

Online trading as a secretary problem^{*}

Elias Koutsoupias and Philip Lazos

University of Oxford

Abstract. We consider the online problem in which an intermediary trades identical items with a sequence of n buyers and n sellers, each of unit demand. We assume that the values of the traders are selected by an adversary and the sequence is randomly permuted. We give competitive algorithms for two objectives: welfare and gain-from-trade.

1 Introduction

We study the problem of facilitating trade between n buyers and n sellers that arrive online. We consider one of the simplest settings in which each trader, buyer or seller, is interested in trading a single item, and all items are identical. Each trader has a value for the item; a seller will sell to any price higher than its value and a buyer will buy for any price lower than its value. Upon encountering a trader, the online algorithm makes an irrevocable price offer, the trader reveals its value and, if the value is at the correct side of the offered price the item is traded. After buying an item from a seller, the online algorithm can store it indefinitely to sell it to later buyers. Of course, the online algorithm can only sell to a buyer if it has at least one item at the time of the encounter.

We consider online algorithms that offer prices based on the sequence of past values and we assume that the online algorithm knows only the number of buyers and sellers, but not their values. The values of the sellers and buyers are selected *adversarially and are randomly permuted*. In that respect, the problem is a generalization of the well-known secretary problem. The secretary problem corresponds to the special case in which there are only buyers, the algorithm starts with a single item, and the objective is to maximize the total welfare, which is to give the value to a buyer with as high value as possible.

Extending this to both sellers and buyers, creates a substantially richer setting. One of the most important differences between the two settings is that besides the objective of maximizing the total welfare, we now have the objective of maximizing the gain-from-trade. For both objectives, the algorithm must buy from sellers with low values and sell to buyers with high values. The objective is that at the end, the items end up at the hands of the traders, sellers or buyers, with the highest values. The welfare of a solution is defined as the value of the buyers and sellers that have an item. The gain-from-trade of a solution is the difference between the welfare at the end of the process minus the welfare at the beginning. At optimality the two objectives are interchangeable: an algorithm achieves the maximum welfare if and only if it achieves the maximum gain-from-trade. But for approximate solutions, the two objectives are entirely different, with the gain-from-trade being the most demanding one.

The Bayesian version of the problem, in which the values of the buyers and sellers are drawn from known probability distributions has been extensively considered in the literature. Optimal mechanisms for bilateral trading, that is, the offline case of a single seller and a single buyer, were first analysed by Myerson and Satterthwaite in [18] and played a pivotal role in the development of the area (see the section Related Work). The online Bayesian case was considered in [10], where the values are drawn from a known distribution but the sequence is adversarially ordered.

A generalization of our model is when the items are not identical and each buyer has different value for each one of them, i.e., each seller has a value for its item and each buyer has a vector of values, one for every pair buyer-seller. This is also a generalization of the well-studied online maximum-matching problem [15,13]. One can cast the online maximum-matching problem as the version in which the

^{*} Supported by the ERC Advanced Grant 321171 (ALGAME).

sellers arrive first and have zero value for their item. The optimal online algorithm for this problem has competitive ratio $1/e$, when the objective is the welfare (which in the absence of seller values is identical to the gain-from-trade). Our model is incomparable to the online maximum-matching problem: it is simpler in the sense that the items are identical (a single value for each buyer instead of a vector of buyer-item values), and at the same time more complicated in that the items are not present throughout the process, but they are brought to the market by sellers that have their own utility. The fact that in our model the buyer-item values are related, allows for a much better competitive ratio regarding the welfare, (almost) 1 instead of $1/e$. More importantly, our algorithm is truthful, while in contrast, no good truthful algorithm is known for the online maximum-matching problem, which remains one of the main open problems of the area. On the other hand, the introduction of sellers poses new challenges, especially with respect to the objective of the gain-from-trade.

There are also similarities between our model and the extension of the classical secretary problem to k secretaries. From an influential result by Kleinberg [14] we know that this problem has competitive ratio $1 - 1/\sqrt{k}$ which is asymptotically tight, and can be transformed into a truthful algorithm. This result depends strongly on the knowledge of k . In our case the equivalent measure, the *number of trades* is not known from the beginning and has to be learned, with a degree of precision that is crucial, especially for the gain-from-trade objective. The fact that the gain-from-trade is not monotone as a function of time highlights the qualitative difference between the two models; the gain-from-trade temporarily *decreases* when the algorithm buys an item, with the risk of having a negative gain at the end. More generally, with the mix of buyers and sellers, wrong decisions are penalized more harshly and the monotone structure of the problem is disrupted.

Our results We consider the case when both the number of buyers and the number of sellers is n . For welfare we show a competitive ratio of $1 - \tilde{O}(n^{-1/3})$, where \tilde{O} hides logarithmic factors.

Actually we can compare an online algorithm with two offline benchmarks: the *optimal* benchmark, in which all trades between buyers and sellers are possible, independently of their order of appearance, and the expected *sequential optimal* in which an item can be transferred from a seller to a buyer only if the seller precedes the buyer in the order.

Our online algorithm achieves a competitive ratio of $1 - \tilde{O}(n^{-1/3})$ against the optimal benchmark. To achieve this, it has a small sampling phase of length $\tilde{O}(n^{2/3})$ to estimate the *median* of the values of all traders, and then uses it as a price for the remaining traders. But if the optimal number of trades is small, such a scheme will fail to achieve competitive ratio almost one, because with constant probability there will not have enough items to sell to buyers with high value. To deal with this risk, the algorithm not only samples values at the beginning but it additionally buys sufficiently many items, $\tilde{O}(n^{2/3})$, from the first sellers¹. The number $\tilde{O}(n^{2/3})$ of bought items balances the potential loss of the welfare that results from removing items from sellers to the expected loss from not having enough items for buyers of high values.

The term $O(n^{-1/3})$ in the competitive ratio seems to be optimal for a scheme that fixes the price after the sampling phase and relates to the number of items needed to approximate the median to a good degree. It may be possible to improve this term to $O(n^{-1/2})$ by a more adaptive scheme, as in the case of the k -secretary problem [14]. Finally, it may be possible to remove the logarithmic factors from the competitive ratio, but we have opted for simplicity and completeness.

For the objective of gain-from-trade, we give a truthful algorithm that has a constant competitive ratio, assuming that the algorithm starts with an item. The competitive ratio is high, approximately 10^3 , but it drops to a small constant when the optimal number of trades is sufficiently high. The additional assumption of starting with an item is necessary, because without it, no online algorithm can achieve a bounded competitive ratio.

¹ Buying from the first sellers cannot be done truthfully unless the algorithm knows an upper bound on their value. But this is not necessary since there is an alternative that has minor effects on the competitive ratio: the algorithm offers each seller the maximum value of the sellers so far. This is a truthful scheme that buys from all but a logarithmic number of sellers, in expectation.

The main difficulty of designing an online algorithm for gain-from-trade is that even a single item that is left unsold at the end has dramatic effects on the gain-from-trade. The online algorithm must deal with the case of many traders, large welfare, but few optimal trades and small gain-from-trade.

To address this problem, our algorithm, unlike the case of welfare, has a large sampling phase. It uses this phase to estimate the number of optimal trades and two prices for trading with buyers and sellers. If the expected number of optimal trades is high, the algorithm uses the two prices for trading with the remaining traders. But if the number is small, it runs the secretary algorithm with the item that it starts with.

The analysis needs high concentration bounds on the expected number of trades to minimize the risk of having items left unsold. Our algorithm is ordinal, in the sense that it uses only the order statistics of the values not the actual values themselves. This leaves little space for errors and it may be possible that cardinal algorithms that use the actual values can do substantially better.

2 Related Work

The bilateral trade literature was initiated by Myerson and Satterthwaite in their seminal paper [18]. They investigated the case of a single seller-buyer pair and proved their famous impossibility result: there exists no truthful, individually rational and budget balanced mechanism that also maximizes the welfare (and consequently, the gain from trade). Subsequent research studied how well these two objectives can be approximated by relaxing these conditions. Blumrosen and Mizrahi [2] devised a $1/e$ -approximate, Bayesian incentive compatible mechanism for the gain from trade assuming the buyer's valuation is monotone hazard rate. Brustle et al. expanded in this direction in [3] for arbitrary valuations and downwards closed feasibility constraints over the allocations. In the case where there are multiple, unit demand, buyers and sellers, McAfee provided a weakly budget balanced, $1 - 1/k$ approximate mechanism for the gain from trade in [16], where k is the number of trades in the optimal allocation. This was later extended to be strongly budget balanced by Segal-Halevi et al. in [19]. McAfee also proved a simple 2-approximation to the gain from trade if the buyer's median valuation is above the seller's [17]. This was significantly improved by Colini-Baldeschi et al. in [5] to $1/r$ and $O(\log(1/r))$, where r is the probability that the buyer's valuation for the item is higher than the seller's. Recently, Giannakopoulos et al. [10] studied an online variant of this setting where buyers and sellers are revealed sequentially by an adversary and have known prior distributions on the value of the items.

The random order model we are using has its origins in the well-known secretary problem, where n items arrive in online fashion and our goal is to maximize the probability of selecting the most valuable, without knowing their values in advance. The matroid secretary problem was introduced by Babaioff et al. [1]. In this setting, we are allowed to select more than item, provided our final selection satisfies matroid constraints. A variety of different matroids have been studied, with many recent results presented by Dinitz in [7]. Of particular interest to our problem are secretary problems on bipartite graphs. Here, the left hand side vertices of the graph are fixed and the right hand side vertices (along with their incident) edges appear online. The selected edges must form a (incomplete) matching and the goal is to maximize the sum of their weights. Babaioff et al. in [1] provided a $4d$ -competitive algorithm for the transversal matroid with bounded left degree d , which is a special case of the online bipartite matching where all edges connected to the same left hand side vertex have equal value. This was later improved to 16 by Dimitrov and Plaxton [6]. The case where all edges have unrelated weights was first considered by Korula and Pal in [15] who designed a 8-competitive algorithm, which was later improved to the optimal $1/e$ by Kesselheim et al. [13]. Another secretary variant which is close to our work is when the online selects k items instead of one, where Kleinberg [14] showed an asymptotically tight algorithm with competitive ratio $1 - O(\sqrt{1/k})$.

The wide range of applications of secretary models (and the related prophet inequalities) have led to the design of posted price mechanisms, that are simple to describe, robust, truthful and achieve surprisingly good approximation ratios. Hajiaghayi et al. introduced prophet inequality techniques in online auction in [12]. The k -choice secretary described above was then studied in [11] which combined with [14] yielded an asymptotically optimal, truthful mechanism. For more general auction settings,

posted-price mechanisms have been used by Chawla et al. in [4] for unit demand agents and expanded by Feldman et al. in [9] for combinatorial auctions and [8] for online budgeted settings.

3 Model and Notation

The setting of the *random intermediation* problem consists of sets $B = \{b_1, \dots, b_n\}$ and $S = \{s_1, \dots, s_n\}$ containing the valuations of the buyers and sellers. For convenience, we assume that they are all distinct. The intermediary interacts with a uniformly random permutation σ of $B \cup S$ which is presented to him one agent at a time, over $2n$ steps. The intermediary has no knowledge of $\sigma(t)$ before step t . We use b^i and s^j to denote the i -th *highest* valued seller and j -th *lowest* valued seller respectively.

We study *posted price* mechanisms that upon seeing the identity of agent t offer price p_t . This price can not depend on the entire valuation function; only the values within $\sigma(1) \dots \sigma(t-1)$ which are revealed at this point. We buy or sell one item from sellers or buyers who accept our price, respectively. Of course, we can only sell items if we have stock available. Formally, let κ_t be the number of items at time t . Starting with κ_0 items (with $\kappa_0 = 0$ for welfare and 1 for the gain-from-trade):

$$\kappa_{t+1} = \begin{cases} \kappa_t + 1, & \text{if } \sigma(t) \in S \wedge \sigma(t) \leq p_t \\ \kappa_t - 1, & \text{if } \sigma(t) \in B \wedge \kappa_t \geq 1 \wedge \sigma(t) \geq p_t \\ \kappa_t & \text{otherwise.} \end{cases}$$

The set of sellers from whom we bought items during the algorithm's execution is

$$T_S = \{s \in S \mid \exists t \sigma(t) = s \leq p_t\}$$

and similarly the set of buyers we sold to is $T_B = \{b \in B \mid \exists t \sigma(t) = b \geq p_t \wedge \kappa_t > 0\}$. Notice that these are random variables, depending on σ .

The social *welfare* of online algorithm A is the sum of the valuations of all agents *with* items. In particular, after executing A it is: $\mathcal{W}_A(S, B) = \mathbb{E} \left[\sum_{s \in S \setminus T_S} s + \sum_{b \in T_B} b \right]$. The gain from trade (or GFT) produced by algorithm A throughout the run is the difference between the final and starting welfare: $GFT_A(S, B) = \mathbb{E} \left[\sum_{b \in T_B} b - \sum_{s \in T_S} s \right]$.

We are interested in the *competitive ratio* of our online algorithm A compared to the offline algorithm OPT . In this setting there are two different offline algorithms to compare against: optimal offline and sequential offline. They both know S, B , but the first can always achieve the maximum welfare, whereas the second operates under the same constraints as we, namely he can only perform trades permitted by σ , which is unknown. We say that algorithm A is ρ -competitive for welfare (or gain from trade) if for any B, S we have:

$$\mathcal{W}_A(B, S) \geq \rho \cdot \mathcal{W}_{OPT}(S, B) - \alpha, \tag{1}$$

for some fixed $\alpha \geq 0$.

Often we will refer to the *matching* between a set of buyers and a set of sellers. Let $M(S, B) = \{\{S_1\} \cup \{B_1\}\}$, where $S_1 \subseteq S, B_1 \subseteq B$ is the set of sellers and buyers with whom we trade (or are matched, in the sense that the items move from sellers to buyers) in a welfare maximizing allocation and $m(S, B)$ the optimal gain from trade. Note that this does *not* contain pairs: only the set of each side of the matching. Similarly, let $M(S, B, q, p)$ be the matching generated by only trading with sellers valued below q and buyers above p . In a slight abuse of notation, we will use $|M(S, B)| = |S_1|$ for the size of the matching and $M(S, B) \cap M(S', B') = \{\{S_1 \cap S'_1\} \cup \{B_1 \cap B'_1\}\}$.

4 Welfare

In order to approximate the welfare, the online algorithm uses a sampling phase to find the median price, in an attempt to move items from agents below the median to more valuable ones above it. The

two main challenges, in terms of its performance, are estimating the median with a small sample and missing trades due to the online nature of the input. Before we delve into the actual algorithm, it is useful to state two probability concentration results, similar to the familiar Azuma-Hoeffding inequality, but for the setting where sampling happens *without* replacement as is our case. Their proof is in the Appendix.

Lemma 1. *Let $\mathcal{X} = \{x_1, \dots, x_N\}$ where $x_i \in \{0, 1\}$ and $x_1 = x_2 = \dots = x_m = 1$. Consider sampling n values of \mathcal{X} uniformly at random **without** replacement and let X_i be the value of the i -th draw. For $Y = \sum_{i=1}^n X_i$, we have that for any $\epsilon > 0$:*

$$\Pr[Y \geq (1 + \epsilon) \mathbb{E}[Y]] \leq e^{-2\epsilon^2 \max\{m, n\} \frac{mn}{N^2}} \quad (2)$$

and

$$\Pr[Y \leq (1 - \epsilon) \mathbb{E}[Y]] \leq e^{-2\epsilon^2 \max\{m, n\} \frac{mn}{N^2}}. \quad (3)$$

Similarly, we often encounter a situation where we are interested in the number of trades between n sellers and n buyers, arriving in a uniformly random permutation. Assuming we buy from all sellers, occasionally we would encounter a buyer without having any items at hand. This results shows that even though this is the case, few trades are lost.

Lemma 2. *The number of trades $M(\sigma)$ in a uniformly random sequence containing n buyers and n sellers satisfies:*

$$\mathbb{E}[M] \geq \frac{n-1}{n} \left(n - \sqrt{2n \log n} \right), \quad (4)$$

assuming all sellers are valued below all buyers.

All the machinery is now in place analyze sequential algorithms in this setting. We first show a key property of the offline algorithm.

Proposition 1. *The optimal offline algorithm sets a price p , equal to the median of all the agents' valuations and trades items from sellers valued below p to buyers valued above p .*

Proof. Since there are only n items available, if we could freely redistribute the items we would choose the top n agents with highest valuations. Let p be the value of the n -th most valuable agent. If there are k buyers valued above p we have $n - k$ buyers and k sellers valued below it. Thus, buying from all sellers below p and selling to all buyers above it is an optimal algorithm.

However, the optimal sequential offline algorithm would not just trade at this price. For instance, if there is 1 buyers and $n - 1$ sellers above p and 1 seller and $n - 1$ buyers below, trading at this price would give a $1/2$ probability of transferring the item, since only one transfer increases the welfare and the agents have to appear in the right order. Therefore, if that buyer has a much larger valuation than anyone else, this algorithm would only be $1/2$ -competitive. However, we can modify this approach with a bias towards buying more items than needed, in order to maximize the probability of finding high valued buyers. See Lemma 9 in the Appendix that quantifies the relation between the optimal and the sequentially optimal welfare.

The next step is to design an online algorithm without knowing p or $|M(S, B)|$ beforehand. The algorithm is as follows:

1. Record the first $8n^{2/3} \log n$ agents and calculate their median p' . Buy from all sellers during this sampling phase.
2. After the sampling phase:
 - (a) Buy from seller s if $s \leq p'$.
 - (b) Sell to buyer b if an item is available and $b \geq p'$.

For the analysis of this algorithm, we first need a concentration result on the sample median p' .

Lemma 3. *Let $X = \{1, \dots, 2n\}$ and select $8n^{2/3} \log n$ elements from X without replacement. Then, their sample median M satisfies:*

$$\Pr[|M - n| \geq n^{2/3}] \leq O\left(\frac{1}{n}\right). \quad (5)$$

Proof. We have that:

$$\Pr[M \geq n + n^{2/3}] = \Pr[\text{more than } 2n^{2/3} \log n \text{ elements picked } \geq n + n^{2/3}]$$

Since we are sampling without replacement, this is equivalent to selecting $8n^{2/3} \log n$ elements uniformly at random from X' containing $n + n^{2/3}$ 0's and $n - n^{2/3}$ 1's and having their sum be greater than $4n^{2/3} \log n$. Using X' , $\epsilon = n^{2/3}/(n - n^{2/3})$ and taking $8n^{2/3} \log n$ samples in Lemma 1, we have:

$$\begin{aligned} \Pr[M \geq n + n^{2/3}] &= \Pr[Y \geq (1 + \epsilon) \mathbb{E}[Y]] = \Pr[Y \geq n^{2/3}] \\ &\leq \exp\left(-2 \left(\frac{n^{2/3}}{(n - n^{2/3})}\right)^2 (n - n^{2/3}) \frac{8n^{2/3} \log n (n - n^{2/3})}{4n^2}\right) \\ &\leq O\left(\frac{1}{n}\right). \end{aligned}$$

By symmetry, the same holds for $\Pr[M \leq n - n^{2/3}]$: just reverse the ordering of the agents.

This shows that our sample median p' might have at most $n^{2/3}$ agents more on one side compared to the true median p . However, this loss is negligible asymptotically. We now show that buying from sellers during the sampling phase, before considering any buyers, can only increase the number of trades. The proof is in the Appendix.

Lemma 4. *Let σ be a sequence containing n buyers and n sellers and $s\sigma'$ the exact same sequence where a seller has been moved to the front. Then we have:*

$$M(s\sigma') \geq M(\sigma)$$

where $M(\sigma)$ is the number of items sold.

Theorem 1. *This algorithm is $\left(1 - \tilde{O}\left(\frac{1}{n^{1/3}}\right)\right)$ -competitive for welfare.*

Proof. As before, let $M = |M(S, B)|$ be the size of the optimal offline matching. The following analysis assumes that the event of Lemma 3 did not occur and p and p' split the agents in two sets, differing by at most $n^{2/3}$. Given this, we analyze the algorithm in three steps. First show that we never buy too many items from highly valued sellers, therefore we keep most of the sellers' contribution to the final welfare. Then we show that we always match a high proportion of the valuable buyers by considering two cases: if there are few such buyers then they are matched to the sellers we obtained during the sampling phase, otherwise we have enough sellers below p' to match them to.

We introduce some notation useful to the analysis: let W be the set containing the top $n - n^{2/3}$ highest valued agents. Then let S_W, B_W be the number of sellers and buyers respectively in W and S'_W, B'_W be how many of them appeared after the sampling phase. valued agents. To show the competitiveness of our algorithm, it suffices find the fraction of W that is achieved at the end of the sequence: being $(1 - \tilde{O}(1/n^{1/3}))$ -competitive against W , the top $n - n^{2/3}$ agents, implies a ratio of

$$\left(1 - \tilde{O}(1/n^{1/3})\right) \cdot \frac{n - n^{2/3}}{n} = 1 - \tilde{O}(1/n^{2/3})$$

against all n agents above the median and therefore the optimal offline.

We first show that we never lose too much welfare by buying from sellers. Given p' , the only occasion on which a seller in W is bought is if he is amongst the first $8n^{2/3} \log n$ sellers. This event is clearly independent from the condition on p' , meaning in expectation we keep

$$\mathbb{E}[S'_W] = S_W \left(1 - \frac{8n^{2/3} \log n}{n}\right) = S_W \left(1 - \frac{8 \log n}{n^{1/3}}\right)$$

highly valued sellers. Therefore, enough of the sellers' original value is kept. For the number of items I_S bought during the sampling phase, the following holds by Lemma 1:

$$\Pr \left[I_S \leq \left(1 - \frac{1}{2}\right) 4n^{2/3} \log n \right] \leq e^{-2\frac{1}{4} 8n^{2/3} \log n \frac{n^2}{4n^2}} \leq e^{-n^{2/3}}. \quad (6)$$

To analyze the number of buyers in W matched, we consider two cases.

$B_W \leq n^{2/3} \log n$: We first need to find $\mathbb{E}[B'_W]$, which is slightly more complicated, since we have conditioned on p' approximating the median. Given p' , at least $4n^{2/3} \log n$ agents were above the median value during the sampling phase. Note that any of the buyers in W are above the median. Therefore, any of the agents in the upper $4n^{2/3} \log n$ half of the sampling phase could be replaced by a buyer in W . Thus, in a random permutation, we have:

$$\mathbb{E}[B'_W] \geq B_W \left(1 - \frac{4n^{2/3} \log n}{n - n^{2/3}}\right).$$

We might also consider up to $n^{2/3}$ extra buyers, if p' underestimated p . However, given that $I_S \geq 2n^{2/3} \log n$ with high probability, every buyer in B'_W will be matched with an item, giving the claimed competitive ratio for this case.

$B_W > n^{2/3} \log n$: Let $k \geq B_W$ be the number of trades the optimal offline algorithm would perform. Since the median might be underestimated, the number of sellers we consider is at least $k - n^{2/3}$ and buyers at most $k + n^{2/3}$. We show that, with the help of the extra items we bought during sampling, we have more items than buyers in total, with high probability. Let $S_{p'}, B_{p'}$ the number of sellers and buyers below and above p' after the sampling phase. By Lemma 1 we expect to find

$$\Pr[S_{p'} \leq \left(1 - \sqrt{\frac{\log k}{k}}\right)(k - n^{2/3}) \frac{2n - 8n^{2/3} \log n}{2n}] \quad (7)$$

$$\leq \exp\left(-2 \frac{\log k}{k} (k - n^{2/3}) \frac{(2n - 8n^{2/3} \log n)^2}{4n^2}\right) \leq O\left(\frac{1}{k}\right). \quad (8)$$

Similarly we have

$$\Pr[B_{p'} \geq \left(1 + \sqrt{\frac{\log k}{k}}\right)(k + n^{2/3}) \frac{2n - 8n^{2/3} \log n}{2n}] \leq O\left(\frac{1}{k}\right). \quad (9)$$

It is important to note that these quantities are almost equal, other than a $n^{2/3}$ factor which is insignificant compared to k . Then, with high probability:

$$B_{p'} - S_{p'} \leq \left(1 + \sqrt{\frac{\log k}{k}}\right)(k + n^{2/3}) - \left(1 - \sqrt{\frac{\log k}{k}}\right)(k - n^{2/3}) \quad (10)$$

$$\leq 2(\sqrt{k \log k} + n^{2/3}) \quad (11)$$

$$\leq 3(\sqrt{n \log n}), \quad (12)$$

given that $k \leq n$. Therefore, since we bought at least $2n^{2/3} \log n$ items during the sampling phase, the total number of items bought is higher than the total number of buyers considered for n large enough. Also, by Lemma 4 having these items ready before encounter buyers is beneficial.

Therefore, we get a lower bound on the true number of buyers in $B_{p'}$ that acquire an item using Lemma 2. The number of items sold in expectation is at least:

$$\mathcal{M} \geq \frac{B_{p'} - 1}{B_{p'}} \left(B_{p'} - \sqrt{2B_{p'} \log B_{p'}} \right). \quad (13)$$

However, we are interested only in the fraction of buyers in B_W who acquired an item. The algorithm does not differentiate between any buyer above p' , the sequence is uniformly random and all buyers in B_W are contained within the top $k + n^{2/3}$ buyers. By lower bounding $B_{p'}$ with Lemma 1:

$$\Pr[B_{p'} \leq (1 - \sqrt{\frac{\log k}{k}})(k + n^{2/3}) \frac{2n - 2n^{2/3} \log n}{2n}] \leq O\left(\frac{1}{k}\right), \quad (14)$$

and using (13), the fraction of buyers in B_W matched is at least:

$$\frac{\mathcal{M}}{k + n^{2/3}} \geq 1 - \tilde{O}\left(\frac{1}{n^{2/3}}\right), \quad (15)$$

with high probability, since $k \geq n^{2/3} \log n$.

5 Gain from Trade

Compared to the welfare, the gain from trade is a more challenging objective. The main reason is that even for large n , the actual trades that maximize the GFT can be very few and quite well hidden. Moreover, buying from a single seller and being unable to sell could completely shatter the GFT, while it could have very little effect on the welfare.

First of all, the setting has to be slightly changed. We give the online algorithm one extra, free item at the beginning to ensure that at least one buyer can acquire an item, even when the initial sampling has been inconclusive. For fairness, the offline algorithm is also provided with this starting item. We show that this modification is absolutely necessary to study this setting under competitive analysis.

Theorem 2. *Starting with no items, the competitive ratio for the GFT is arbitrarily high.*

Proof. Consider two different valuations: $s_1 = c > 0, b_1 = c + \epsilon$ and $\hat{s}_1 = c, \hat{b}_1 = c - \epsilon, \hat{s}_2 < \hat{b}_1 - \epsilon'$. In both cases, remaining buyers and sellers have values of 0 and ∞ , thus cannot be part of any transaction increasing the total welfare.

Let $p = \Pr[s_1 \in T_S \mid \sigma^{-1}(b_1) > \sigma^{-1}(s_1)]$ be the probability of the online algorithm buying from s_1 , conditioned on b_1 arriving later. This must be $p > 0$, otherwise his expected gain from trade will be 0, compared to the $\epsilon/2$ generated by the offline.

However, in the second instance the algorithm should buy from \hat{s}_2 instead of \hat{s}_1 . But, if \hat{s}_1 appears first, the first algorithm should buy from him too, as the information received so far is the same:

$$\hat{p} = \Pr[\hat{s}_1 \in T_S \mid \sigma^{-1}(\hat{b}_1) > \sigma^{-1}(\hat{s}_1) \wedge \sigma^{-1}(\hat{s}_2) > \sigma^{-1}(\hat{s}_1)] \geq p \Pr[\sigma^{-1}(\hat{s}_2) > \sigma^{-1}(\hat{s}_1)] = \frac{p}{2} > 0. \quad (16)$$

So the online algorithm has a positive chance of buying the item from the wrong seller. Assuming in all other cases maximum gain from trade is extracted, we have:

$$GFT_A(S, B) \leq \hat{p}(-\epsilon) + (1 - \hat{p})(\epsilon' - \epsilon). \quad (17)$$

Since \hat{p} is independent of ϵ, ϵ' , careful selection of these parameters can force the expected gain from trade to be negative, leading to arbitrarily bad competitive ratio.

In any case, no online algorithm can perform well in both instances.

To avoid the previous pitfall, we assume the intermediary starts with one item. Roughly, the algorithm starts by estimating the total volume of trades in an optimal matching by observing the first segment of the sequence. Using this information, two prices \hat{p}, \hat{q} are computed, to be offered to agents in the second part. This being an ordinal mechanism, the goal is to maximize the number of trades *and leave no item unsold*. The online algorithm $A(c, \epsilon, N)$ contains parameters whose values will be specified later.

Input: A sequence σ of length $2n$, appearing online.

Output: A matching between buyers and sellers.

With probability $\frac{1}{2}$ ignore sellers and sell the item as in the normal secretary, otherwise continue ;

Split the sequence into two segments such that $\sigma = \sigma_1\sigma_2$, with $|\sigma_1| = c \cdot 2n$;

Let S_1, B_1 denote the sellers and buyers of σ_1 ;

Calculate the welfare maximizing matching $M(S_1, B_1)$;

if $|M(S_1, B_1)| \leq N$ **then**

 | Sell the item to the highest buyer as in the normal secretary problem and stop;

end

Set \hat{p}, \hat{q} which only keep $(1 - \epsilon) \cdot c \cdot |M(S_1, B_1)|$ many matched pairs;

$i \leftarrow c \cdot 2n$;

$k \leftarrow \emptyset$;

$M \leftarrow \emptyset$;

/ For the first half of σ_2 , buy and sell items, keeping at most one in stock */*

while $i \leq c \cdot 2n + (1 - c) \cdot 2n/2$ **do**

 | **if** $\sigma(i)$ is a seller, $k = \emptyset$ and $\sigma(i) \leq \hat{q}$ **then**

 | $k \leftarrow \sigma(i)$;

end

 | **if** $\sigma(i)$ is a buyer, $k \neq \emptyset$ and $\sigma(i) \geq \hat{p}$ **then**

 | Sell to $\sigma(i)$;

 | $k \leftarrow \emptyset$;

end

 | $i \leftarrow i + 1$;

end

For the second half of σ_2 , just try to sell the last remaining item, if any;

The idea is to use the first part of the sequence to estimate the matching $M(S, B)$. If a large (in terms of pairs) GFT maximizing matching is observed, it is likely that a proportionate fraction of it will be contained in the second half. In that case, sellers and buyers are matched in non overlapping pairs to avoid buying too many items.

Before moving on to the analysis of the algorithm, we need a small lemma on the structure of revenue maximizing matchings, to explain the prices set.

Lemma 5. *For any S, B and $S_1 \subseteq S, B_1 \subseteq B$:*

1. $m(S, B)$ can be obtained by setting two threshold prices p, q and trading with buyers above and sellers below them.
2. Choosing $\hat{p} > p$ and $\hat{q} < q$ such that $|M(S, B, \hat{q}, \hat{p})| \geq \alpha |M(S, B)|$ for $\alpha < 1$ yields $m(S, B, \hat{q}, \hat{p}) \geq \alpha m(S, B)$.
3. $|M(S, B)| \geq |M(S_1, B_1)|$ and $m(S, B) \geq m(S_1, B_1)$.

Proof. For Property 1, assume $s < b < \hat{s} < \hat{b}$, such that $s, b, \hat{s}, \hat{b} \in M(S, B)$. But, instead of two matches we can just match \hat{b} to s instead: $\hat{b} - s > b - s + \hat{b} - \hat{s}$, thus any such pair of matched agents cannot be part of $M(S, B)$. Setting $q = \max \{s \in M(S, B)\}$ and $p = \min \{b \in M(S, B)\}$ we have $q < p$ and the result follows.

Property 2 follows because $M(S, B, \hat{q}, \hat{p})$ contains the α highest value pairs for $M(S, B)$. Property 3 is straightforward.

Theorem 3. $A(c = 0.3, \epsilon = 0.2758, N = 114)$ is $O(1)$ -competitive for the gain from trade.

Proof. Let $z = |M(S, B)|$. We bound the gain from trade for the case where σ_1, σ_2 contain their analogous proportion of $M(S, B)$ and show that the losses are insignificant otherwise. In particular, let

$$\begin{aligned} f(c, \epsilon, z) &= \Pr \left[\frac{|M(S, B) \cap M(S_1, B_1)|}{|M(S, B)|} \geq c(1 - \epsilon) \wedge \frac{|M(S, B) \cap M(S_2, B_2)|}{|M(S, B)|} \geq (1 - c)(1 - \epsilon) \right] \end{aligned}$$

be the *well mixed* probability, where an ϵ -approximate chunk of the matching appears in both parts. The two events are not independent. To bound $f(c, \epsilon, z)$, it suffices to study the distribution of $S_M = \{s \in M(S, B)\}$ and $B_M = \{b \in M(S, B)\}$, the sets of agents comprising the optimal matching. By Lemma 5, we know that any seller in S_M can be matched to any buyer in B_M . Since we only care about the *size* of the matching in σ_1 and σ_2 , not its actual value, we can rewrite $f(c, \epsilon, z)$ as:

$$f(c, \epsilon, z) = \Pr \left[\frac{|S_M \cap S_1|}{|S_M|} \geq c(1 - \epsilon) \wedge \frac{|B_M \cap B_1|}{|B_M|} \geq c(1 - \epsilon) \wedge \right. \quad (18)$$

$$\left. \frac{|S_M \cap S_2|}{|S_M|} \geq (1 - c)(1 - \epsilon) \wedge \frac{|B_M \cap B_2|}{|B_M|} \geq (1 - c)(1 - \epsilon) \right], \quad (19)$$

which is easier to handle.

It is useful to think the input as being created in two steps: first the *volume* of agents in S_1, B_1, S_2, B_2 is chosen and *afterwards* their exact values are randomly assigned. As such, a lower bound on the size of the offline matching provides the same bound on its gain from trade. We begin by bounding $f(c, \epsilon, z)$.

Lemma 6. *The probability the matching is well-mixed is*

$$f(c, \epsilon, z) \geq 1 - 2(e^{-2\epsilon^2 z c^2} + e^{-2\epsilon^2 z (1-c)^2})$$

The proof is straightforward, see the Appendix.

Let p and q be the prices achieving the matching $M(S, B)$, by Lemma 5. We need to show that the prices \hat{p}, \hat{q} computed achieve a constant approximation of $m(S_2, B_2)$. Since $M(S, B)$ is well mixed and by using Lemma 5 we have that:

$$|M(S, B)| \geq |M(S_1, B_1)| \geq |M(S_1, B_1, q, p)| \geq (1 - \epsilon) \cdot c \cdot |M(S, B)|, \quad (20)$$

where the second inequality holds since $M(S_1, B_1)$ is a gain from trade maximizing matching and the third because at least a $(1 - \epsilon) \cdot c$ fraction of $M(S, B)$ appeared in σ_1 . In particular, we have that $M(S_1, B_1, q, p) \subseteq M(S_1, B_1)$ is the highest value part of $M(S_1, B_1)$ and $M(S_1, B_1, \hat{q}, \hat{p}) \subseteq M(S_1, B_1, q, p)$, thus $\hat{q} \leq q$ and $\hat{p} \geq p$ leading to:

$$|M(S_1, B_1, \hat{q}, \hat{p})| \geq (1 - \epsilon)^2 c^2 |M(S, B)| \quad (21)$$

by Eq. (20). We now need to find just how many of the trades in $M(S_2, B_2, \hat{p}, \hat{q})$ are achieved by our algorithm. Let $\hat{S}_2 = \{s \mid s \in S_2 \wedge s < \hat{q}\}$ and $\hat{B}_2 = \{b \mid b \in B_2 \wedge b > \hat{p}\}$. We need a high probability guarantee on the size of \hat{S}_2 and \hat{B}_2 .

Lemma 7. *Assuming the matching is well mixed:*

$$\Pr \left[|\hat{S}_2| \geq ((1 - c)(1 - \epsilon) - \frac{1}{2}) |S_M| \right] \geq 1 - 2^{-c^2 (1 - \epsilon)^2 |S_M|}.$$

The proof is in the Appendix.

Clearly, Lemma 7 holds for buyers as well. The proof is almost identical, keeping in mind that buyers are ordered the opposite way.

At this point we have a clear indication of how many sellers and buyers the prices \hat{p}, \hat{q} cover in the second part of the sequence. Since this is an ordinal mechanism, we want to maximize the number of trades *provided no item is left unsold*. There are no a priori guarantees on the welfare increase of each trade, even a single unsold item ruins our gain from trade guarantees, in the worst case.

Lemma 8. *Let $N = |M(S_2, B_2, \hat{q}, \hat{p})|$ and $M = |S_2| + |B_2| - N$. Then, the probability that no item is left unsold is at least $1 - 2^{-N}$. Moreover, the expected number of trades in this case is:*

$$\frac{\frac{N+M/2-1}{2N+M-1} \cdot \frac{N}{2} - \frac{N}{2^N}}{1 - 2^{-N}} \approx N/4 \quad (22)$$

The proof is in the Appendix.

Everything is now in place to provide a lower bound on the gain from trade of the matching calculated by the algorithm. Assuming $z = |M(S, B)|$, we can compose Lemma 6, Lemma 7 and Lemma 8 to show that with probability at least

$$J(c, \epsilon, z) = f(c, \epsilon, z) \cdot (1 - 2^{-c^2(1-\epsilon)^2z}) \cdot (1 - 2^{-((1-c)(1-\epsilon) - \frac{1}{2})z}) \quad (23)$$

the matching has size at least

$$\frac{((1-c)(1-\epsilon) - \frac{1}{2})z}{4}. \quad (24)$$

The matching is not a uniformly random subset of $M(S, B)$, but it is skewed to contain higher value trades since $\hat{p} > p$ and $\hat{q} < q$. Taking into account that we run a simple secretary algorithm with probability 1/2 and assuming we lose the highest valued seller s^* in our matching when the agents are not well mixed the GFT is:

$$\frac{1}{2e}b^1 + \frac{J(c, \epsilon, z)}{2} \cdot \frac{((1-c)(1-\epsilon) - \frac{1}{2})m(S, B)}{4} - \frac{1 - J(c, \epsilon, z)}{2}s^* \quad (25)$$

whereas the offline GFT is at most

$$m(S, B) + b^1. \quad (26)$$

To upper bound the competitive ratio we analyze two different cases:

1. If $z < N$:

In this case the algorithm would never detect a sufficiently sized matching.

$$\rho \geq \frac{\frac{1}{e}b^1}{m(S, B) + b^1} \geq \frac{1}{e} \cdot \frac{b^1}{(z+1)b^1} = \frac{1}{e(z+1)} \quad (27)$$

2. If $N \leq z < N \frac{1}{c(1-\epsilon)}$. In the well-mixed case, the online algorithm will not detect a matching and fall back to secretary. Therefore, the competitive ratio is:

$$\rho \geq \frac{\frac{1}{2e}b^1 + \frac{1}{2}(f(c, \epsilon, z)\frac{1}{e}b^1 - (1 - f(c, \epsilon, z))s^*)}{m(S, B) + b^1} \geq \frac{1 - e + (1 + e)f(c, \epsilon, z)}{2e(z+1)}. \quad (28)$$

given that $c < 1/e$ and the sampling phase for the secretary continues.

3. $z \geq N \frac{1}{c(1-\epsilon)}$. Now in the well mixed case a large enough matching is found. We compare We have:

$$\frac{\frac{1}{2e}b^1 - \frac{1 - J(c, \epsilon, z)}{2}s^*}{b^1} \geq \frac{1}{2e} - (1 - J(c, \epsilon, z)), \quad (29)$$

and

$$\frac{\frac{J(c, \epsilon, z)}{2} \cdot \frac{((1-c)(1-\epsilon) - \frac{1}{2})m(S, B)}{4}}{m(S, B)} \geq \frac{J(c, \epsilon, z) \cdot ((1-c)(1-\epsilon) - \frac{1}{2})m(S, B)}{8}. \quad (30)$$

Therefore, the competitive ratio is:

$$\begin{aligned} \rho &\geq \frac{\frac{1}{2e}b^1 + \frac{J(c, \epsilon, z)}{2} \cdot \frac{((1-c)(1-\epsilon) - \frac{1}{2})m(S, B)}{4} - \frac{1 - J(c, \epsilon, z)}{2}s^*}{m(S, B) + b^1} \\ &\geq \min \left\{ \frac{1}{2e} - (1 - J(c, \epsilon, z)), \frac{J(c, \epsilon, z) \cdot ((1-c)(1-\epsilon) - \frac{1}{2})m(S, B)}{8} \right\}. \end{aligned}$$

Therefore, c , ϵ and N are selected to maximize the minimum amongst all cases of z , which is picked by the adversary. Computationally, we find that setting $c = 0.3$, $\epsilon = 0.2758$ and $N = 114$ yields $\rho \geq 1/1434$.

If we are given that $|M(S, B)|$ will be large, then this algorithm can be adapted to have greatly improved competitive ratio. In particular, setting $c = \epsilon = 0.01$ achieves $\rho \geq 1/17$ as $|M(S, B)| \rightarrow \infty$.

References

1. Moshe Babaioff, Nicole Immorlica, and Robert Kleinberg. Matroids, secretary problems, and online mechanisms. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, pages 434–443, Philadelphia, PA, USA, 2007. Society for Industrial and Applied Mathematics.
2. Liad Blumrosen and Yehonatan Mizrahi. Approximating gains-from-trade in bilateral trading. In Yang Cai and Adrian Vetta, editors, *Web and Internet Economics*, pages 400–413, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg.
3. Johannes Brustle, Yang Cai, Fa Wu, and Mingfei Zhao. Approximating gains from trade in two-sided markets via simple mechanisms. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 589–590. ACM, 2017.
4. Shuchi Chawla, Jason D. Hartline, David L. Malec, and Balasubramanian Sivan. Multi-parameter mechanism design and sequential posted pricing. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 311–320. ACM, 2010.
5. Riccardo Colini-Baldeschi, Paul Goldberg, Bart de Keijzer, Stefano Leonardi, and Stefano Turchetta. Fixed price approximability of the optimal gain from trade. In Nikhil R. Devanur and Pinyan Lu, editors, *Web and Internet Economics*, pages 146–160, Cham, 2017. Springer International Publishing.
6. Nediako B. Dimitrov and C. Greg Plaxton. Competitive weighted matching in transversal matroids. *Algorithmica*, 62(1):333–348, Feb 2012.
7. Michael Dinitz. Recent advances on the matroid secretary problem. *SIGACT News*, 44(2):126–142, June 2013.
8. Alon Eden, Michal Feldman, and Adi Vardi. Online random sampling for budgeted settings. In Vittorio Bilò and Michele Flammini, editors, *Algorithmic Game Theory*, pages 29–40, Cham, 2017. Springer International Publishing.
9. Michal Feldman, Nick Gravin, and Brendan Lucier. Combinatorial auctions via posted prices. In *Proceedings of the Twenty-sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pages 123–135, Philadelphia, PA, USA, 2015. Society for Industrial and Applied Mathematics.
10. Yiannis Giannakopoulos, Elias Koutsoupias, and Philip Lazos. Online market intermediation. In *Proceedings of the 44th International Colloquium on Automata, Languages, and Programming*, ICALP '17, 2017. Full version in CoRR: abs/1703.09279.
11. Mohammad Taghi Hajiaghayi, Robert Kleinberg, and David C. Parkes. Adaptive limited-supply online auctions. In *Proceedings of the 5th ACM conference on Electronic commerce*, pages 71–80. ACM, 2004.
12. Mohammad Taghi Hajiaghayi, Robert Kleinberg, and Tuomas Sandholm. Automated online mechanism design and prophet inequalities. In *AAAI*, volume 7, pages 58–65, 2007.
13. Thomas Kesselheim, Klaus Radke, Andreas Tönnis, and Berthold Vöcking. An optimal online algorithm for weighted bipartite matching and extensions to combinatorial auctions. In Hans L. Bodlaender and Giuseppe F. Italiano, editors, *Algorithms – ESA 2013*, pages 589–600, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
14. Robert Kleinberg. A multiple-choice secretary algorithm with applications to online auctions. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 630–631, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics.
15. Nitish Korula and Martin Pál. Algorithms for secretary problems on graphs and hypergraphs. In *Proceedings of the 36th International Colloquium on Automata, Languages and Programming: Part II*, ICALP '09, pages 508–520, Berlin, Heidelberg, 2009. Springer-Verlag.
16. R. Preston McAfee. A dominant strategy double auction. *Journal of economic Theory*, 56(2):434–450, 1992.
17. R. Preston McAfee. The gains from trade under fixed price mechanisms. *Applied Economics Research Bulletin*, 1(1):1–10, 2008.
18. Roger B Myerson and Mark A Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265 – 281, 1983.
19. Erel Segal-Halevi, Avinatan Hassidim, and Yonatan Aumann. Sbba: a strongly-budget-balanced double-auction mechanism. In *International Symposium on Algorithmic Game Theory*, pages 260–272. Springer, 2016.

6 Appendix

6.1 Missing proofs

Proof (Proof of Lemma 1). Let $Y_i = \mathbb{E}[Y|X_1, \dots, X_i]$ be the Doob martingale of Y , exposing the choices of the first i draws. Clearly we have that $|Y_{i+1} - Y_i| \leq 1$, since the knowledge of one draw cannot change the expectation by more than 1. Applying Azuma's inequality, we obtain:

$$\Pr[Y_n - Y_0 \geq t] \leq e^{-\frac{t^2}{n}}. \quad (31)$$

Let Z_j for $1 \leq j \leq m$ indicate if x_j was chosen. Since only these x_j contribute to Y , we have that $Y = \sum_{i=1}^m Z_i$. Repeating the previous martingale construction, we get:

$$\Pr[Y_m - Y_0 \geq t] \leq e^{-\frac{t^2}{m}}. \quad (32)$$

But, we know that $Y_0 = \mathbb{E}[Y] = \mathbb{E}[\sum_{i=1}^m Z_i] = m \frac{n}{N}$. Setting $t = \epsilon m \frac{n}{N}$ in both (31) and (32) and using $Y_n = Y_m = Y$ we obtain:

$$\Pr[Y \geq (1 + \epsilon) \mathbb{E}[Y]] \leq e^{-2\epsilon^2 \max\{m, n\} \frac{mn}{N^2}}. \quad (33)$$

Concentration in the opposite direction is found through exactly the same analysis.

Note that this result is not superfluous: by immediately applying Hoeffding's inequality for sampling with replacement, we would obtain:

$$\Pr[Y \geq (1 + \epsilon) \mathbb{E}[Y]] \leq e^{-2\epsilon^2 \frac{m^2 n}{N^2}},$$

which is only tight if m is large compared to N . The concentration should intuitively work if n is large as well: imagine $n = N$.

Proof (Proof of Lemma 2). Since we buy from all sellers and attempt to sell to all buyers, let X_t be 1 if at step t a seller is encountered and -1 if it is a buyer. We define the following martingale, with $Y_0 = 0$:

$$Y_{t+1} = \begin{cases} Y_t + X_t & \text{if } Y_t > 0 \\ Y_t - X_t & \text{if } Y_t < 0 \\ \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases} & \text{if } X_t = 1 \text{ and } Y_t = 0 \\ Y_t & \text{otherwise} \end{cases}$$

Basically, Y_t keeps track of the unsold items: sellers pull away from 0 and buyers towards 0. Clearly, the number of unsold items at time t is at most $|Y_t|$ and $|Y_{t+1} - Y_t| < 1$. Thus, by Azuma's inequality we have:

$$\Pr[|Y_{2n} - Y_0| \geq \sqrt{2n \log n}] \leq e^{-2 \frac{4n \log n}{2 \cdot 2^n}} = \frac{1}{n}. \quad (34)$$

Since n items are bought, we have:

$$\mathbb{E}[M] \geq \frac{n-1}{n} \left(n - \sqrt{2n \log n} \right) \quad (35)$$

Proof (Proof of Lemma 4). Let b' be the first buyer not to receive an item in σ . Clearly, if b' doesn't exist then the number of items sold in both cases is n . Assume we sell the item bought from s only if it is the last item left. Then, it is sold to b' : otherwise b' would not be the first buyer not to be sold an item is σ . There are two cases:

1. If s appears in σ before b' : both sequences continue identically as we have no items in stock after b' .
2. If s appears after b' in σ : there is one fewer seller in $s\sigma'$ after b' , since s was moved to the front. However, this can result in at one lost sale.

Proof (Proof of Lemma 6). Continuing from (18) we have:

$$\begin{aligned}
f(c, \epsilon, z) &= \Pr \left[\frac{|S_M \cap S_1|}{|S_M|} \geq c(1 - \epsilon) \wedge \frac{|B_M \cap B_1|}{|B_M|} \geq c(1 - \epsilon) \wedge \right. \\
&\quad \left. \frac{|S_M \cap S_2|}{|S_M|} \geq (1 - c)(1 - \epsilon) \wedge \frac{|B_M \cap B_2|}{|B_M|} \geq (1 - c)(1 - \epsilon) \right] \\
&\geq 1 - \Pr \left[\frac{|S_M \cap S_1|}{|S_M|} \leq c(1 - \epsilon) \right] - \Pr \left[\frac{|B_M \cap B_1|}{|B_M|} \leq c(1 - \epsilon) \right] - \\
&\Pr \left[\frac{|S_M \cap S_2|}{|S_M|} \leq (1 - c)(1 - \epsilon) \right] - \Pr \left[\frac{|B_M \cap B_2|}{|B_M|} \leq (1 - c)(1 - \epsilon) \right] \tag{36} \\
&\geq 1 - 2(e^{-2\epsilon^2 z c^2} + e^{-2\epsilon^2 z (1-c)^2}), \tag{37}
\end{aligned}$$

where (36) follows by taking the complement and a union bound and (37) by applying Lemma 1 individually for each event.

Proof (Proof of Lemma 7). In the well-mixed case, we have that

$$|S_2 \cap S_M| \geq (1 - c)(1 - \epsilon)|S_M| \text{ and } |S_1 \cap S_M| \geq c(1 - \epsilon)|S_M| \tag{38}$$

which leads to

$$|S_1 \cap S_M| \leq (1 - (1 - c)(1 - \epsilon))|S_M|. \tag{39}$$

To get a lower bound on the size, we have:

$$\Pr \left[|\hat{S}_2| \geq \frac{|S_M|}{2} - |S_1 \cap S_M| \right] \geq \Pr \left[|\hat{S}_2| \geq ((1 - c)(1 - \epsilon) - \frac{1}{2})|S_M| \right] \tag{40}$$

$$\geq \Pr [\hat{q} \geq \text{median}(S_M)] \tag{41}$$

$$\geq 1 - 2^{-c^2(1-\epsilon)^2|S_M|} \tag{42}$$

where Eq. (40) follows from Eq. (38). Eq. (41) follows since if \hat{q} is greater than the median, then at worst case all elements from $S_1 \cap S_M$ are less than \hat{q} , which still leaves plenty of sellers in S_2 . Eq. (42) follows since draws are not actually independent, but this works in the inequality's favor. From Eq. (21) we know \hat{q} is greater than at least a $c^2(1 - \epsilon)^2$ fraction of sellers. Since the 'bad case' is choosing all sellers below the median, this happens with higher probability if each draw is *with* rather than *without* replacement, leading to the result.

Proof (Proof of Lemma 8). We begin by calculating the probability of having an unsold item, which is easy: it is at most as much as the probability of not encountering a buyer within the last $(1 - c) \cdot 2n/2$ agents. Using a similar argument as Lemma 7, this probability is at most 2^{-N} .

We now need to calculate the expected number of trades. Let X_i be a random variable indicating that an item was *sold* to the i -th agent. We have:

$$\begin{aligned}
\Pr[X_i = 1] &= \Pr[\text{previous transaction was buying} \wedge X_i \text{ is a buyer}] \\
&= \frac{N}{2N} \cdot \frac{N}{2N + M - 1} = \frac{1}{2} \cdot \frac{N}{2N + M - 1},
\end{aligned}$$

since the previous transaction being buying from a seller occurs with probability $\frac{N}{2N}$ as there are N sellers in $M(S_2, B_2, \hat{p}, \hat{q})$ for $2N$ total agents and the sequence is shuffled. The second fraction has $2N + M - 1$ for the denominator, taking into account that one seller has already been used.

By linearity of expectation, the total number of trades X is:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E} \left[\sum_{i=2}^{(2N+M)/2} X_i \right] = (N + M/2 - 1) \mathbb{E}[X_i] \\ &= (N + M/2 - 1) \frac{1}{2} \cdot \frac{N}{2N + M - 1} \\ &= \frac{N + M/2 - 1}{2N + M - 1} \cdot \frac{N}{2} \end{aligned}$$

We can use this to calculate the expected number of trades in the case where nothing is left unsold:

$$\begin{aligned} \mathbb{E}[X | \text{No items unsold}] &= \frac{\mathbb{E}[X] - \mathbb{E}[X | \text{Unsold items}] \Pr[\text{Unsold Items}]}{\Pr[\text{No unsold items}]} \\ &\geq \frac{\mathbb{E}[X] - \frac{N}{2^{2N}}}{1 - 2^{-N}} \end{aligned}$$

6.2 Optimal vs sequential optimal

The following lemma quantifies the relation between the optimal and the sequentially optimal welfare.

Lemma 9. *The optimal sequential online algorithm is $\left(1 - O\left(\frac{\log n}{n^{1/3}}\right)\right)$ -competitive against the optimal offline for welfare.*

Proof. The optimal online algorithm adjusts the price p according to the following two cases, where $M = |M(S, B)|$.

1. $M \geq n^{2/3}$. In this case the same price p is used. At the end, the online algorithm will still keep the highest valued $n - M$ sellers and by Lemma 2 will match

$$\frac{M-1}{M} \left(M - \sqrt{2M \log M} \right)$$

buyers in expectation. The offline optimum will of course keep the highest $n - M$ sellers and M buyers, leading to a competitive ratio of at most:

$$\frac{M-1}{M} \left(1 - \frac{\sqrt{2 \log M}}{\sqrt{M}} \right) = 1 - O\left(\frac{\log n}{n^{1/3}}\right).$$

2. $M < n^{2/3}$.

In this case, suppose two prices are used: p_S to buy from the lowest $n^{2/3}$ sellers and p_B to sell to the highest $n^{2/3}$ buyers. For the buyers, the online does at least as well as the previous case. In particular, it obtains a uniformly random sample of size at least $n^{2/3} - \sqrt{2n^{2/3} \log n^{2/3}}$ by Lemma 2, amongst the top $n^{2/3}$ buyers with probability at least $(n^{2/3} - 1)/n^{2/3}$. Since the M buyers matched by the optimal offline are contained within the highest $n^{2/3}$ buyers, the ratio just from buyers remains the same as before.

From the sellers side, the online keeps the highest $n - n^{2/3}$ sellers, while the offline keeps at most n , for a ratio at most $1 - 1/n^{1/3}$.

Combining both cases, the ratio is asymptotically at most:

$$1 - O\left(\frac{\log n}{n^{1/3}}\right). \quad (43)$$

Note that the choice of $n^{2/3}$ to separate the two cases is optimal.