The Average-Value Allocation Problem

Kshipra Bhawalkar
Google Research, Mountain View, USA

Zhe Feng
Google Research, Mountain View, USA

Anupam Gupta
NYU & Google Research, New York City & Mountain View, USA

Aranyak Mehta
Google Research, Mountain View, USA

David Wajc[1]
Technion, Haifa, Israel

Di Wang
Google Research, Mountain View, USA

Abstract

We initiate the study of centralized algorithms for welfare-maximizing allocation of goods to buyers subject to *average-value constraints*. We show that this problem is NP-hard to approximate beyond a factor of $\frac{e}{e-1}$, and provide a $\frac{4e}{e-1}$-approximate offline algorithm. For the online setting, we show that no non-trivial approximations are achievable under adversarial arrivals. Under i.i.d. arrivals, we present a polytime online algorithm that provides a constant approximation of the optimal (computationally-unbounded) online algorithm. In contrast, we show that no constant approximation of the ex-post optimum is achievable by an online algorithm.

2012 ACM Subject Classification Theory of computation → Approximation algorithms analysis; Theory of computation → Online algorithms

Keywords and phrases Resource allocation, return-on-spend constraint, approximation algorithm, online algorithm

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2024.13

Category APPROX

Related Version *Full Version*: `http://arxiv.org/abs/2407.10401` [9]

Funding *Anupam Gupta*: Supported in part by NSF awards CCF-1955785 and CCF-2006953.
David Wajc: Supported in part by a Taub Family Foundation "Leader in Science and Technology" fellowship. Work done while the author was visiting Google Research.

1 Introduction

Allocating goods to buyers so as to maximize social welfare is one of the most central problems in economics. This problem, even under linear utilities, is complicated by buyers' various constraints and the manner in which items are revealed.

In this work we introduce the *average-value allocation* problem (AVA). Here, we wish to maximize social welfare (total value of allocated items), while guaranteeing for each buyer j an *average* value of allocated items of at least ρ_j. Formally, if the value of item i for buyer j is v_{ij}, and $x_{ij} \in \{0,1\}$ indicates whether item i is allocated to buyer j, we wish to maximize the social welfare $\sum_{ij} v_{ij}\, x_{ij}$, subject to each item being allocated to at most one buyer (i.e., $\sum_j x_{ij} \leq 1$), and to the "average value" constraint:

[1] corresponding author

$$\forall j, \quad \sum_i v_{ij}\, x_{ij} \geq \rho_j \cdot \Big(\sum_i x_{ij} \Big). \tag{1.1}$$

Average-value constraints arise naturally in numerous situations. E.g., consider settings when goods are to be distributed among "buyers", and the (fixed) cost of distributing, receiving, or deploying each such good allocated is borne by the recipient. Each buyer wants their average value for their goods to be at least some parameter ρ_j. This parameter ρ_j allows to convert between units, and so this fixed cost for each buyer can be in money, time, labor, or any other unit. So, for example, for allocation and distribution of donations to a charitable organization, a certain value-per-item is required to justify the time contributed by volunteers, or the money spent by government in the form of subsidies. In other words, the amount of "benefit" per task allocated to an individual j should be above the threshold ρ_j, so that even if some of the tasks are individually less rewarding (i.e., they have benefit less than ρ_j, the total amount of happiness they get overall justifies their workload.

In addition to this average-value constraint on the allocation, we may also consider side-constraints (such as the well-studied budget constraints), but for now we defer their discussion and focus on on the novel constraint (1.1). At first glance, the AVA problem may seem similar to other packing problems in the literature, but there is a salient difference – it is not a packing problem at all! Indeed, if buyer i gets some subset $S_i = \{j \mid x_{ij} = 1\}$ of items in some feasible allocation, it is possible that a subset $S' \subseteq S_i$ of this allocation is no longer feasible, since its average value may be lower. Given that this packing (subset-closedness) property is crucial to many previous results on allocation problems, their techniques do not apply. Hence, we have to examine this problem afresh, and we ask: *how well can the average-value allocation be approximated?* We investigate this question, both in the offline and online settings.

1.1 Our Results and Techniques

Recall that the AVA problem seeks to maximize the social welfare $\sum_{ij} v_{ij}x_{ij}$ subject to each item going to at most one buyer, and also the novel average-value constraint (1.1) above. Our first result rules out polynomial-time exact algorithms for AVA in an offline setting, or even a PTAS, showing that this problem is as hard to approximate as the MAX-COVERAGE problem:

▶ **Theorem 1** (Hardness of AVA). *For any constant $\varepsilon > 0$, the AVA problem is NP-hard to $(\frac{e}{e-1} - \varepsilon)$-approximate.*

We then turn our attention to positive results, and give the following positive result for the problem:

▶ **Theorem 2** (Offline AVA). *There exists a randomized polynomial-time algorithm for the AVA problem which achieves an approximation factor of $\frac{4e}{e-1}$.*

To prove Theorem 2, we would like to draw on techniques used for traditional packing problems, but the non-traditional nature of this problem means we need to investigate its structure carefully. A key property we prove and leverage throughout is the existence of approximately-optimal solutions of a very special kind: each buyer gets a collection of "bundles", where a bundle for buyer j consists of a single item i with positive $v_{ij} - \rho_j$ (i.e., contributing positively to the average-value constraint (1.1)) and some number of items i with negative $v_{ij} - \rho_j$, such that they together satisfy the AVA constraint. Given this structure we can focus on partitioning items among bundles, and allocating bundles to buyers. Note that this partitioning and allocation have to happen simultaneously, since the values (i.e., v_{ij}) and whether it contributes positively or negatively (i.e., $v_{ij} - \rho_j$) depend on the buyer and bundle under consideration. We show how algorithms for GAP (generalized assignment problem) with matroid constraints [13] can be used.

Relax-and-Round. In order to extend our results from the offline to the online settings, and to add in side-constraints, we then consider linear programming (LP) based relax-and-round algorithms for the AVA problem. The LP relaxations take advantage of the structural properties above, as they try to capture the best bundling-based algorithms (and hence to approximate the optimal solution of any kind). Once we have fractional solutions to the LP, we can then round these in both offline and online settings to get our feasible allocations.

Our first rounding-based algorithm, given in §4, is in the offline setting, and yields another $O(1)$-approximate algorithm for AVA, qualitatively matching the result from Theorem 2. While the constants are weaker, the result illustrates our ideas, and allows us to support additional side-constraints (more on this in §1.1.1).

Online Algorithms. We then turn to online AVA, where items arrive over T timesteps, and must be allocated to buyers as soon as they arrive. We want to maintain feasible solutions to the AVA at all times. We show that under adversarial arrivals, only trivial $O(T)$ approximations are possible. This forces us to focus our attention on i.i.d. arrivals. Our first result is a time-efficient approximation of the optimum (computationally-unbounded) online algorithm:

▶ **Theorem 3** (Online AVA: Approximating the Optimal Online IID Algorithm). *There exists a randomized polynomial-time online algorithm for the AVA problem which achieves a constant factor of the value achieved by the optimal (computationally-unbounded) online algorithm.*

To approximate the optimum online algorithm, we provide an LP capturing a constraint only applicable to online algorithms, inspired by such constraints from the secretary problem and prophet inequality literatures [12, 34]. We then provide a two-phase online algorithm achieving a constant approximation of this LP, analyzed via a coupling with an imaginary algorithm that may violate AVA constraints and allocate items to several buyers.

We then turn our attention to approximating the ex-post optimum (a.k.a., getting a competitive ratio for the observed sequence). In contrast, we show that when comparing with the ex-post optimum, no such constant approximation ratio is possible, but we give matching upper and lower bounds. (Due to lack of space, this is deferred to Appendix A.)

▶ **Theorem 4** (Online AVA: Ex-post Guarantees (Informal)). *There exist families of online i.i.d. AVA instances on which any online algorithm is $\Omega\left(\frac{\log T}{\log\log T}\right)$-competitive. In contrast, there exists an online algorithm matching this bound asymptotically (on all instances).*

The lower bound is proved by giving an example using a balls-and-bins process (and its anti-concentration). Then we formulate an LP capturing this kind of anti-concentration, using which we match the lower bound, under some mild technical conditions (see Appendix A for details).

1.1.1 Generalizations

There are many interesting generalizations of the basic problem. For example, there might exist "budgets" which limit the number of items any buyer can receive; or more generally we may have costs on items which must sum to at most the buyer's budgets. These costs could be different for different buyers, and in different units than those captured by constraint (1.1). These constraints are the natural ones considered in packing problems; in general, we can consider the AVA constraint as being a non-packing constraint on the allocation that can supplemented with other conventional packing constraints. As we show in §4.3, our relax-and-round algorithm extends seamlessly to accommodate such side constraints, provided any individual item has small cost compared to the relevant budgets.

Another natural generalization is *return-on-spend* (RoS) constraints, which have been central to much recent work on advertisement allocation (see [25, 20]) and §1.2). We call the problem *generalized AVA* (GenAVA) and define it as follows: the objective is to maximize social welfare, but now the average value is measured in a more general way. Indeed, the allocation of item i to buyer j can incur a different "cost" c_{ij}, and the average-value constraint becomes the following ROS constraint:

$$\forall j, \quad \sum_i v_{ij}\, x_{ij} \geq \rho_j \cdot \left(\sum_i \mathbf{c_{ij}}\, x_{ij} \right). \tag{1.2}$$

In contrast to AVA, we show that allowing general costs c_{ij} in the generalized AVA problem in (1.2) makes it as hard as one of the hardest combinatorial problems – computing a maximum clique in a graph. In particular, we show that it is NP-hard to $n^{1-\varepsilon}$-approximate GenAVA with n buyers, for any constant $\varepsilon > 0$. In Appendix B we show that similar hardness persists even for stochastically generated inputs, and the problem remains hard even if we allow for bicriteria approximation.

1.2 Related Work

Resource allocation is one of the most widely-studied topics in theoretical computer science. Here we briefly discuss some relevant lines of work.

Packing/Covering Allocation Problems. The *budgeted allocation problem* or AdWords of [32] is NP-hard to approximate within some constant [14], and constant approximations are known even online [32, 11, 28]. The *generalized assignment problem (GAP)* [22] and its extension, the *separable assignment problem*, have constant approximations in both offline [23, 13] and (stochatic) online settings [30]. In both cases, arbitrarily-good approximations are impossible under adversarial online arrivals, even under structural assumptions allowing for an offline PTAS (e.g., "small" bids) [32]. However, assuming both small bids and *random-order* (or *i.i.d.*) arrivals allows us to achieve $(1-\varepsilon)$-competitiveness [16, 18, 30, 26, 2]. Some such allocation problems are also considered with concave or convex utilities [17, 7]. As noted above, many results and techniques for (offline and online) packing and covering constraints are not applicable to our problem, which is neither a packing nor covering problem in the conventional sense.

RoS constraints in online advertising. Return-on-spend constraints as defined in (1.2) have received much attention in recent years in the context of online advertising. Several popular autobidding products allow advertisers to provide campaign-level RoS constraints with a goal to maximize their volume or value of conversions (sales) [25, 20]). Fittingly, there has been much interest in understanding the RoS setting along various directions, including optimal bidding [1], mechanism design [8, 24], and on welfare properties at equilibrium [1, 15, 31]. In these results, distributed bidding based algorithms are shown to achieve a constant fraction of the optimal welfare. However, note that the per-item costs in the autobidding setting are *endogenous* (set via auction dynamics) whereas in our allocation problem there is no pricing mechanism and the costs are *exogenous.* Our results about the hardness of the generalized AVA show that under exogenous prices, such allocation problems do not admit constant (or even sublinear) approximation guarantees.

Approximating the optimum online algorithm. Our online i.i.d. results relate to a recent burgeoning line of work on approximation of the optimum online algorithm via restricted online algorithms. This includes restriction to polynomial-time algorithms (as in our case) [34, 33, 10, 3, 29], fair algorithms [5], order-unaware algorithms [19] and inflexible algorithms [4, 35], and more. These works drive home the message that approximating the optimum online algorithm using restricted algorithms is hard, but can often lead to better approximation than possible when comparing to the (unattainable) benchmark of the ex-post optimum. We echo this message, showing that for our problem under i.i.d. arrivals, a constant-approximation of the optimum online algorithm (using polytime algorithms) is possible, but is impossible when comparing to the optimum offline solution.

1.3 Problem Formulation

In the *average-value-constrainted allocation problem (AVA)*, allocating item i to buyer j yields a value of v_{ij}. Each buyer j requires that the average value they obtain from allocated items be at least ρ_j. We wish to (approximately) maximize the total social welfare, or sum of values obtained by the buyers, captured by the following integer LP:

$$\begin{aligned}
\max \quad & \sum_{(i,j)\in E} v_{ij}\, x_{ij} && \text{(AVA-ILP)}\\
\text{s.t.} \quad & \sum_{i} v_{ij}\, x_{ij} \geq \rho_j \cdot \sum_{i} x_{ij} && \forall \text{ buyers } j\\
& \sum_{j} x_{ij} \leq 1 && \forall \text{ items } i\\
& x_{ij} \in \{0,1\} && \forall \text{ items } i, \text{ buyers } j.
\end{aligned}$$

An instance $\mathcal{I}$ of AVA can be captured by a bipartite graph (I, J, E), with a set I of items and set J of buyers, and edges $E \subseteq I \times J$, capturing all buyer-item pairs with non-zero value. For $i \in I$ and $j \in J$, edge (i, j) has value v_{ij}. We say edge (i, j) is a P*-edge* (positive edge) if it has non-negative *excess* $v_{ij} - \rho_j \geq 0$, and an N*-edge* otherwise, in which case we refer to $v_{ij} - \rho_j < 0$ as its *deficit.* An item i is a P*-item* if *all* its edges in E are P-edges, and an N*-item* if *all* its edges in E are N-edges: naturally, some items may be neither P-items or N-items. We will call an instance *unit-*ρ if $\rho_j = 1$ for all buyers.[2]

In the online setting, the n buyers and their ρ_j values are known a priori, but items i are revealed one at a time, together with their value v_{ij} for each buyer j, and an algorithm must decide what buyer to allocate an item to (if any), immediately and irrevocably on

[2] Such instances capture the core difficulty of the AVA problem, and our examples (except those for GenAVA in Section B) are unit-ρ instances, so one can WLOG take $\rho_j = 1$ in the first read.

arrival. In the online i.i.d. setting, T items are drawn (one after another) i.i.d. from a known distribution over m known item *types*, with type i drawn with probability q_i. We say an edge type (i, j) is an *N-edge type* or a *P-edge type* if $v_{ij} - \rho_j < 0$ or $v_{ij} - \rho_j \geq 0$, respectively.

1.4 Paper Outline

We begin in §2 by proving some structural lemmas regarding AVA, including an unintuitive non-linear dependence of the welfare on the amount of supply. In §3 we present the improved algorithm for the offline setting giving Theorem 2. In §4 we present our LP-rounding algorithm for AVA in an offline setting. We also discuss the approach's extendability, allowing to incorporate additional constraints, in §4.3. Building on this offline rounding-based algorithm, in §5 we present a constant-approximation of the optimum online algorithm. In the interest of space, we defer the discussion of competitive ratio bounds to Appendix A, and our hardness results to Appendix B.

2 The Structure of Near-optimal Solutions for AVA

In this section, we show how to partition any feasible allocation of AVA instances into structured subsets (which we call *permissible bundles*). This bundling-based structure will prove useful for all of our algorithms.

▶ **Definition 5** (Bundling). *A set S of edges incident on buyer j is a* permissible bundle *if*
nolistsep *S consists of a single P-edge $(i^\star, j)$ and zero or more N-edges (i, j), and*
noliistsep *the edges in S satisfy the average-value constraint, i.e., $\sum_{(i,j)\in S} v_{ij} \geq \rho_j \cdot |S|$.*
A bundling-based *solution is one that can be partitioned into a collection of permissible bundles.*

Clearly, no bundling-based solution can be better than the best unconstrained solution, but in the following lemma we show a converse, up to constant factors. (Throughout, we use the shorthand notation $v \cdot x := \sum_{ij} v_{ij} x_{ij}$ for any vector $x \in \mathbb{R}^E$.)

▶ **Lemma 6** (Good Bundling-Based Solution). *Let x^* be a solution to an instance of AVA. Then, there exists a bundling-based solution $\widehat{x}$ of value at least $v \cdot \widehat{x} \geq \frac{1}{2}\, v \cdot x^*$.*

As a corollary, the best bundling-based solution is a 2-approximation, and so we will strive to approximate such bundling-based solutions.

We prove a strengthening of Lemma 6 which also addresses online settings.

▶ **Definition 7** (Committed Bundling). *An online algorithm is a* committed *bundling-based algorithm if its solution consists of permissible bundles, and items can only be added to bundles; in particular, it* commits *to the allocation of each item to a particular bundle, and does not move items between permissible bundles.*

▶ **Lemma 8** (Online Bundling-Based Solution). *Let x^* be a solution to an instance of AVA, with x^* revealed online and (all interim partial solutions) satisfying the average-value constraints throughout. Then there exists a solution $\widehat{x}$ that is the output of a committed online bundling-based algorithm, of value at least $v \cdot \widehat{x} \geq \frac{1}{2}\, v \cdot x^*$.*

Proof. For each buyer j, consider the edges $S := \{(i, j) \mid x^*_{ij} = 1\}$ corresponding to items assigned to buyer j in solution x^*, in order of addition to the solution x^*, namely $e_1, e_2, \ldots, e_{|S|}$, with $e_k = (i_k, j)$. We now show how a committed online algorithm can output a collection of permissible bundles of at least half the value from among the edges in S; doing this for each buyer proves the result.

Consider i_k, i.e., the k-th item allocated to j by x^*, if e_k is a P-edge (i.e. $v_{i_k,j} \geq \rho_j$), we denote $p = i_k$, open (create) a bundle $B_p = \{(j,p)\}$ and allocate appropriately in the new solution $\widehat{x}$. When $e_k = (i_k, j)$ is an N-edge, if e_k can be added to some open bundle B_p of j while keeping it permissible, we add (i_k, j) to B_p in solution $\widehat{x}$; otherwise, we pick some open bundle B_p of j and mark it as closed (and never add more edges to this bundle). Since x^* is feasible throughout the online arrival, for any $k \in [1, |S|]$ we have that $\sum_{\ell \leq k} v_{i_\ell,j} \geq k \cdot \rho_j$, and since we allocate all P-edges of x^* in $\widehat{x}$ and only allocate a subset of the N-edges, we find that there must always be some open bundle of j when considering an N-edge e_k. Therefore, the above (committed) bundling-based online algorithm is well-defined. Now, each bundle is closed by at most one N-edge (i,j), and so we can charge the N-edges (i,j) allocated in x^* but not in $\widehat{x}$ to the P-edge (p,j) in the bundle B_p that they closed. But by definition of the P-edge and N-edge, we know $v_{pj} \geq \rho_j \geq v_{ij}$. Therefore, denoting by x^*_D the part of the solution x^* that is *discarded* in $\widehat{x}$ and by x^*_p and x^*_n the value of the P-edges and N-edges allocated by both x^* and the new solution $\widehat{x}$, we have that $v \cdot x^*_D \leq v \cdot x^*_p$. Hence,

$$v \cdot x^* = v \cdot x^*_D + v \cdot (x^* - x^*_D) \leq 2\, v \cdot x^*_p + x^*_n \leq 2\, v \cdot (x^*_p + x^*_n). \tag{2.3}$$

That is, the obtained bundles of the solution $\widehat{x} = x^*_p + x^*_n$ constitute a 2-approximation. ◀

▶ Remark 9. This loss of a factor of two in the value is tight. To see this, consider a single-buyer unit-ρ AVA instance. There are $\frac{1}{\varepsilon}$ N-edges each with value $1-\varepsilon$ and $\frac{1}{\varepsilon(1-\varepsilon)}$ P-edges each with value $1+\varepsilon(1-\varepsilon)$. It is feasible to allocate all items to the buyer, and (arbitrarily close to) half the value of this solution is given by N-edges, but any permissible bundle contains no N-edges as any single P-edge doesn't have enough excess to cover the deficit of any N-edge.

For our algorithms it will be convenient if each item is incident only on P-edges, or only on N-edges, thus removing the ambiguity about whether to use these as the single P-edge in a permissible bundle. Fittingly, we call such instances *unambiguous*. For example, when all buyers have the same average-value constraint (i.e. $\forall j : \rho_j = \rho$), for any item i incident on a P-edge (i.e., $\exists j : v_{ij} \geq \rho$), we can trivially drop all N-edges of the item (i.e., drop (i,j') where $v_{ij'} < \rho$) since there is no reason to allocate any N-edge instead of a P-edge of i, and so making such instances unambiguous comes with no cost. As we now show, any instance of AVA in general can be made unambiguous while still preserving a bundling-based allocation that is constant-approximate for the original instance.

▶ **Lemma 10** (Bundling Unambiguous Sub-Instances). *Given an AVA instance $\mathcal{I} = (I, J, E)$, dropping all of the P-edges or all the N-edges of each item $i \in I$ independently with probability $1/2$ results in an unambiguous sub-instance $\mathcal{I}' = (I, J, E')$ (where $E' \subseteq E$), admitting a bundling-based solution x' which is 4-approximate for $\mathcal{I}$.*

Proof. Let x^* be an optimal solution for $\mathcal{I}$. If we denote by x^*_p and x^*_n the characteristic vector for P-edges and N-edges allocated by both x^* and $\widehat{x} = x^*_p + x^*_n$ as in the proof of Lemma 8, then, by the penultimate inequality of Equation (2.3), we have that $v \cdot x^* \leq 2\, v \cdot x^*_p + v \cdot x^*_n$. Now, consider the solution x' consisting of all P-edges allocated in $\widehat{x}$ that were not dropped and all non-dropped N-edges allocated in bundle S whose P-edge was also not dropped. We therefore have that this new solution has value precisely $\frac{1}{2}\, v \cdot x^*_p + \frac{1}{4}\, v \cdot x^*_n$, and so, by Equation (2.3), we have that x' is a 4-approximation, since

$$v \cdot x^* \leq 4 \cdot \left(\frac{1}{2}\, v \cdot x^*_p + \frac{1}{4}\, v \cdot x^*_n \right) = 4\, v \cdot x'.$$ ◀

We also provide an alternative, deterministic method to find such an unambiguous sub-instance. However, since our algorithms are randomized, we defer discussion of this method to the full version.Note in unambiguous instances, every item is either a P-item or an N-item.

2.1 Welfare is non-linear in supply

In this section we provide a bound on the multiplicative gain in welfare in terms of increased supply. This will prove useful later. For now, it illustrates non-linearity of the AVA problem in its supply. (This is in contrast to other allocation problems where the welfare is at best linear in the supply.)

To motivate this bound, consider the outcome of creating k copies of each item in an AVA instance. Clearly, the welfare increases by a factor of at least k, as we can just repeat the optimal allocation for the original instance k times. However, as the following example illustrates, welfare can be *super-linear* in the supply size increase for AVA.

▶ **Example 11.** Consider a unit-ρ instance of k-buyer AVA with a single P-item of value $1 + k\varepsilon$ for all buyers and k many N-items, with the i-th N-items having value zero for all buyers except for one distinct buyer i, to whom it has value $1 - \varepsilon$. In this instance $\mathsf{OPT} \approx 2$, since the P-item can only be allocated to a single buyer, who can then only be allocated one N-item, while in the instance obtained by creating k copies of each item we can allocate a P-item to each buyer together with k many N-items, and so for this instance $\mathsf{OPT} \approx k^2$, i.e., increasing supply k-fold increases the welfare $(k^2/2)$*-fold.*

The following lemma shows that the above example is an extreme case, and for a k-fold increase in supply, an $O(k^2)$-fold increase in welfare is best possible.

▶ **Lemma 12** (Supply Lemma). *Let $\mathcal{I} = (I, J, E)$ be an AVA instance, and let $\mathcal{I}' = (I', J, E')$ be an instance with the same buyer set and underlying costs and values obtained by copying each item in $\mathcal{I}$ some k times.*

$$\mathsf{OPT}(\mathcal{I}') \leq O(k^2) \cdot \mathsf{OPT}(\mathcal{I}).$$

Proof. Since bundling-based solutions are nearly optimal up to a constant factor of 2, we can start with an optimal bundling-based allocation $\mathcal{A}'$ for $\mathcal{I}'$ and randomly (and independently) associate the items of $\mathcal{I}$ with one of their k copies in $\mathcal{I}'$, allocating them as in $\mathcal{A}'$. Finally, we remove all non-permissible obtained bundles to obtain allocation $\mathcal{A}$ for $\mathcal{I}$. For each copy i' of an item i, if i' is allocated in a P-edge in $\mathcal{A}'$, the probability that i is associated with i' (and thus assigned to the same buyer by $\mathcal{A}$) is precisely $1/k$. In contrast, if i' is allocated in an N-edge by $\mathcal{A}'$, the probability that $\mathcal{A}$ allocates i the same way as i' is precisely $1/k^2$, as this requires both i to be assigned to the same bundle (associated with the same copy) and the P-edge of this bundle to similarly be assigned to the same bundle. The lemma then follows by linearity of expectation. ◀

3 Offline Algorithm via Reduction to Matroid-Constrained GAP

In this section we provide an improved constant-approximation for AVA in the *offline* setting; we will show in Appendix B.1 that the problem is hard to approximate to better than $\frac{e}{e-1}$.

▶ **Theorem 13.** *There exists a $(\frac{4e}{e-1} + o(1))$-approximate randomized algorithm for AVA.*

The algorithm proceeds by reducing AVA to GAP with matroid constraints. Recall that an instance of the *generalized assignment problem* (GAP) consists of n elements that can be packed into m bins. Packing an element e into a bin b gives a value v_{eb} and uses up s_{eb} space in that bin. If we let $y_{eb} \in \{0, 1\}$ denote the indicator for whether element e is assigned to bin b, then naturally $\sum_b y_{eb} \leq 1$. Each bin has unit size, and so the size of elements assigned to bin b is at most 1: in other words, $\sum_e s_{eb}\, y_{eb} \leq 1$. The goal is to maximize the total value of the assignment $\sum_{eb} v_{eb}\, y_{eb}$. [23] gave a $(1-1/e)$-approximation for this problem. [13] gave the same approximation for an extension of the problem, where the opened subset of bins must be an independent set in some given matroid $\mathcal{M}$.

▶ **Theorem 14.** *There exists a randomized polynomial-time algorithm that, for any unambiguous AVA instance, outputs a solution with expected value at least $\big(1 - 1/e - o(1)\big)$ times the optimal bundling-based solution.*

Proof. Given an unambiguous AVA instance (i.e., one where each item is incident on only P-edges or only N-edges), we construct an instance of Matroid-Bin GAP as follows:

1. *Elements and bins:* For each P-item p and buyer j, construct a bin (p, j) in the GAP instance. The elements of the GAP instance are exactly the items of the AVA instance.
2. *Values/sizes of P-items:* Assigning a P-item p to bin (p, j) yields value v_{pj} and uses zero space; Assigning P-item p to a bin (p', j) with $p \neq p'$ yields value zero and uses $1 + \varepsilon$ space.
3. *Values/sizes of N-items:* Assigning N-item i to bin (p, j) yields value v_{ij} and uses $\frac{\rho_j - v_{ij}}{v_{pj} - \rho_j}$ space.
4. *Matroid on the bins:* Finally, the matroid $\mathcal{M}$ on the bins is a partition matroid, requiring that we choose at most one bin from $\{(p, j) \mid j \in B\}$, for each P-item p.

The construction above results in a value-preserving one-to-one correspondence between feasible GAP solutions which are *maximal*, i.e., where each P-item p is assigned to some bin, and permissible bundling-based solutions to the AVA instance. Indeed, for any feasible bundling-based solution to the AVA instance, fix a bundle (p, j) containing the item set S. The value of placing the items in S in the bin (p, j) is precisely $\sum_{i \in S} v_{ij}$. Summing over all bins, we find that both solutions (to the AVA and GAP instance) have the same value. On the other hand, the GAP solution is feasible since for each P-item p we open up at most one bin (p, j) (thus respecting the matroid constraint) and moreover each bin's size constraint is respected due to the per-bundle average-value constraint and the zero size of p in bin (p, j), implying that $\sum_{i \in S} s_{i,(p,j)} = \sum_{i \in S \setminus \{p\}} \frac{\rho_j - v_{ij}}{v_{pj} - \rho_j} \leq 1$. Similarly, starting with a maximal solution to the GAP instance, the single bin (p, j) into which p is placed has its average-value constraint satisfied (note that p cannot be placed in a bin (p', j) for $p' \neq p$, where its size is $1 + \varepsilon$), and the value of the bundles obtained this way is the same as the GAP solution's value. Now the $(1 - 1/e - o(1))$-approximation algorithm for GAP with matroid constraints [13] gives the same approximation for AVA on unambiguous instances. ◀

Theorem 14 combined with Lemma 10 completes the proof of Theorem 13.

4 An Offline Algorithm via Relax-and-Round

Let us now present an LP-rounding based algorithm for AVA. This more sophisticated algorithm yields another constant-approximate offline algorithm, which also allows to incorporate additional side constraints, see Section 4.3). Moreover, this section's algorithm also provides a template for our main *online* algorithms.

The natural starting point for an LP-rounding based algorithm, the LP relaxation obtained by dropping the integrality constraints of (AVA-ILP), turns out to be a dead end. This relaxation has an integrality gap of $\Omega(n)$ on n-buyer instances,[3] even for unit-ρ, as shown by the reinspecting the instance of Example 11.

[3] Recall that an LP relaxation's *integrality gap* is the difference in objective between its best fractional and integral solutions.

▶ **Example 15.** Consider an n-buyer unit-ρ instance with a single P-item p of value $1+n\varepsilon$ for all buyers, and n N-items, with the i-th N-item having zero value for all buyers except for buyer j_i, for whom its value is $1-\varepsilon$. An assignment $x_{pj} = \frac{1}{n}$ for all buyers j and $x_{ij_i} = 1$ for every N-item i gives value $n+1$ for the LP relaxation of (AVA-ILP), while clearly the optimal integral solution has value ≈ 2.

Therefore, to obtain any constant approximation via LP rounding, we need a tighter relaxation. To this end, we rely on Lemmas 6 and 10, and provide the following relaxation for *bundling-based* solutions for unambiguous AVA instances. This LP has decision variables x_{ijp} for (P or N)-item i, buyer j and P-item p. Informally, these correspond to the probability that i is allocated to j in the bundle with P-item p, which we denote by jp. (Note: this polynomially-sized LP is clearly poly-time solvable.)

$$\begin{aligned}
\max \quad & \sum_{i,j,p} v_{ij}\, x_{ijp} & & \text{(Bundle-LP)} \\
\text{s.t.} \quad & \sum_{i} (\rho_j - v_{ij})\, x_{ijp} \le 0 & \forall j,p \quad & (4.4) \\
& \sum_{j,p} x_{ijp} \le 1 & \forall i \quad & (4.5) \\
& x_{ijp} \le x_{pjp} & \forall i,j,p \quad & (4.6) \\
& x_{p'jp} = 0 & \forall j, P\text{-item } p' \ne p \quad & (4.7) \\
& x_{ijp} \ge 0 & \forall i,j,p \quad &
\end{aligned}$$

Intuitively, the bundling, and in particular Equation (4.6), will allow us to overcome the integrality gap example above. We formalize this intuition later by approximately rounding this LP, but first we show that (Bundle-LP) is a relaxation of bundling-based allocations for unambiguous AVA instances.

▶ **Lemma 16.** *For any unambiguous AVA instance, the value of* (Bundle-LP) *is at least as high as that of the optimal bundling-based allocation.*

Proof. Fix a (randomized) bundling-based allocation algorithm $\mathcal{A}$. Let Y_{ijp} be the indicator for $\mathcal{A}$ having allocated item i in bundle jp. We argue that Y_{ijp} satisfy the constraints of (Bundle-LP), realization by realization. Consequently, by linearity of expectation, so do their marginals, $\mathbb{E}[Y_{ijp}]$. Constraint (4.4) holds since $\mathcal{A}$ satisfies the average-value constraint for each bundle. Constraint (4.5) holds since each item is allocated at most once. Constraint (4.6) holds because bundle jp must be opened for i to be allocated in it. Constraint (4.7) holds since permissible bundles have a single P-item in them. Finally, non-negativity of $\mathbf{Y}$ is trivial. We conclude that $\mathbb{E}[\mathbf{Y}]$ is a feasible solution to the above LP, with objective precisely $\sum_{ijp} v_{ij}\, \mathbb{E}[Y_{ijp}]$. The lemma follows. ◀

We now turn to rounding this LP. To this end, we consider a two-phase algorithm, whose pseudo-code is given in Algorithm 1. In Phase I we *open bundles*, letting each P-item p pick a single buyer j with probability x_{pjp},[4] and opening the bundle jp. In Phase II we *enrich the bundles*, by adding N-items to them. Specifically, for each N-item i, we create a set S_i containing each open bundle jp independently with probability $\alpha \cdot \frac{x_{ijp}}{x_{pjp}}$, where $\alpha \in [0,1]$

[4] Since Constraint (4.5) is tight for every P-item in any optimal LP solution, $\{x_{pjp}\}_j$ is a distribution over buyers.

is a parameter to be specified later. Then, if this set S_i contains a single bundle jp and adding i to this bundle would not violate the average-value constraint restricted to the bundle (denoted by $\mathsf{BundleAV}_{jp}$), i.e., this bundle would remain permissible, then we allocate i to the bundle jp. Otherwise, we leave i unallocated.

Algorithm 1 Offline rounding of Bundle-LP.

1: Make the instance unambiguous as in Lemma 10
2: Let $\mathbf{x}$ be an optimal solution to (Bundle-LP) for the obtained unambiguous instance
3: **for each** P-item p **do** ▷ Phase I
4:     Pick j according to distribution $\{x_{pjp}\}_{j=1,\ldots,n}$ and open bundle jp
5: **for each** N-item i **do** ▷ Phase II
6:     $S_i \leftarrow \emptyset$
7:     **for each** bundle jp, **with probability** $\alpha \cdot \frac{x_{ijp}}{x_{pjp}}$ **do**
8:         **if** jp was opened in Phase I **then**
9:             $S_i \leftarrow S_i \cup \{jp\}$
10:     **if** $|S_i| = 1$ **then**
11:         **if** the only bundle $jp \in S_i$ remains permissible after adding i to it **then**
12:             Allocate i to jp

Algorithm 1 clearly outputs a feasible allocation, since it only allocates N-items i to a bundle jp if this would not violate the average-value constraint of the bundle, and hence by linearity the average-value constraint of the buyer remains satisfied. Moreover, the algorithm is well-defined; in particular, the probability spaces defined in lines 4 and 7 are valid, by constraints (4.5) for P-item p, and (4.6) for triple i, j, p, respectively. We turn to analyzing this algorithm's approximation ratio. For this, we will lower bound the probability of each item i to be allocated in bundle jp in terms of x_{ijp}.

By Section 4, each P-item p is assigned in bundle jp precisely with probability x_{pjp}. Consequently, the expected value Algorithm 1 obtains from P-items is precisely their contribution to the LP solution's value. It remains to understand what value we get from N-items.

4.1 Allocation of N-items

To bound the contribution of N-items, we consider any tuple of N-item i, buyer j and P-item p. Note that N-item i is assigned to bundle jp if and only if all the four following events occur:

1. $\mathcal{E}_1$: the event that bundle jp is open, which happens with probability x_{pjp}.
2. $\mathcal{E}_2$: the event that the Bernoulli$(\alpha \cdot \frac{x_{ijp}}{x_{pjp}})$ in Section 4 comes up heads for jp.
3. $\mathcal{E}_3$: the event that $S_i \setminus \bigcup_{j'=1,\ldots,n}\{j'p\} = \emptyset$.
4. $\mathcal{E}_4$: the event that jp would remain permissible if we were to add i to bundle jp.

We note that events $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ are all independent, as they depend on distinct (and independent) coin tosses. So, for example, $\mathbb{P}r[S_i \ni jp] = \mathbb{P}r[\mathcal{E}_1 \wedge \mathcal{E}_2] = \mathbb{P}r[\mathcal{E}_1] \cdot \mathbb{P}r[\mathcal{E}_2] = \alpha \cdot x_{ijp}$. Moreover, we have the following simple bound on $\mathbb{P}r[\mathcal{E}_3]$.

▶ **Lemma 17.** $\mathbb{P}r\left[\bigwedge_{\ell=1}^{3} \mathcal{E}_\ell\right] = \prod_{\ell=1}^{3} \mathbb{P}r[\mathcal{E}_\ell] \geq (1-\alpha) \cdot \alpha \cdot x_{ijp}$.

Proof. The first equality follows from independence of $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$. We therefore turn to lower bounding $\mathbb{P}r[\mathcal{E}_3]$. Since $\mathbb{P}r[X > 0] \leq \mathbb{E}[X]$ for any integer random variable $X \geq 0$, we know

$$\mathbb{P}r[\overline{\mathcal{E}_3}] \leq \mathbb{E}\left[\left|S_i \setminus \bigcup_{j'} \{j'p\}\right|\right] = \sum_{p' \neq p} \sum_{j'} \alpha \cdot x_{ij'p'} \leq \alpha,$$

where the equality follows from $\mathbb{P}r[S_i \ni j'p'] = \alpha \cdot x_{ij'p'}$ by the above, and the last inequality follows from Constraint (4.5). Since $\mathbb{P}r[\mathcal{E}_1] \cdot \mathbb{P}r[\mathcal{E}_2] = \alpha \cdot x_{ijp}$, the lemma follows. ◀

A challenge. As noted above, $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ are independent, resulting in a simple analysis for the probability $\mathbb{P}r\left[\bigwedge_{\ell=1}^{3} \mathcal{E}_\ell\right] = \prod_{\ell=1}^{3} \mathbb{P}r[\mathcal{E}_\ell]$. Unfortunately, lower bounding $\mathbb{P}r[\mathcal{E}_4 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3]$ is more challenging, due to possible *negative correlations* between $\mathcal{E}_4$ and $\mathcal{E}_3$. To see this, note that $\mathcal{E}_3 \wedge \mathcal{E}_1$ implies $S_i = \{jp\}$, and this event can be positively correlated with previous N-items i' having $S_{i'} = \{jp\}$, thus making it more likely that jp won't be able to accommodate i under $\mathsf{BundleAV}_{jp}$.

We can overcome this challenge of negative correlations, provided (i, j) has small deficit compared to (p, j)'s excess. (We address the large deficit case separately later.) Specifically, by coupling our algorithm with an algorithm that allocates more often and does not suffer from such correlations, we can lower bound this conditional probability as follows.

▶ **Lemma 18.** *Let $\beta \in [0, 1]$. If i, j, p are such that $\rho_j - v_{ij} \leq \beta \cdot (v_{pj} - \rho_j)$, then*

$$\mathbb{P}r[\mathcal{E}_4 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] \geq 1 - \frac{\alpha}{1-\beta}.$$

Proof. Consider an imaginary algorithm $\mathcal{A}'$ that allocates every N-item i' into every bundle $j'p' \in S_{i'}$, even when $|S_{i'}| > 1$ (so we may over-allocate) and even if this violates the $\mathsf{BundleAV}_{j'p'}$ constraint. Coupling $\mathcal{A}'$ with Algorithm 1 by using the same randomness for both algorithms, we have that item i' is allocated to bin $j'p'$ by $\mathcal{A}'$ with probability precisely $\mathbb{P}r[S_{i'} \ni j'p'] = \alpha \cdot x_{i'j'p'}$. In particular, $\mathcal{A}'$ only allocates more items than Algorithm 1.

We denote by N'_{jp} the set of N-items allocated to bundle jp by $\mathcal{A}'$. Now, let $\mathcal{E}'_4$ be the event that $\sum_{i' \in N'_{jp} \setminus \{i\}} (\rho_j - v_{i'j}) \leq (1 - \beta) \cdot (v_{pj} - \rho_j)$, that is, the deficit of N-items other than i that $\mathcal{A}'$ allocated to the bundle jp together only consumes at most a $(1-\beta)$ fraction of p's excess for j. By the small deficit assumption on i, j, p, we know that event $\mathcal{E}'_4$ is sufficient for $\mathsf{BundleAV}_{jp}$ to be satisfied if Algorithm 1 were to add i to jp. Thus, $\mathcal{E}'_4$ implies $\mathcal{E}_4$ in any realization (of the randomness), since $\mathcal{A}'$ only allocates more items to each bin than Algorithm 1. On the other hand, we also have that both $\mathcal{E}'_4$ and $\mathcal{E}_1$ are independent of both $\mathcal{E}_2 \wedge \mathcal{E}_3$, since the latter combined event depends on an independent random coin toss ($\mathcal{E}_2$) and events concerning other bundles jp', which are both independent of the randomness concerning bundle jp. (Here we use that $\mathcal{A}'$ allocates i to jp whenever $S_i \ni jp$, regardles of other bundles $j'p'$ belonging to S_i.) Consequently, by standard applications of Bayes' Law, we obtain the following.

$$\mathbb{P}r[\mathcal{E}'_4 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] = \mathbb{P}r[\mathcal{E}'_4 \mid \mathcal{E}_1].$$

As the imaginary algorithm $\mathcal{A}'$ assigns i' to jp (i.e. $i' \in N'_{jp}$) iff $S_{i'} \ni jp$, we know that

$$\begin{aligned}\mathbb{E}\left[\sum_{i' \in N'_{jp}} (\rho_j - v_{i'j}) \,\middle|\, \mathcal{E}_1\right] &= \sum_{i' \neq p} (\rho_j - v_{i'j}) \cdot \mathbb{P}r[S_{i'} \ni jp \mid \mathcal{E}_1] \\ &= \alpha \cdot \sum_{i' \neq p} (\rho_j - v_{i'j}) \frac{x_{i'jp}}{x_{pjp}} \leq \alpha \cdot (v_{pj} - \rho_j).\end{aligned}$$

Above, the second equality follows from linearity and $\mathbb{P}r[S_{i'} \ni jp \mid \mathcal{E}_1] = \alpha \cdot \frac{x_{ij'p'}}{x_{pjp}}$, and the inequality follows from the average-value constraint for bundle jp (i.e. Equation (4.4)) in our LP. Therefore, by Markov's inequality

$$\mathbb{P}r\left[\sum_{i' \in N'_{jp} \setminus \{i\}} (\rho_j - v_{i'j}) > (1-\beta) \cdot (v_{pj} - \rho_j) \;\middle|\; \mathcal{E}_1\right] \leq \frac{\mathbb{E}\left[\sum_{i' \in N'_{jp} \setminus \{i\}} (\rho_j - v_{i'j}) \;\middle|\; \mathcal{E}_1\right]}{(1-\beta) \cdot (v_{pj} - \rho_j)} \leq \frac{\alpha}{1-\beta},$$

and thus $\mathbb{P}r\left[\mathcal{E}'_4 \mid \mathcal{E}_1\right] \geq 1 - \frac{\alpha}{1-\beta}$. Recalling that $\mathcal{E}'_4$ implies $\mathcal{E}_4$ in any realization, we conclude with the desired bound, as follows.

$$\mathbb{P}r[\mathcal{E}_4 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] \geq \mathbb{P}r[\mathcal{E}'_4 \mid \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] = \mathbb{P}r[\mathcal{E}'_4 \mid \mathcal{E}_1] \geq 1 - \frac{\alpha}{1-\beta}.$$ ◀

Lemma 18 and the preceding discussion yield a lower bound on the probability of an N-item i being successfully allocated to a bundle jp when i's deficit is small relative to the excess of the P-item p. For the large deficit case, no such bound holds. However, as we now observe (with proof deferred to the full version), large-deficit edges contribute a relative small portion of the allocation's value in the optimal LP solution.

▶ **Lemma 19.** *Let $\beta \in [0,1]$. For any bundle jp, let L^{β}_{jp} denote the set of β-large* deficit *N-items for bundle jp, i.e., N-item i with $\rho_j - v_{ij} > \beta \cdot (v_{pj} - \rho_j)$. Then,*

$$\sum_{j,p} \sum_{i \in L^{\beta}_{jp}} v_{ij}\, x_{ijp} \leq \frac{1}{\beta} \sum_{j,p} v_{pj}\, x_{pjp}.$$

4.2 Completing the analysis

We are now ready to bound the approximation ratio of Algorithm 1.

▶ **Theorem 20.** *Algorithm 1 with $\alpha = 0.3$ is a 32-approximation for AVA.*

Proof. Let $\beta \in [0,1]$ be some constant to be determined and let $\gamma = \gamma(\alpha, \beta) := \alpha \cdot (1-\alpha) \cdot \left(1 - \frac{\alpha}{1-\beta}\right)$. Denote N_{jp} by the set of N-items allocated to bundle jp by the algorithm. By Lemmas 17 and 18 we have for bundle jp and N-item $i \notin L^{\beta}_{jp}$ that

$$\mathbb{P}r[i \in N_{jp}] = \mathbb{P}r\left[\mathcal{E}_4 \;\middle|\; \bigwedge_{\ell=1}^{3} \mathcal{E}_\ell\right] \mathbb{P}r\left[\bigwedge_{\ell=1}^{3} \mathcal{E}_\ell\right] \geq \left(1 - \frac{\alpha}{1-\beta}\right) \cdot \alpha \cdot (1-\alpha) \cdot x_{ijp} = \gamma \cdot x_{ijp}.$$

Therefore, by linearity of expectation and Lemma 19, the expected value of the (feasible) random allocation of Algorithm 1 is at least

$$\sum_{j,p} v_{pj}\, x_{pjp} + \gamma \sum_{i,j,p: i \neq p} v_{ij}\, x_{ijp} - \gamma \sum_{j,p} \sum_{i \in L^{\beta}_{jp}} v_{ij}\, x_{ijp}$$
$$\geq \left(1 - \frac{\gamma}{\beta}\right) \sum_{j,p} v_{pj}\, x_{pjp} + \gamma \sum_{i,j,p: i \neq p} v_{ij}\, x_{ijp}.$$

So, this algorithm's output has value at least a $\min\{1 - \frac{\gamma}{\beta},\, \gamma\}$ fraction of the optimal LP value; i.e., it is a $1/\min\{1 - \frac{\gamma}{\beta},\, \gamma\}$-approximation. Taking $\alpha \approx 0.3$ and $\beta \approx 0.156$ (optimized by an off-the-shelf numerical solver) yields a ratio of $1/0.13 < 8$. The theorem then follows from Lemma 16 and Lemma 10. ◀

4.3 Extension: adding side constraints

Before moving on to our online algorithms, we note that the LP-based approach allows us to incorporate additional constraints seamlessly. For example, our LP and algorithm, with minor modifications, allow to approximate allocation problems with both the average-value constraint and $O(1)$ many budget constraints (for every buyer), corresponding to different resources. More formally, for a cost function ℓ (e.g., corresponding to storage, time, or other costs), each buyer j has some budget $B_j^{(\ell)}$, and the ℓ-cost of allocation to buyer j must not exceed this budget. That is, for $x_{ij} \in \{0,1\}$ an indicator for item i being allocated to buyer j, we have

$$\forall j, \quad \ell\text{-cost}_j = \sum_i \ell_{ij}\, x_{ij} \le B_j^{(\ell)}. \tag{4.8}$$

The *small-cost* assumption (a.k.a. the small-bids assumption for online AdWords [32]) stipulates that no particular item has high cost compared to the budget, i.e. $\max_{ij} \ell_{ij}/B_j^{(\ell)} \le \varepsilon \to 0$.

▶ **Theorem 21.** *There exists a constant-approximate algorithm for AVA and any constant number of budget constraints (for every buyer) subject to the small-bids assumption.*

We defer the proof of the above result to the full version. The same arguments in this section extend to our online algorithms, but are omitted for brevity.

5 Online Algorithms: Approximating the Online Optimum

In this section and the next we study AVA in the online i.i.d. setting (see Section 1.3 for definition and notation). Specifically, in this section we provide a polynomial-time online algorithm which provides a constant approximation of the optimal online algorithm.

First, by Lemma 6, we have that the optimal online algorithm is approximated within a factor two by a bundling-based online algorithm which is *committed*. As we will show, the following LP provides a relaxation for the value of the best such online algorithm. Our LP consists of variables x_{ijp} for each item type $i \in [m]$, buyer $j \in [n]$ and item type p such that (p, j) is a P-edge.

$$\begin{aligned}
\max \quad & \sum_{i,j,p} v_{ij}\, x_{ijp} && && \text{(OPTon-Bundle-LP)}\\
\text{s.t.} \quad & \sum_i (\rho_j - v_{ij})\, x_{ijp} \le 0 && \forall\ P\text{-edge type } (p,j) && (5.9)\\
& \sum_{j,p} x_{ijp} \le q_i \cdot T && \forall \text{ item type } i && (5.10)\\
& x_{ijp} \le x_{pjp} \cdot q_i \cdot T && \forall\ N\text{-edge type } (i,j), P\text{-edge type } (p,j) && (5.11)\\
& x_{p'jp} = 0 && \forall\ P\text{-edge types } (p,j) \ne (p',j) && (5.12)\\
& x_{ijp} \ge 0 && \forall \text{ item type } i,\ P\text{-edge type } (p,j) &&
\end{aligned}$$

▶ **Lemma 22.** (OPTon-Bundle-LP) *has value which is at least half the expected value of any online AVA algorithm under i.i.d. arrivals (from the same distribution used in the LP), where item type i is drawn with probability q_i.*

Proof. First, by the Online Bundling Lemma (Lemma 8), the best committed online bundling-based algorithm 2-approximates the best online algorithm. We therefore turn to showing that (OPTon-Bundle-LP) is a relaxation of the value of the best committed bundling-based online algorithm, $\mathcal{A}$. Let x_{ijp} be the average number of times a copy of item type i is allocated in a copy of bundle jp by $\mathcal{A}$. Constraint (5.9) follows by linearity of expectation, together with the fact that each opened copy of bundle jp must satisfy the average-value constraint. Constraint (5.10) simply asserts that i is allocated at most as many times as it arrives. Constraint (5.11) holds for a committed online algorithm (that guarantees feasibility with probability 1), for the following reason: for every copy of bundle jp opened, no items can be placed in that bundle before it is opened. But the expected number of copies of i to be assigned after any bundle jp is opened is at most the number of arrivals of i after this bundle is opened and is at most $q_i \cdot T$, which upper-bounds the ratio between x_{ijp} and x_{pjp}. All other constraints hold similarly to their counterparts in the proof of Lemma 16. ◀

Note. Constraint (5.11) is reminiscent of constraints bounding the optimal online algorithm in the secretary problem literature [12] and prophet inequality literature [34].

The outline of our algorithm is similar to that of Algorithm 1, though as it does not have random access to the different items throughout, it first allocates P-edges in the first $T/2$ arrivals, and only then allocates N-edges in the last $T/2$ arrivals. To distinguish between bundles opened at different times, we now label copies of bundle type jp (i.e., items allocated to buyer j with single P-edge of type (p, j)) opened at time t by jpt. The algorithm's pseudocode is given in Algorithm 2.

Note that in our online algorithms (here and in Appendix A), the LPs are based on distributions that can be ambiguous in the sense that each item type in the distribution can have both P-edges and N-edges, and we don't explicitly modify the distribution to make it unambiguous. However, our algorithm effectively makes each realized instance (of T sampled items) unambiguous, as we ignore all N-edges incident to the first $T/2$ items and vice versa for the last $T/2$ items.

Algorithm 2 Online rounding of bundling-based LP.

Let $\mathbf{x}$ be an optimal solution to Equation (OPTon-Bundle-LP)
for all arrivals $t = 1, \dots, T/2$, of type p **do**
    Pick a j according to the distribution $\{\frac{x_{pjp}}{q_p \cdot T}\}_{j=1,\dots,n}$ and open bundle jpt
for all arrival $t^\star = T/2+1, \dots, T$ of type i **do**
    $S_{it^\star} \leftarrow \emptyset$
    for all bundles jpt, **with probability** $\frac{\alpha \cdot x_{ijp}}{x_{pjp} \cdot q_i \cdot T}$ **do**
        if bundle jpt is open **then**
            $S_{it^\star} \leftarrow S_{it^\star} \cup \{jpt\}$
    if $|S_{it^\star}| = 1$ **then**
        if $jpt \in S_{it^\star}$ remains permissible after adding $it^\star$ to it **then**
            Allocate $it^\star$ to jpt

5.1 Analysis

In what follows we provide a brief overview of the relevant events in the analysis of Algorithm 2, deferring proofs reminiscent of the analysis of Algorithm 1 to the full version.

First, the value obtained from P-edges by Algorithm 2 is clearly half that of the LP, by linearity of expectation. In particular, we create $x_{pjp}/2$ copies of bundle jp in expectation. The crux of the analysis is in bounding our gain from N-edges.

To bound the contribution of N-edges, we note that a copy of item i at time $t^\star > T/2$, which we denote by $it^\star$, is assigned to bundle jpt if and only if all the five following events (overloading notation from Section 4) occur:

1. $\mathcal{E}_0$: the event that $it^\star$ is the realized item at time $t^\star$, which happens with probability q_i
2. $\mathcal{E}_1$: the event that bundle jpt is open, which happens with probability $q_p \cdot \frac{x_{pjp}}{q_p \cdot T} = \frac{x_{pjp}}{T}$.
3. $\mathcal{E}_2$: the event that the Bernoulli($\frac{\alpha \cdot x_{ijp}}{x_{pjp} \cdot q_i \cdot T}$) in Section 5 comes up heads for jpt.
4. $\mathcal{E}_3$: the event that $S_{it^\star} \setminus \bigcup_{j'p'} \bigcup_{t' \neq t} \{j'p't'\} = \emptyset$.
5. $\mathcal{E}_4$: the event that jpt would remain permissible if we were to add $it^\star$ to bundle jpt.

Similarly to the events we studied when anlyzing our offline Algorithm 1, the events $\mathcal{E}_0, \mathcal{E}_1, \mathcal{E}_2$ are independent, as are the events $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$. However, $\mathcal{E}_3$ is not independent of $\mathcal{E}_0$ (in particular, it occurs trivially if $\mathcal{E}_0$ does not). Nonetheless, bounding $\mathbb{P}r\left[\bigwedge_{\ell=0}^{3} \mathcal{E}_\ell\right]$ is not too hard. The following lemma, whose proof essentially mirrors that of Lemma 17, and is thus deferred to the full version, provides a bound on the probability of all first four events occurring.

▶ **Lemma 23.** $\mathbb{P}r[\mathcal{E}_0 \wedge \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] \geq \alpha \cdot (1 - \alpha/2) \cdot \frac{x_{ijp}}{T^2}$.

As with our offline Algorithm 1, the challenge in the analysis is due to possible negative correlations between $\mathcal{E}_4$ and $\mathcal{E}_3$. Similarly, we overcome this challenge of negative correlations, provided (i, j) has small deficit compared to (p, j)'s excess, by coupling with an algorithm with no such correlations. (We address large-deficit (i, j) later.) The obtained syntactic generalization of Lemma 18, whose proof is deferred to the full version, is the following.

▶ **Lemma 24.** *Let $\beta \in [0, 1]$. If i, j, p are such that $\rho_j - v_{ij} \leq \beta \cdot (v_{pj} - \rho_j)$, then*

$$\mathbb{P}r[\mathcal{E}_4 \mid \mathcal{E}_0 \wedge \mathcal{E}_1 \wedge \mathcal{E}_2 \wedge \mathcal{E}_3] \geq 1 - \frac{\alpha}{2(1-\beta)}.$$

Lemma 24 and the preceding discussion yield a lower bound on the probability of a copy of item i be allocated to a bundle jpt at time $t^\star$ if i, j, p is in the small deficit case as the above lemma. For large-deficit items, no such bound holds. However, large-deficit edges contribute a small portion of the allocation's value. Specifically, Lemma 19, holds for (OPTon-Bundle-LP) as well, since the only constraint that this lemma's proof relied on was Constraint (4.4), which is identical to Constraint (5.9) in (OPTon-Bundle-LP).

We are now ready to bound the approximation ratio of Algorithm 1.

▶ **Theorem 25.** *Algorithm 1 with $\alpha = 0.64$ is a polynomial-time algorithm achieving a 57-approximation of the optimal online algorithm for AVA under known i.i.d. arrivals.*

Proof. That the algorithm runs in polynomial time follows from its description, together with the LP (OPTon-Bundle-LP) having polynomial size (in the distribution size). The analysis is essentially identical to that of Theorem 20, with the following differences. First, we recall that the expected number of copies of bundle jp opened is $\frac{T}{2} \cdot q_p \cdot \frac{x_{pjp}}{q_p \cdot T} = \frac{1}{2}\, x_{pjp}$. Next, by Lemmas 23 and 24, the probability that copy $it^\star$ of small-deficit item i for bundle jpt is allocated to it is at least $\gamma \cdot \frac{x_{ijp}}{T^2}$, for $\gamma = \gamma(\alpha, \beta) := \frac{\alpha}{2} \cdot \left(1 - \frac{\alpha}{2}\right) \cdot \left(1 - \frac{\alpha}{2(1-\beta)}\right)$. Again, linearity of expectation and summation over all $(t, t^\star) \in [T/2] \times (T/2, T]$ in combination with Lemma 19 implies that for any $\beta \in [0, 1]$, the gain of Algorithm 2 is at least

$$\frac{1}{2}\left(\sum_{j,p} v_{pj}\, x_{pjp} + \frac{\gamma}{4} \sum_{i,j,p:i\neq p} v_{ij}\, x_{ijp} - \frac{\gamma}{4}\sum_{j,p}\sum_{i\in L^{\beta}_{jp}} v_{ij}\, x_{ijp}\right)$$
$$\geq \left(\left(\frac{1}{2} - \frac{\gamma}{4\beta}\right) \sum_{j,p} v_{pj}\, x_{pjp} + \frac{\gamma}{4} \sum_{i,j,p:i\neq p} v_{ij}\, x_{ijp}\right).$$

Therefore, by Lemma 22, Algorithm 2 yields a $2/\min\{\frac{1}{2} - \frac{\gamma}{4\beta}, \frac{\gamma}{4}\}$-approximation. This expression is optimized by $\alpha \approx 0.64$ and $\beta \approx 0.0766$, yielding a ratio of $\approx \frac{2}{0.0355} < 57$, as claimed. ◀

References

1. Gagan Aggarwal, Ashwinkumar Badanidiyuru, and Aranyak Mehta. Autobidding with constraints. In *International Conference on Web and Internet Economics*, pages 17–30. Springer, 2019.
2. Shipra Agrawal and Nikhil R Devanur. Fast algorithms for online stochastic convex programming. In *Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms*, pages 1405–1424, 2014.
3. Nima Anari, Rad Niazadeh, Amin Saberi, and Ali Shameli. Nearly optimal pricing algorithms for production constrained and laminar bayesian selection. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 91–92, 2019.
4. Nick Arnosti and Will Ma. Tight guarantees for static threshold policies in the prophet secretary problem. *Operations Research*, 2022.
5. Makis Arsenis and Robert Kleinberg. Individual fairness in prophet inequalities. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, page 245, 2022.
6. Yossi Azar, Andrei Z. Broder, Anna R. Karlin, and Eli Upfal. Balanced allocations. *SIAM J. Comput.*, 29(1):180–200, 1999.
7. Yossi Azar, Niv Buchbinder, TH Hubert Chan, Shahar Chen, Ilan Reuven Cohen, Anupam Gupta, Zhiyi Huang, Ning Kang, Viswanath Nagarajan, Joseph Naor, et al. Online algorithms for covering and packing problems with convex objectives. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 148–157. IEEE, 2016.
8. Santiago Balseiro, Yuan Deng, Jieming Mao, Vahab Mirrokni, and Song Zuo. Robust auction design in the auto-bidding world. *Advances in Neural Information Processing Systems*, 34:17777–17788, 2021.
9. Kshipra Bhawalkar, Zhe Feng, Anupam Gupta, Aranyak Mehta, David Wajc, and Di Wang. The average-value allocation problem, 2024. `arXiv:2407.10401`.
10. Mark Braverman, Mahsa Derakhshan, and Antonio Molina Lovett. Max-weight online stochastic matching: Improved approximations against the online benchmark. In *23rd ACM Conference on economics and Computation*, pages 967–985, 2022.
11. Niv Buchbinder, Kamal Jain, and Joseph Seffi Naor. Online primal-dual algorithms for maximizing ad-auctions revenue. In *European Symposium on Algorithms*, pages 253–264, 2007.
12. Niv Buchbinder, Kamal Jain, and Mohit Singh. Secretary problems via linear programming. *Mathematics of Operations Research*, 39(1):190–206, 2014.
13. Gruia Calinescu, Chandra Chekuri, Martin Pal, and Jan Vondrák. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6):1740–1766, 2011.
14. Deeparnab Chakrabarty and Gagan Goel. On the approximability of budgeted allocations and improved lower bounds for submodular welfare maximization and gap. *SIAM Journal on Computing (SICOMP)*, 39(6):2189–2211, 2010.

15 Yuan Deng, Jieming Mao, Vahab Mirrokni, and Song Zuo. Towards efficient auctions in an auto-bidding world. In *Proceedings of the Web Conference 2021*, WWW '21, pages 3965–3973. Association for Computing Machinery, 2021.

16 Nikhil R. Devanur and Thomas P. Hayes. The adwords problem: online keyword matching with budgeted bidders under random permutations. In *ACM Conference on Electronic Commerce*, pages 71–78, 2009. doi:10.1145/1566374.1566384.

17 Nikhil R. Devanur and Kamal Jain. Online matching with concave returns. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 137–144. ACM, 2012. doi:10.1145/2213977.2213992.

18 Nikhil R. Devanur, Kamal Jain, Balasubramanian Sivan, and Christopher A. Wilkens. Near optimal online algorithms and fast approximation algorithms for resource allocation problems. In *ACM Conference on Electronic Commerce*, pages 29–38, 2011. doi:10.1145/1993574.1993581.

19 Tomer Ezra, Michal Feldman, Nick Gravin, and Zhihao Gavin Tang. "who is next in line?" on the significance of knowing the arrival order in bayesian online settings. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3759–3776, 2023.

20 Auto-bidding products support page. https://www.facebook.com/business/help/1619591734742116, 2022. Accessed: 2023-07-12.

21 Uriel Feige. A threshold of $\ln n$ for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

22 Jon Feldman, Monika Henzinger, Nitish Korula, Vahab S. Mirrokni, and Clifford Stein. Online stochastic packing applied to display ad allocation. In *ESA (1)*, pages 182–194, 2010. doi:10.1007/978-3-642-15775-2_16.

23 Lisa Fleischer, Michel X Goemans, Vahab S Mirrokni, and Maxim Sviridenko. Tight approximation algorithms for maximum separable assignment problems. *Mathematics of Operations Research*, 36(3):416–431, 2011.

24 Negin Golrezaei, Ilan Lobel, and Renato Paes Leme. Auction design for roi-constrained buyers. In *Proceedings of the Web Conference 2021*, WWW '21, pages 3941–3952, 2021.

25 Auto-bidding products support page. https://support.google.com/google-ads/answer/2979071, 2022. Accessed: 2023-07-12.

26 Anupam Gupta and Marco Molinaro. How the experts algorithm can help solve LPs online. *Math. Oper. Res.*, 41(4):1404–1431, 2016.

27 Johan Håstad. Clique is hard to approximate within $n^{1-\epsilon}$. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 627–636, 1996.

28 Zhiyi Huang, Qiankun Zhang, and Yuhao Zhang. Adwords in a panorama. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1416–1426, 2020.

29 Kristen Kessel, Ali Shameli, Amin Saberi, and David Wajc. The stationary prophet inequality problem. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 243–244, 2022.

30 Thomas Kesselheim, Klaus Radke, Andreas Tönnis, and Berthold Vöcking. Primal beats dual on online packing lps in the random-order model. *SIAM J. Comput.*, 47(5):1939–1964, 2018.

31 Aranyak Mehta. Auction design in an auto-bidding setting: Randomization improves efficiency beyond VCG. In *Proceedings of the ACM Web Conference 2022*, pages 173–181, 2022.

32 Aranyak Mehta, Amin Saberi, Umesh Vazirani, and Vijay Vazirani. Adwords and generalized online matching. *Journal of the ACM (J.ACM)*, 54(5):22, 2007.

33 Joseph (Seffi) Naor, Aravind Srinivasan, and David Wajc. Online dependent rounding schemes. *arXiv preprint arXiv:2301.08680*, 2023.

34 Christos Papadimitriou, Tristan Pollner, Amin Saberi, and David Wajc. Online stochastic max-weight bipartite matching: Beyond prophet inequalities. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 763–764, 2021.

35 Sebastian Perez-Salazar, Mohit Singh, and Alejandro Toriello. The iid prophet inequality with limited flexibility. *arXiv preprint arXiv:2210.05634*, 2022.

36 David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 681–690, 2006.

A Online Algorithms: Approximating the Offline Optimum

In this section we look at the lower and upper bounds of the competitive ratio for online algorithms, i.e. the approximation of the ex-post optimum allocation's value, and we consider both the adversarial and i.i.d. cases.

Adversarial arrival. In this setting, we note that no online algorithm can be $o(T)$-competitive. To see this, consider the unit-ρ instance where the first $T-1$ arriving items have value $1-\varepsilon$ for all $n = T$ buyers, followed by a single item at the end with value $1+\varepsilon T$ for a single adversarially chosen buyer and value 0 for all other buyers. Any online algorithm cannot allocate any of the first $T-1$ items due to the average-value constraint, and thus can only get value $1+\varepsilon T$ from the last item. In contrast, the ex-post optimum can allocate all items to one buyer and collect value $T+1-\varepsilon$. On the other hand, a competitive ratio of T is trivial to achieve for online AVA, by simply allocating any item i with a P-edge (i,j) greedily to the buyer j yielding the highest value. This is a feasible allocation and has value equal to the highest-valued edge in the T-item instance, which is obviously at least a $1/T$ fraction of the optimal allocation's value.

The rest of this section will therefore be dedicated to AVA with i.i.d. arrivals, as in Section 5, but now focusing on approximating the ex-post optimum. We start with the following result lower bounding the competitive ratio.

▶ **Lemma 26.** *There exists a family of uniform online i.i.d. unambiguous unit-ρ AVA instances with $n = m = T \geq 2$ growing, on which every online algorithm's approximation ratio of the ex-post optimum is at least $\Omega\left(\frac{\ln n}{\ln\ln n}\right) = \Omega\left(\frac{\ln m}{\ln\ln m}\right) = \Omega\left(\frac{\ln T}{\ln\ln T}\right)$.*

Proof. Let $\varepsilon = \frac{1}{T}$. Consider an instance with T buyers $j_1, \ldots, j_T$, where all buyers have $\rho = 1$, and T item types. Each item type $i \in [T-1]$ is an N-item, with value $1-\varepsilon$ for buyer j_i and value zero for all others. (So, buyer j_T has zero value for all N-items.) The single P-item type T has value $1+\varepsilon T$ for all buyers. The T arrival types are drawn uniformly from these T types, and consequently there is a single arrival of each type in expectation. Now, an online algorithm (that guarantees average-value constraints in any outcome) can only allocate N-items to a buyer after the buyer was allocated a P-item. But since each N-item appears only once in expectation (and hence at most once after the arrival of a P-item type), each allocation of a P-item (and N-items) to a buyer yields expected value at most $1+\varepsilon T+1-\varepsilon = 3-\varepsilon$ to an online algorithm. Since only one P-item arrives in expectation, an online algorithm accrues value at most $3-\varepsilon$ in expectation on this instance family.

In contrast, the event $\mathcal{E}$ that a single P-item arrived satisfies $\mathbb{P}r[\mathcal{E}] = T \cdot \frac{1}{T} \cdot (1-\frac{1}{T})^{T-1} \geq (1-\frac{1}{T})^T \geq \frac{1}{4}$. Conditioned on $\mathcal{E}$, we have a multi-nomial distribution for the number of arrivals A_i's of the N-item types. Therefore, by standard anti-concentration arguments for the classic balls and bins process [6], we have

$$\mathbb{P}r\left[\max_i A_i \geq \frac{\ln T}{\ln\ln T} - 1 \;\middle|\; \mathcal{E}\right] = 1 - o(1).$$

Consequently, the offline algorithm which, if event $\mathcal{E}$ occurs, allocates the single P-item and all copies of $i^\star := \arg\max_i A_i$ to $j_{i^\star}$ yields expected value at least $\mathbb{E}[\max_i A_i \mid \mathcal{E}] \cdot \mathbb{P}r[\mathcal{E}] = \Omega\left(\frac{\ln T}{\ln\ln T}\right)$. Consequently, this asymptotic ratio also lower bounds any online algorithm's approximation ratio of the ex-post optimum. The full lemma statement follows, since $n = m = T$. ◀

A.1 A matching algorithm assuming constant expected arrivals

Lemma 26 relied on anti-concentration. If the expected number of arrivals A_i of each item type i is at least some constant $\Gamma > 0$, namely $\mathbb{E}[A_i] = q_i \cdot T \geq \Gamma$ (e.g., in Lemma 26 we had $q_i \cdot T = 1$ for every i), then this anti-concentration is tight. In particular, we have the following, by standard Chernoff bounds and union bound (see the full version for proof).

▶ **Observation 27.** *If $\mathbb{E}[A_i] \geq \Gamma$ for all $i \in [m]$ and $\kappa := \frac{6}{\min(1,\Gamma)} \cdot \frac{\ln T}{\ln\ln T}$, then*

$$\mathbb{P}r\left[\max_i A_i \geq \kappa \cdot q_i \cdot T\right] \leq \frac{1}{T^2}.$$

We will show that if the distribution satisfies the assumption on all $\mathbb{E}[A_i] \geq \Gamma = \Theta(1)$, we can show an asymptotically matching upper-bound $O(\frac{\ln T}{\ln\ln T})$ of the competitive ratio.

Our first ingredient towards this proof will, naturally, be another LP, this time capturing possible anti-concentration of arrivals. Similar to (OPTon-Bundle-LP), the LP has one variable x_{ijp} for each item type $i \in [m]$, buyer $j \in [n]$ and item type p such that (p, j) is a P-edge.

$$\begin{aligned}
\max \quad & \sum_{i,j,p} v_{ij}\, x_{ijp} && && \text{(OPToff-Bundle-LP)}\\
\text{s.t.} \quad & \sum_i (\rho_j - v_{ij})\, x_{ijp} \leq 0 && \forall\ P\text{-edge type } (p,j) && \text{(A.13)}\\
& \sum_{jp} x_{ijp} \leq 2 \cdot \lceil q_i \cdot T \rceil && \forall \text{ item type } i && \text{(A.14)}\\
& x_{ijp} \leq x_{pjp} \cdot \lceil q_i \cdot T \cdot \kappa \rceil && \forall\ N\text{-edge type } (i,j), P\text{-edge type } (p,j) && \text{(A.15)}\\
& x_{p'jp} = 0 && \forall\ P\text{-edge types } (p,j) \neq (p',j) && \text{(A.16)}\\
& x_{ijp} \geq 0 && \forall \text{ item type } i,\ P\text{-edge type } (p,j) &&
\end{aligned}$$

▶ **Lemma 28.** *Fix an AVA instance with i.i.d. arrivals satisfying $q_i \cdot T \geq \Gamma = \Theta(1)$ for all $i \in [m]$. Let OPT be the ex-post optimal value and let $V[\mathsf{OFF}]$ be the value of* (OPToff-Bundle-LP)*. Then,*

$$\mathbb{E}[\mathsf{OPT}] \leq O(V[\mathsf{OFF}]).$$

Proof. By Lemma 16, we can restrict to the optimal ex-post bundling-based solution and just lose a factor of 2 in the approximation ratio. We start with a trivial upper-bound on the value of OPT in any outcome of the i.i.d. arrivals. Consider the instance with exactly one copy of each item type from the support of the distribution. The best bundling-based offline solution for this instance is upper-bounded by (Bundle-LP) (Lemma 16), and this value is clearly upper bounded by $V[\mathsf{OFF}]$ since the constraints for (Bundle-LP) are tighter than those of (OPToff-Bundle-LP). Under T i.i.d. arrivals, each item can appear at most T times, and thus by the Supply Lemma (Lemma 12) applied to the instance with a single occurrence per item type, we find that the following bound holds deterministically.

$$\mathsf{OPT} \leq O(T^2) \cdot V[\mathsf{OFF}].$$

Next, let $\mathcal{E}$ be the event that no item type i has more than $\lceil q_i \cdot T \cdot \kappa \rceil$ arrivals. By Observation 27, $\mathbb{P}r[\mathcal{E}] \geq 1 - \frac{1}{T^2}$. Conditioned on $\mathcal{E}$, consider the expected number of times (over the randomness of the i.i.d. arrivals) that the ex-post optimal bundling-based solutions allocate an item of type i to a copy of bundle jp, and denote this value by x_{ijp}. We will argue that such x_{ijp}'s form a feasible solution for (OPToff-Bundle-LP). Since the expected value of the ex-post optimal bundling-based solution conditioned on $\mathcal{E}$ is simply $\sum_{i,j,p} v_{ij}\ x_{ijp}$, this immediately gives that $\mathbb{E}[\mathsf{OPT} \mid \mathcal{E}] \leq 2 \cdot V[\mathsf{OFF}]$.

The proof that x_{ijp} constructed above is feasible follows essentially the same argument as Lemma 22. The average-value constraint (A.13) holds by linearity of expectation because the ex-post (bundling-based) optimum for any outcome satisfies the average-value constraint. Constraint (A.14) holds since the expected times we allocate items of type i cannot exceed i's expected number of occurrences, which is bounded by $\mathbb{E}[A_i \mid \mathcal{E}] \leq \frac{\mathbb{E}[A_i]}{\mathbb{P}r[\mathcal{E}]} \leq \frac{q_i \cdot T}{1-1/T^2} \leq 2 \cdot q_i \cdot T \leq 2 \cdot \lceil q_i \cdot T \rceil$. Constraint (A.15) holds since whenever a bundle jp is opened in the ex-post optimum for any outcome, conditioned on $\mathcal{E}$ we have at most $q_i \cdot T \cdot \kappa$ items of type i, which is a trivial upperbound on how many items of type i can be allocated to bundle jp, and thus cap the ratio between x_{ijp} and x_{pjp}.

Combining the above arguments together with linearity of expectation, the lemma follows.

$$\mathbb{E}[\mathsf{OPT}] = \mathbb{E}[\mathsf{OPT}|\mathcal{E}] \cdot \mathbb{P}r[\mathcal{E}] + \mathbb{E}\left[\mathsf{OPT}|\overline{\mathcal{E}}\right] \cdot \mathbb{P}r\left[\overline{\mathcal{E}}\right] \leq O(V[\mathsf{OFF}]). \qquad \blacktriangleleft$$

We make the simple observation that (OPTon-Bundle-LP) and (OPToff-Bundle-LP) only differ at the RHS of the constraints, with the most crucial difference being in the constraints upper bounding x_{ijp}/x_{pjp}, where they differ by a factor of $\frac{\lceil q_i \cdot T \cdot \kappa \rceil}{q_i \cdot T} = O(\kappa)$ (using that $\Gamma = \Omega(1)$). As we prove in the full version, scaling down any feasible solution of the latter LP by $O(\kappa)$ yields a feasible solution to the former LP, leading to the following observation.

▶ **Observation 29.** *Fix an AVA instance with i.i.d. arrivals, satisfying $q_i \cdot T \geq \Gamma = \Theta(1)$ for all item type i. Then, $V[\mathsf{OFF}]$ and $V[\mathsf{ON}]$, the values of* (OPToff-Bundle-LP) *and* (OPTon-Bundle-LP) *(respectively) satisfy $V[\mathsf{OFF}] \leq O\left(\frac{\ln T}{\ln \ln T}\right) \cdot V[\mathsf{ON}]$*

In our proof of Theorem 25, we showed that Algorithm 2 achieves value at least $\Omega(V[\mathsf{ON}])$. Consequently, Lemma 28 and Observation 29 imply the following result.

▶ **Theorem 30.** *Algorithm 2 is an $O\left(\frac{\ln T}{\ln \ln T}\right)$-competitive online algorithm for AVA under T known i.i.d. arrivals with each item type arriving an expected constant number of times.*

▶ Remark 31. Under the stronger assumption that $\mathbb{E}[A_i] = q_i \cdot T = \Omega(\ln(mT)/\varepsilon^2)$ for each of the m item types i (e.g., if T grows while the distribution $\{q_i\}$ remains fixed), the number of arrivals of each item is more concentrated: it is $\mathbb{E}[A_i] \cdot (1 \pm \varepsilon)$ w.h.p. Consequently, natural extensions of the arguments above, with a smaller blow-up of the RHS of the constraints in (OPTon-Bundle-LP), imply that Algorithm 2's competitive ratio improves to $O(1)$ in this case.

B Hardness Results

In this section we provide hardness of approximation results for AVA and stark impossibility results for the generalization to GenAVA.

B.1 Max-Coverage hardness of AVA

Here we prove that AVA is as hard as the Max-Coverage problem, even if restricted to the unit-ρ case.

▶ **Theorem 32** (Hardness of AVA). *For any constant $\varepsilon > 0$, it is NP-hard to approximate AVA to a factor better than $\left(\frac{e}{e-1} + \varepsilon\right)$ even for unit-ρ instances.*

Proof. We give a reduction from "balanced" instances of the Max-Coverage problem. Such an instance consists of a set system with n elements and m sets, with each set containing n/k elements. A classic result of [21] shows that for each $\delta > 0$, there exist n and $k \le n\delta$, such that it is NP-hard to distinguish between the following two cases: (a) there exists a perfect partition, i.e., k sets in the set system that cover all n elements (YES-instances), and (b) no collection of k sets from the set system cover more than $n(1 - 1/e + \delta)$ elements (NO-instances). We now define a unit-ρ AVA instance consisting of:

1. m buyers, where each buyer i_S corresponds to a set S in the set system,
2. k identical *choice items*, which have value $1 + (\varepsilon/2) \cdot n/k$ for every buyer, and
3. n distinct *element items*, one for each element e, which has value $1 - (\varepsilon/2)$ for the buyers i_S such that set S contains element e, and value zero for the other buyers.

For a YES-instance of Max-Coverage, there is a solution with value $k + n$: we can assign both the choice and element items to the buyers corresponding to the k sets in the perfect partition, thereby getting us value $n + k$. (The excess for each choice item can subsidize the deficit for the n/k element items assigned to that buyer.) On the other hand, for a NO-instance, the k buyers/sets selected by the choice items can give value k and also subsidize at most $n(1 - 1/e + \delta)$ element items with deficit. (No other items with deficit can be chosen.) Setting $\delta = \varepsilon/2$ means the NO-instances have value at most $k + n(1 - 1/e + \delta) + n\varepsilon/2 \le n(1 - 1/e + \varepsilon)$. This gives a gap between instances with value at least n and at most $n(1 - 1/e + \varepsilon)$, proving the theorem. ◀

B.2 Clique hardness of GenAVA

Next, we prove that approximating GenAVA defined in (1.2) is as hard as approximating the maximum independent set number in a graph. Recall that the objective in GenAVA is to maximize welfare $\sum_{ij} v_{ij} x_{ij}$ subject to the more general return-on-spend (ROS) constraints:

$$\forall j, \quad \sum_i v_{ij}\, x_{ij} \ge \rho_j \cdot \left(\sum_i c_{ij}\, x_{ij} \right). \tag{B.17}$$

Without loss of generality, we scale c_{ij} and ensure that all $\rho_j = 1$. We show the hardness even for the case where costs depend only on the items, i.e., $c_{ij} = c_i$ for each item i. (The case where $c_{ij} = c_j$ for each buyer j is much easier – equivalent to the AVA problem – because we can just fold the c_j term into the ρ_j threshold.)

▶ **Theorem 33** (Hardness of GenAVA). *For any constant $\varepsilon > 0$, it is NP-hard to approximate GenAVA for n-buyer instances with $\Omega(n^2)$ items to better than a factor of $n^{1-\varepsilon}$.*

The proof uses a reduction from the Maximum Independent Set problem. The reduction proceeds as follows: given a graph $G = (V, E)$ with $|V| = n$, define $M := 2|E|/n^{\varepsilon}$, and construct the following GenAVA instance.

1. For each vertex $v \in V$, there is a buyer j_v with $\rho_{j_v} = 1$.
2. For each vertex $v \in V$, there is a *vertex item* i_v with item cost $c_i := M + \deg(v)$, where $\deg(v)$ is v's degree in G; it has value M for the buyer j_v, and zero value for all other buyers.
3. For each edge $e = (u, v) \in E$, there is an *edge item* i_e having zero cost; it has value 1 for buyers j_u and j_v, and zero value for all others.

Proof of Theorem 33. If vertex item i_v is allocated to buyer j_v, then by the constraints above, all edge items j_e with $e \ni v$ must be allocated to i_v. Thus, the set of vertices $U \subseteq V$ whose buyers are sold their respective vertex item is an independent set in G. Conversely, U can be taken to be any independent set. Thus, the maximum value obtained by allocating vertex items is precisely $M \cdot \alpha(G)$. On the other hand, any optimal allocation must allocate all edge items, as this does not violate any of the ROS constraints. Combining the above, we have that $OPT = \alpha(G) \cdot M + |E|$, where $\alpha(G)$ is the independence number of G, i.e., the size of the maximum independent set of G.

Finally, we use the result that for any constant $\varepsilon > 0$, it is NP-hard to distinguish between the following two scenarios for an n-node graph G: (a) G contains a clique on $n^{1-\varepsilon}$ nodes (YES instances), and (b) G contains no clique on $n^{\varepsilon}/2$ nodes (NO instances) [27, 36]. This means that it is NP-hard to distinguish between instances of GenAVA with value at least $n^{1-\varepsilon} \cdot M$ (corresponding to YES instances) from those with value at most $(n^{\varepsilon}/2) \cdot M + |E| = n^{\varepsilon} \cdot M$ corresponding to the NO instances, and hence proves the claim. ◀

The above hardness construction can, with small changes, show the following hardness results. We defer these additional results' proofs, as well as algorithms showing the (near) tightness of our lower bounds for general GenAVA, to the full version.

▶ **Theorem 34** (Hardness of i.i.d. GenAVA). *For any constant $\varepsilon > 0$, it is NP-hard to $n^{1-\varepsilon}$-approximate GenAVA in n-buyer instances with* $\text{poly}(n)$ *items drawn i.i.d. from a known distribution.*

▶ **Theorem 35** (Hardness of Bicriteria GenAVA). *For any $\varepsilon > 0$, it is NP-hard to obtain a solution (which can even be infeasible) to GenAVA that achieves an objective value at least $\tilde{\Omega}(\sqrt{\varepsilon})$ times the optimal value (i.e. an $\tilde{O}(1/\sqrt{\varepsilon})$-approximation), while guaranteeing the cost for each buyer is at most $1 + \varepsilon$ times their total value, assuming the UGC.*[5]

[5] As usual, the soft-Oh notation hides polylogarithmic factors in its argument: i.e., $\tilde{O}(f) = f \cdot \text{poly}\log(f)$.