_T^\star(\mathscr{G}) \ge R_T^\star(\mathscr{G}_1)$) by the Embedding lemma (Lemma B.1) with $\tilde{\mathscr{f}}$ and $\mathscr{g}$ as the identities, and $\hbar$ as the inclusion. Define $\mathcal{Z}_1 := \{0\}$ and $\varphi_1\colon \mathcal{X}\times\mathcal{Y} \to \mathcal{Z}_1$, $\big(p,(s,b)\big) \mapsto 0$. Let $(Y_t)_{t\in\mathbb{N}}$ be the adversary's actions in $\mathcal{G}_1$. Now, since for all $t \in \mathbb{N}$, any two scenarios $\mathbb{P}$ and $\mathbb{Q}$ of game $\mathscr{G}_1$, and each $p \in [0,1]$, $\mathbb{P}_{\varphi(p,Y_t)} = \mathbb{Q}_{\varphi(p,Y_t)}$, then by the well-known Skorokhod representation (Williams, 1991, Section 17.3), for each $t \in \mathbb{N}$ and each $p \in [0,1]$ there exists $\gamma_{t,p}\colon [0,1] \to \{0,1\}^2$ such that for any scenario $\mathbb{P}$ of game $\mathscr{G}_1$, $\mathbb{P}_{\varphi(x,Y_t)} = (\mu_L)_{\gamma_{t,x}}$. Thus, the Simulation lemma (Lemma B.2) with $\mathcal{I} = \varnothing$ and $\mathcal{U} = \mathcal{X}$ implies that the game $\mathscr{G}_2 := (\mathcal{X},\mathcal{Y},\mathcal{Z}_2,\rho,\varphi_2,\mathscr{P}_1)$ is easier than $\mathscr{G}_1$ (i.e., $R_T^\star(\mathscr{G}_1) \ge R_T^\star(\mathscr{G}_2)$). Define $\rho_3\colon \mathcal{X}\times\mathcal{Y} \to [0,1]$, $\big(p,(s,b)\big) \mapsto (b-s)\mathbb{I}\big\{s \le \frac{3}{8} \le b\big\}\mathbb{I}\big\{p \le \frac{1}{2}\big\} + (b-s)\mathbb{I}\big\{s \le \frac{5}{8} \le b\big\}\mathbb{I}\big\{p > \frac{1}{2}\big\}$ and $\mathscr{G}_3 := (\mathcal{X},\mathcal{Y},\mathcal{Z}_2,\rho_3,\varphi_2,\mathscr{P}_1)$. By the Embedding lemma with $\tilde{\mathscr{f}},\mathscr{g},\hbar$ as the identities, we have that the game $\mathscr{G}_3$ is easier than the game $\mathscr{G}_2$ (i.e., $R_T^\star(\mathscr{G}_2) \ge R_T^\star(\mathscr{G}_3)$). Finally, consider the game $\mathscr{G}_4 := \big(\{1,2\},\{1,2\},\{0\},\rho_4,\varphi_4,\mathscr{P}_4\big)$, where in matrix notation, $\rho_4 = \big[\rho(i,j)\big]_{i,j\in\{1,2\}}$ and $\varphi_4 = \big[\varphi(i,j)\big]_{i,j\in\{1,2\}}$ are given by

$$\rho_4 := \begin{bmatrix}1/3 & 1/4\\ 1/4 & 1/3\end{bmatrix}, \qquad \varphi_4 := \begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix},$$

and $\mathscr{P}_4$ is the set of all measures of the form $(1-\lambda)\delta_1 + \lambda\delta_2$, for $\lambda \in [0,1]$. Using again the Embedding lemma, this time with $\tilde{\ell}\colon [0,1] \to \{1,2\}$, $p \mapsto \mathbb{I}\{p \le 1/2\} + 2\mathbb{I}\{1/2 < p\}$, $g\colon \{0\} \to \{0\}$, $i \mapsto i$, and $\hbar\colon \otimes_{t\in\mathbb{N}} \big((1-\lambda)\delta_1 + \lambda\delta_2\big) \otimes \boldsymbol{\mu}_L \mapsto \boldsymbol{\mu}_\lambda \otimes \boldsymbol{\mu}_L$, we obtain that $\mathscr{G}_4$ is easier than $\mathscr{G}_3$ (i.e., $R_T^\star(\mathscr{G}_3) \ge R_T^\star(\mathscr{G}_4)$). This last game has (trivially) minimax regret at most $\big(\frac{1}{3} - \frac{1}{4}\big)\frac{T}{2}$. In conclusion, we proved that $R_T^\star(\mathscr{G}) \ge R_T^\star(\mathscr{G}_4) \ge \frac{1}{24}T$. □

F Linear Lower Bound Under Realistic Feedback (iv)

In this section, we prove that in the realistic-feedback case, no strategy can achieve sublinear regret without any limitations on how concentrated the distributions of the valuations of the seller and buyer are, not even if they are independent of each other (iv).

The idea of the proof is that if the two distributions are very concentrated in a small region, finding an optimal price is like finding a needle in a haystack. Each strategy that (at each time step) receives as feedback only a finite number of bits, as in our realistic setting, can assign positive probability to at most a countable set of points. Thus one could find concentrated distributions of the buyer and seller that have a unique optimal point in which the strategy has zero probability of posting prices at all time steps, and such that *all* other prices suffer large regret.

Theorem (Theorem 4.4)**.** *In the realistic-feedback stochastic (iid) setting with independent valuations (iv), for all horizons $T \in \mathbb{N}$, the minimax regret satisfies*

$$R_T^\star \ge \frac{1}{8}T\,.$$

Proof. To lighten the notation, for any $n \in \mathbb{N}$ and a family $(\lambda_k)_{k\in\mathbb{N}}$, we let $\lambda_{1:n} := (\lambda_1, \dots, \lambda_n)$. Fix an arbitrary horizon $T \in \mathbb{N}$. Recalling Appendix A.2, the realistic-feedback stochastic (iid) setting with independent valuations (iv) is a game $\mathscr{G} := (\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \rho, \varphi, \mathscr{P})$, where $\mathcal{X} = [0,1]$, $\mathcal{Y} = [0,1]^2$, $\mathcal{Z} = \{0,1\}^2$, $\rho = \mathrm{GFT}$, $\varphi\colon \big(p, (s,b)\big) \mapsto \big(\mathbb{I}\{s \le p\}, \mathbb{I}\{p \le b\}\big)$, and $\mathscr{P} = \mathscr{P}_{\mathrm{iv}}$. Let $\mathcal{S}$ be the set of scenarios of $\mathscr{G}$. Fix a strategy α for game $\mathscr{G}$ and let $\varepsilon \in (0,1)$. Define $\bar{\alpha}_1 := \alpha_1$, $\nu_1 := (\mu_L)_{\bar{\alpha}_1}$, and for each $t \in \mathbb{N}$ and $z_1, \dots, z_t \in \{0,1\}^2$,

$$\bar{\alpha}_{t+1,z_{1:t}}\colon [0,1]^{t+1} \to [0,1], \quad u_{1:t+1} \mapsto \alpha_{t+1}(u_{1:t+1}, z_{1:t}) \qquad \text{and} \qquad \nu_{t+1,z_{1:t}} := (\otimes_{s=1}^{t+1}\mu_L)_{\bar{\alpha}_{t+1,z_{1:t}}}\,.$$

Define also the set $A_1 := \big\{x \in [0,1] \mid \nu_1[\{x\}] > 0\big\}$ and, for each $t \in \mathbb{N}$, the union $A_{t+1} := \bigcup_{z_{1:t}\in\{0,1\}^2} \big\{x \in [0,1] \mid \nu_{t,z_{1:t}}[\{x\}] > 0\big\}$. Note that, for each $t \in \mathbb{N}, A_t$ is countable, being the union of 4^{t-1} countable sets. Then $A := \bigcup_{t\in\mathbb{N}} A_t$ is countable. Since $B := [\frac{1-\varepsilon}{2}, \frac{1+\varepsilon}{2}]$ has the power of continuum, we have that the same holds for $B \setminus A$. In particular, $B \setminus A$ is non-empty. Pick $x^\star \in B \setminus A$ and define $\mu_S := \frac{1}{2}\delta_0 + \frac{1}{2}\delta_{x^\star}$, $\mu_B := \frac{1}{2}\delta_{x^\star} + \frac{1}{2}\delta_1$, and $\mathbb{P} := (\otimes_{t\in\mathbb{N}}(\mu_S \otimes \mu_B)) \otimes \boldsymbol{\mu}_L \in \mathcal{S}$. Then for each $t \in \mathbb{N}$, we have that

$$\mathbb{E}_{\mathbb{P}}\big[\rho(x^\star, Y_t)\big] = \frac{x^\star + (1-x^\star) + 1}{4}\,.$$

On the other hand, $\mathbb{P}[X_1 = x^\star] = \nu_1[\{x^\star\}] = 0$ and for each $t \in \mathbb{N}$, we have that

$$\begin{aligned}
\mathbb{P}[X_{t+1} = x^\star] &= \mathbb{P}\big[\alpha_{t+1}(U_1, \dots, U_{t+1}, Z_1, \dots, Z_t) = x^\star\big] \\
&= \sum_{z_1,\dots,z_t\in\{0,1\}^2} \mathbb{P}\big[\alpha_{t+1}(U_1, \dots, U_{t+1}, z_1, \dots, z_t) = x^\star \cap Z_1 = z_1 \cap \dots \cap Z_t = z_t\big] \\
&\le \sum_{z_1,\dots,z_t\in\{0,1\}^2} \mathbb{P}\big[\alpha_{t+1}(U_1, \dots, U_{t+1}, z_1, \dots, z_t) = x^\star\big] = \sum_{z_1,\dots,z_t\in\{0,1\}^2} \nu_{t+1,z_1,\dots,z_t}\big[\{x^\star\}\big] = 0\,,
\end{aligned}$$

which in turn gives

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\big[\rho(X_t,Y_t)\big] &= \frac{\mathbb{E}_{\mathbb{P}}\big[\rho\big(X_t,(0,x^\star)\big)\big]+\mathbb{E}_{\mathbb{P}}\big[\rho\big(X_t,(x^\star,1)\big)\big]+\mathbb{E}_{\mathbb{P}}\big[\rho\big(X_t,(0,1)\big)\big]+\mathbb{E}_{\mathbb{P}}\big[\rho\big(X_t,(x^\star,x^\star)\big)\big]}{4}\\ &= \frac{x^\star\mathbb{P}_{X_t}\big[[0,x^\star]\big]+(1-x^\star)\mathbb{P}_{X_t}\big[[x^\star,1]\big]+1}{4} = \frac{x^\star\mathbb{P}_{X_t}\big[[0,x^\star)\big]+(1-x^\star)\mathbb{P}_{X_t}\big[(x^\star,1]\big]+1}{4}\\ &\le \frac{\max(x^\star,1-x^\star)+1}{4} = \frac{x^\star+(1-x^\star)+1-\min(x^\star,1-x^\star)}{4}\,.\end{aligned}$$

So, if $T\in\mathbb{N}$ we get

$$R_T^{\mathbb{P}}(\alpha)=\mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^T\rho(x^\star,Y_t)-\sum_{t=1}^T\rho(X_t,Y_t)\right]\ge\frac{\min(x^\star,1-x^\star)}{4}T\ge\frac{1-\varepsilon}{8}T.$$

Since ε was arbitrary, we get, for all $T\in\mathbb{N}$, $R_T^{\mathcal{S}}(\alpha)=\sup_{\mathbb{P}\in\mathcal{S}}R_T^{\mathbb{P}}(\alpha)\ge\sup_{\varepsilon\in(0,1)}\frac{1-\varepsilon}{8}T=T/8$. Since α was arbitrary we get, for each $T\in\mathbb{N}$, $R_T^\star=\inf_{\alpha\in\mathscr{A}}R_T^{\mathcal{S}}(\alpha)\ge T/8$. □

G Adversarial Setting: Linear Lower Bound Under Full Feedback

In this section, we give a more detailed proof of Theorem 5.1 with a notation consistent to our abstract setting of sequential games.

Theorem (Theorem 5.1)**.** *In the full-feedback adversarial (adv) setting, for all horizons $T\in\mathbb{N}$, we have*

$$R_T^\star\ge\frac{1}{4}T\,.$$

Proof. Recalling Appendix A.2, the full-feedback adversarial (adv) bilateral trade setting is a game $\mathscr{G}:=(\mathcal{X},\mathcal{Y},\mathcal{Z},\rho,\varphi,\mathscr{P})$, where $\mathcal{X}=[0,1]$, $\mathcal{Y}=[0,1]^2$, $\mathcal{Z}=[0,1]^2$, $\rho=\mathrm{GFT}$, $\varphi\colon\big(p,(s,b)\big)\mapsto(s,b)$, and $\mathscr{P}=\mathscr{P}_{\mathrm{adv}}$. Let $\mathcal{S}$ be the set of scenarios of $\mathscr{G}$. Fix a strategy $\alpha\in\mathscr{A}$ and an $\varepsilon\in(0,1/18)$. Define $\bar\alpha_1:=\alpha_1$, $\nu_1:=(\mu_L)_{\bar\alpha_1}$, and

$$\begin{cases}c_1:=\frac12-\frac32\varepsilon,\ d_1:=\frac12-\frac12\varepsilon,\ s_1:=0,\ b_1:=d_1, & \text{if } \nu_1\big[\big[0,\frac12-\frac12\varepsilon\big]\big]\le\frac12\,,\\ c_1:=\frac12+\frac12\varepsilon,\ d_1:=\frac12+\frac32\varepsilon,\ s_1:=c_1,\ b_1:=1, & \text{otherwise.}\end{cases}$$

If $t\in\mathbb{N}$, suppose we defined $\bar\alpha_t,\nu_t,c_t,d_t,s_t,b_t$ and let

$$\bar\alpha_{t+1}:[0,1]^{t+1}\to[0,1],(u_1,\dots,u_{t+1})\mapsto\alpha_{t+1}(u_1,\dots,u_{t+1},(s_1,b_1),\dots,(s_t,b_t)),$$

$\nu_{t+1}:=\big(\otimes_{s=1}^{t+1}\mu_L\big)_{\bar\alpha_{t+1}}$, and

$$\begin{cases}c_{t+1}:=c_t,\ d_{t+1}:=d_t-\frac{2\varepsilon}{3^t},\ s_{t+1}:=0,\ b_{t+1}:=d_{t+1}, & \text{if } \nu_{t+1}\big[\big[0,c_t+\frac{\varepsilon}{3^t}\big]\big]\le\frac12\,,\\ c_{t+1}:=c_t+\frac{2\varepsilon}{3^t},\ d_{t+1}:=d_t,\ s_{t+1}:=c_{t+1},\ b_{t+1}:=1, & \text{otherwise.}\end{cases}$$

Then the sequences $(\bar\alpha_t)_{t\in\mathbb{N}},(\nu_t)_{t\in\mathbb{N}},(c_t)_{t\in\mathbb{N}},(d_t)_{t\in\mathbb{N}},(s_t)_{t\in\mathbb{N}},(b_t)_{t\in\mathbb{N}}$ are well-defined by induction and satisfy:

- for each $t\in\mathbb{N}, d_t-c_t=\frac{\varepsilon}{3^{t-1}}$;
- for each $t\in\mathbb{N}, c_1\le c_2\le c_3\le\cdots\le c_t\le d_t\le\cdots\le d_3\le d_2\le d_1$;
- $\exists!x^\star\in\bigcap_{t=1}^\infty[c_t,d_t]$;
- for each $t\in\mathbb{N},\rho\big(x^\star,(s_t,b_t)\big)=b_t-s_t\ge\frac{1-3\varepsilon}{2}$;
- for each $t\in\mathbb{N},\mathbb{P}\big[\alpha_t\big(U_1,\dots,U_t,(s_1,b_1),\dots,(s_{t-1},b_{t-1})\big)\in[s_t,b_t]\big]\le\frac12$

Now, define $\mathbb{P} := \left(\otimes_{t\in\mathbb{N}}\delta_{(s_t,b_t)}\right)\otimes \boldsymbol{\mu}_L \in \mathcal{S}$. Then, for each $t \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}[\rho(X_t,Y_t)] &= \mathbb{E}_{\mathbb{P}}\Big[\rho\Big(\alpha_t\big(U_1,\ldots,U_t,(s_1,b_1),\ldots,(s_{t-1},b_{t-1})\big),(s_t,b_t)\Big)\Big]\\ &\leq \left(\frac{1}{2}+\frac{3\varepsilon}{2}\right)\mathbb{P}\big[\alpha_t\big(U_1,\ldots,U_t,(s_1,b_1),\ldots,(s_{t-1},b_{t-1})\big)\in[s_t,b_t]\big] \leq \frac{1}{4}+\frac{3\varepsilon}{4}\,,\end{aligned}$$

and so, for each $T \in \mathbb{N}$

$$\begin{aligned}R_T^{\mathbb{P}}(\alpha) &= \mathbb{E}_{\mathbb{P}}\left[\sum_{t=1}^{T}\rho(x^\star,Y_t)-\sum_{t=1}^{T}\rho(X_t,Y_t)\right] = \sum_{t=1}^{T}\rho(x^\star,(s_t,b_t))-\sum_{t=1}^{T}\mathbb{E}_{\mathbb{P}}[\rho(X_t,Y_t)]\\ &\geq \sum_{t=1}^{T}(b_t-s_t)\big(1-\mathbb{P}\big[\alpha_t\big(U_1,\ldots,U_t,(s_1,b_1),\ldots,(s_{t-1},b_{t-1})\big)\in[s_t,b_t]\big]\big) \geq \frac{1-3\varepsilon}{4}T\,.\end{aligned}$$

Since ε was arbitrary, we get, for all $T \in \mathbb{N}$, $R_T^{\mathcal{S}}(\alpha) = \sup_{\mathbb{P}\in\mathcal{S}} R_T^{\mathbb{P}}(\alpha) \geq \sup_{\varepsilon\in(0,1/18)}\frac{1-3\varepsilon}{4}T = \frac{T}{4}$. Since α arbitrarity, we get, for each $T \in \mathbb{N}$, $R_T^\star = \inf_{\alpha\in\mathscr{A}} R_T^{\mathcal{S}}(\alpha) \geq \frac{T}{4}$. □