The Distortion of Threshold Approval Matching

Mohamad Latifian\textsuperscript{1} and Alexandros A. Voudouris\textsuperscript{2}
\textsuperscript{1}University of Toronto, Canada
\textsuperscript{2}University of Essex, UK
latifian@cs.toronto.edu, alexandros.voudouris@essex.ac.uk

Abstract

We study matching settings in which a set of agents have private utilities over a set of items. Each agent reports a partition of the items into approval sets of different threshold utility levels. Given this limited information on input, the goal is to compute an assignment of the items to the agents (subject to cardinality constraints depending on the application) that (approximately) maximizes the social welfare (the total utility of the agents for their assigned items). We first consider the well-known, simple one-sided matching problem in which each of a set of agents is to be assigned exactly one item. We show tight bounds on distortion of deterministic and randomized matching algorithms that are functions of the number of threshold utility levels. We further show that our distortion bounds extend to a more general setting in which there are multiple copies of the items, each agent can be assigned a number of items (even copies of the same one) up to a capacity, and the utility of an agent for an item depends on the number of its copies that the agent is given.

1 Introduction

The assignment of papers to reviewers in conference management systems like CMT, HotCRP, EasyChair and OpenReview is computed using bidding information that classifies the papers into sets based on whether the reviewers are, for example, eager, willing, or not willing to handle them. In a sense, this process defines a collection of threshold levels that the reviewers (or, more generally, agents) can use to partition the papers (or, more generally, items) into associated approval sets based on their preferences (which can be dependent on their experience, their interests, and so on).

Eliciting only threshold approvals rather than more detailed information about the underlying utility preferences of the agents for the items inevitably leads to inefficiency in terms of natural, cardinal objectives such as the well-known social welfare (the total utility). Typically, the loss of efficiency of decision-making methods that have access only to incomplete information is captured by the notion of distortion, which is defined as the worst-case ratio of the maximum possible social welfare over that of the computed solution. The distortion was originally used for social choice settings (such as voting) where decisions are made only based on ordinal information (rankings) [Procaccia and Rosenschein, 2006; Boutilier et al., 2015], but has recently been studied for settings in which different types of information is available or can be elicited (e.g., see [Amanatidis et al., 2021; Amanatidis et al., 2022; Mandal et al., 2019; Mandal et al., 2020; Ma et al., 2021]).

In the matching setting we study in this work, which captures various interesting applications (such as the paper assignment problem in peer-reviewing that we briefly introduced above, as well as general constrained resource allocation), the threshold approvals reported by the agents is a type of information that lies in-between fully cardinal and fully ordinal. Hence, while we cannot hope to achieve full efficiency, we can hope to achieve distortion better than what is possible just with ordinal preferences, depending on how detailed the threshold approvals are. In particular, we are interested in the possible tradeoffs between the distortion and the number of threshold levels both for when allocations are computed deterministically (which is the most natural way of doing so in social choice problems), as well as when randomization can be exploited.

1.1 Our Contribution

We start by considering the fundamental one-sided matching problem (also known as house allocation) to introduce the main ideas of our techniques, before turning to a more general setting. In one-sided matching, there is a set of \( n \) agents with utilities for a set of \( n \) items; we assume that the utilities satisfy the standard unit-sum assumption [Aziz, 2019].

The utilities are private and are not explicitly reported by the agents. Instead, for a number \( t \) of decreasing threshold values, each agent reports a collection of \( t \) approval sets consisting of items of different utility level; in particular, each approval set is associated with a threshold value and includes all items for which the agent has utility that is at least this threshold.

Given the approval sets as input, our goal is to determine a one-to-one matching between agents and items to approximate the social welfare (total utility of the agents for their assigned items) is maximized.

We show tight bounds on the best possible distortion achieved by matching mechanisms for any number \( t \) of...
thresholds. In particular, we show a bound of $\Theta(\sqrt{n})$ for deterministic mechanisms, and a bound of $\Theta(\sqrt[4]{n})$ for randomized ones. The lower bounds are presented in Section 3 and the upper bounds in Section 4. To put the bounds into perspective, we note that just one threshold is sufficient to obtain distortion $\Theta(n)$ for deterministic algorithms, beating the best possible distortion of $\Theta(n^2)$ that can be achieved using ordinal information [Amanatidis et al., 2022]. Similarly, a distortion of $\Theta(\sqrt{n})$ can be achieved with randomization, matching the best possible distortion achieved by ordinal randomized algorithms [Filos-Ratsikas et al., 2014].

In Section 5, we turn our attention to a more general setting where the agents have capacities, indicating the maximum number of items they can receive, and the items have supplies, indicating the number of copies of them that are available. We make a budget-balance assumption that the total capacity is asymptotically of the same order of the total supply; for example, in the paper assignment problem, all papers must receive a number of reviews, and so the total capacity must be sufficiently larger than the total supply. We also assume that, when agents can receive multiple copies of an item, their utility depends on the number of these copies, and thus the copies are not treated as independent items. Our goal is to compute an allocation of items to agents, such that the capacity and the supply constraints are satisfied, and the social welfare is maximized.

As this setting is a generalization of the one-sided matching (in which the number of agents is equal to the number of items, and there are unit capacities and supplies), our lower bounds from Section 3 extend directly. For the upper bounds, we show that, for $t$ thresholds, the best possible distortion achieved by deterministic mechanisms is $O(c \cdot \sqrt{T})$, where the best distortion achieved by randomized mechanisms is $O(c \cdot \sqrt{T})$, where $T$ is the total available supply (or capacity) and $c$ is a parameter that depends either on the maximum capacity or the ratio between the number of items and agents. From this, we get bounds $\Theta(\sqrt{n})$ and $\Theta(\sqrt[4]{n})$ when the capacities and the supplies are constant.

1.2 Related Work

The distortion was originally defined by Procaccia and Rosenschein [2006] to measure the worst-case loss in social welfare when voting decisions are made using only ordinal information. Since then, the distortion has been studied for several different voting problems, including utilitarian voting [Procaccia and Rosenschein, 2006; Boutilier et al., 2015; Caragiannis and Procaccia, 2011; Caragiannis et al., 2017; Ebadian et al., 2022], metric voting [Anshelevich et al., 2018; Caragiannis et al., 2022; Charikar and Ramakrishnan, 2022; Charikar et al., 2024; Gkatzelis et al., 2020; Kizilkaya and Kempe, 2022], and combinations of the two [Gkatzelis et al., 2023]. It has also been studied for social choice problems beyond voting, such as one-sided matching that we also consider in this paper [Filos-Ratsikas et al., 2014; Amanatidis et al., 2022; Amanatidis et al., 2024], as well as other clustering and graph problems [Abramowitz and Anshelevich, 2018; Anshelevich and Sekar, 2016; Burkhardt et al., 2024]. See the survey of Anshelevich et al. [2021] for an introduction to the distortion framework.

Our paper follows a relatively recent stream of papers within the distortion literature that have considered elicitation methods beyond ordinal information. In this direction, Mandal et al.; Mandal et al. [2019; 2020] showed tradeoffs between the best possible distortion and a communication complexity measure (the number of bits the agents can use to report information) for utilitarian voting. Results of similar flavour for metric voting have also been shown, for example, by Kempe [2020].

More related to our work, in a series of papers, Amanatidis et al.; Amanatidis et al.; Amanatidis et al. [2021; 2022; 2024] studied voting and matching settings in which the agents provide ordinal information and, on top of that, are capable of answering value queries about their utilities for specific alternatives. They showed lower and upper bounds on the distortion of deterministic mechanisms that are functions of the number of queries per agent that are of similar to ours in the sense that the distortion decreases with the number of queries; some of their lower bounds (related to the number of queries required to achieve constant distortion) were recently improved by Caragiannis and Fehr [2023]. Another related paper is that of Ma et al. [2021] who considered the one-sided matching problem when the agents can answer binary threshold queries about whether their utility for specific alternatives is larger than appropriately chosen thresholds. Using an approach similar to that of Amanatidis et al., they showed bounds on the distortion of deterministic mechanisms that is a function of the number of queries in terms of the social welfare among matchings that satisfy properties such as Pareto or rank-maximality. Ignoring differences in the models, the elicitation methods in these papers are related to the one we consider since the threshold approval sets can be computed using a number of (value or binary threshold) queries. Hence, our elicitation method is in a sense a bit more demanding. However, for the setting we focus here, we are able to show asymptotically tight bounds not only for deterministic mechanisms, but also for randomized ones, which have not been studied before.

Threshold approvals have also been recently explored in various other works on the distortion of voting mechanisms, most notably by Ebadian et al. [2023] who showed that a single, appropriately chosen threshold is sufficient to achieve a distortion of $O(\sqrt{m})$ in utilitarian single-winner voting with $m$ alternatives. In metric voting, Anshelevich et al. [2024] showed improved distortion bounds using an approval set per agent computed by a threshold value that is relative (rather than absolute) to the distance from the top-ranked alternative. Threshold approvals have also been considered in the context of participatory budgeting by Benadé et al. [2021], and voting under truthfulness constraints by Bhaskar et al. [2018]. All these works use just a single threshold, whereas we here explore the full potential of this elicitation method (for matching problems, rather than voting) using multiple thresholds and show tight bounds on the possible distortion.

2 The One-Sided Matching Problem

We start with the simple one-sided matching setting to express the core idea; in Section 5, we show that our results
extend to a more general setting that more accurately captures applications such as paper assignment in peer reviewing. Let $\mathcal{N}$ be a set of $n$ agents and $\mathcal{M}$ be a set of $n$ items. Agent $i$ has a utility function $u_i : \mathcal{M} \rightarrow [0, 1]$ over the items. We assume that these utility functions satisfy the unit-sum assumption, which means for each agent $i \in \mathcal{N}$, $\sum_{j \in \mathcal{M}} u_i(j) = 1$. Together, these utility functions form the utility profile $\bar{u}$. A matching of the items to the agents is a bijection $A : \mathcal{N} \rightarrow \mathcal{M}$. With a slight abuse of notation, we use $A_i = A(i)$ to refer to the item matched to agent $i$, and also $A(a)$ to refer to the agent matched to item $a$. We define the social welfare of a matching $A$ under utility profile $\bar{u}$ to be the total utility of the agents for the items they are matched to, i.e.,

$$sw(A, \bar{u}) = \sum_{i \in \mathcal{N}} u_i(A_i) = \sum_{a \in \mathcal{M}} u_{A(a)}(a).$$

The goal is to compute a matching with high social welfare in the worst case. For ease of notation, we will drop $\bar{u}$ from $sw(A, \bar{u})$ whenever it is clear from context.

Elicitation Method. In this paper, we focus on eliciting threshold approval votes. A threshold vector $\vec{\tau} = (\tau_1, \ldots, \tau_t)$, we ask each agent $i$ to submit $t$ disjoint threshold approval subsets of $\mathcal{M}$, denoted by $S_{i,1}, \ldots, S_{i,t}$, where $S_{i,k}$ includes the items for which the agent has utility in $[\tau_{k-1}, \tau_k)$, with $\tau_0 := 1$. In other words, $S_{i,k} = \{j \in \mathcal{M} : \tau_{k-1} \geq u_i(j) > \tau_k\}$. All these $n \times t$ threshold approval sets form the input profile $\mathbf{S}$. Note that different utility profiles might induce the same input profile. We say that a utility profile $\bar{u}$ is consistent with an input profile $\mathbf{S}$ (in which say we write $\bar{u} \triangleright \mathbf{S}$) if for each agent $i \in \mathcal{N}$, $k \in [t]$, and $j \in S_{i,k}$, $\tau_{k-1} \geq u_i(j) > \tau_k$.

Mechanisms and Distortion. A mechanism $f$ defines a threshold vector $\vec{\tau}$, takes an input profile $\mathbf{S}$ based on $\vec{\tau}$, and then outputs a matching $f(\mathbf{S})$ of the items to the agents. The distortion of a matching $A$ on input profile $\mathbf{S}$ is defined as:

$$\text{dist}(A, \mathbf{S}) = \sup_{\bar{u} \triangleright \mathbf{S}} \frac{sw(A^*, \bar{u})}{sw(A, \bar{u})},$$

where $A^*$ is the matching with the maximum social welfare with respect to $\bar{u}$. The distortion of a matching mechanism $f$ is defined as the worst case distortion of $f$ on any input profile:

$$\text{dist}(f) = \sup_{\mathbf{S}} \text{dist}(f(\mathbf{S}), \mathbf{S}).$$

3 Lower Bounds

In this section we show lower bounds on the best possible distortion achievable by deterministic and randomized mechanisms for the one-sided matching problem. In particular, for mechanisms that use $t \geq 1$ thresholds, we show a lower bound of $\Omega(\sqrt{n})$ for deterministic mechanisms and a lower bound of $\Omega\left(\frac{\sqrt{n}}{\log n}\right)$ for randomized mechanisms. We start by showing a technical lemma that holds for randomized mechanisms that will be useful in establishing the lower bounds in several cases. For a randomized mechanism $f$, denote by $p(i, a)$ the probability that item $a$ is assigned to agent $i$ according to $f$. Due to lack of space, the proof of the following statement, as well as that of some other ones, can be found in the technical appendix.

Lemma 1. For any subset of items $\mathcal{M} \subseteq \mathcal{M}$, let $A_M$ be the matching of the items in $\mathcal{M}$ to the agents with minimum sum of probabilities with respect to $f$. Then, $\sum_{a \in \mathcal{M}} p(A_M(a), a) \leq 1$.

We are now ready to show the lower bounds via a sequence of lemmas capturing different cases. The first lower bound depends on the ratio of consecutive threshold levels and holds for any mechanism (randomized or deterministic).

Lemma 2. Consider a threshold vector $\vec{\tau} = (\tau_1, \ldots, \tau_t)$, and let $k \in [t]$ be such that $\delta = \tau_{k-1}/\tau_k$ is the largest multiplicative gap between two consecutive thresholds (assuming $\tau_0 = 1$). Then, the distortion of any matching mechanism $f$ that uses $\vec{\tau}$ is $\Omega(\delta)$.

Our next two lemmas provide lower bounds for deterministic and randomized mechanisms, respectively, for when the last threshold level is sufficiently small.

Lemma 3. Consider a threshold vector $\vec{\tau} = (\tau_1, \ldots, \tau_t)$ such that $\tau_t \geq 1/(n-1)$. Then, the distortion of any deterministic matching mechanism $f$ that uses $\vec{\tau}$ is unbounded.

Proof. Consider the input profile $\mathbf{S}$ where the threshold approval sets of any agent are empty, and thus the utility of any agent for any item is at most $\tau_t$. Let $A = f(\mathbf{S})$ be the matching computed by the deterministic matching mechanism $f$, and let $B$ be another matching such that $A(a) \neq B(a)$ for every item $a \in \mathcal{M}$. Consider the utility profile $\vec{u}$ where agents have utility 0 for their matched item in $A$, utility $\tau_t$ for their matched item in $B$, and $(1 - \tau_t)/(n - 2)$ for each of the remaining $n - 2$ items. Note that $\tau_t \geq 1/(n-1) \implies (1 - \tau_t)/(n - 2) \leq \tau_t$, and hence $\vec{u} \triangleright \mathbf{S}$. Since $sw(A, \vec{u}) = 0$ and $sw(B, \vec{u}) = n \cdot \tau_t > 0$, the distortion is unbounded. □

Lemma 4. Consider a threshold vector $\vec{\tau} = (\tau_1, \ldots, \tau_t)$ such that $\tau_t > 1/n$. Then, the distortion of any randomized matching mechanism $f$ that uses $\vec{\tau}$ is $\Omega(n \cdot \tau_t)$.

Proof. Consider the input profile $\mathbf{S}$ where the threshold approval sets of any agent are empty, and thus the utility of any agent for any item is at most $\tau_t$. Let $A_M$ be the matching over $\mathcal{M}$ with minimum sum of probabilities; by Lemma 1, $\sum_{a \in \mathcal{M}} p(A_M(a), a) \leq 1$. Now, consider the utility profile $\vec{u}$ where each agent has utility $\tau_t$ for the item she is matched to according to $A_M$ and utility $(1 - \tau_t)/(n-1)$ for each of the remaining items. Note that $\tau_t \geq 1/n \implies (1 - \tau_t)/(n-1) \leq \tau_t$, and hence $\vec{u} \triangleright \mathbf{S}$. The expected social welfare of the mechanism is

$$E_{A \sim f(\mathbf{S})} \left[ \sum_{a \in \mathcal{M}} u_{A_M(a)}(a) \right] = \sum_{a \in \mathcal{M}} \left( p(A_M(a), a) \cdot \tau_t + (1 - p(A_M(a), a)) \cdot \frac{1 - \tau_t}{n - 1} \right) \leq \tau_t - \frac{1 - \tau_t}{n - 1} + n \cdot \frac{(1 - \tau_t)}{n - 1} = 1.$$
Since $sw(A_M, \bar{u}) = t \cdot \tau_i$, the distortion is $\Omega(n \cdot \tau_i)$.

By appropriately combining Lemmas 2, 3, and 4, we can establish the desired lower bounds on the distortion of the different types of mechanisms.

**Theorem 1.** The distortion of any deterministic matching mechanism $f$ that uses a threshold vector $\bar{\tau}$ of length $t$ is $\Omega(\sqrt{n})$.

**Proof.** If $\tau_i \geq 1/(n-1)$, by Lemma 3, the distortion is bounded. Otherwise, if $\tau_i \leq 1/(n-1)$, let $k \in \arg \max_{j \in [t]} \tau_j \geq 1. Clearly,

$$
\left(\frac{\tau_k - 1}{\tau_k}\right) \geq \prod_{j \in [t]} \frac{\tau_j - 1}{\tau_j} = \frac{1}{\tau_i}
$$

Thus, $\delta = \frac{\tau_k - 1}{\tau_k} \geq \tau_i^{-1/t} \geq \sqrt{n}$, and thus, by Lemma 2, the distortion is $\Omega(\sqrt{n})$.

**Theorem 2.** The distortion of any matching mechanism $f$ that uses a threshold vector $\bar{\tau}$ of length $t$ is $\Omega(\sqrt[4]{n})$.

**Proof.** Suppose that the threshold vector $\bar{\tau}$ is such that $\tau_i > n^{-t/(t+1)}$. Since $n^{-t/(t+1)} \geq n^{-1}$, by Lemma 4, the distortion of $f$ is $\Omega(n \cdot \tau_i) = \Omega(\sqrt[4]{n})$. So, we can now assume that $\tau_i \leq n^{-t/(t+1)}$. As in the proof of Theorem 1, we have that $\delta = \tau_k - 1/\tau_k \geq \tau_i^{-1/t} \geq \sqrt[4]{n}$, and thus, by Lemma 2, the distortion of $f$ is $\Omega(\delta) = \Omega(\sqrt[4]{n})$.

**4 Upper Bounds**

In this section we present asymptotically tight upper bounds for deterministic and randomized matching mechanisms. Our deterministic mechanism (described below) computes a maximum-weight matching by assuming that each agent has the minimum possible utility (according to the thresholds) for all the items in the different approval sets given as input.

**Definition 1.** For $\delta > 1$ and $t \in [n]$, consider the threshold vector $\bar{\tau} = (\delta^{-1}, \delta^{-2}, \ldots, \delta^{-t})$. The deterministic matching mechanism $f_\delta$ uses the threshold vector $\bar{\tau}$ and, given an input profile $S$, constructs the following weighted bipartite graph $G_S$: There are $2n$ nodes in total, consisting of a node $v_i$ for each agent $i \in \mathcal{N}$ on the left side and a node $z_a$ for each item $a \in M$ on the right side. For $i \in \mathcal{N}$, $k \in [t]$ and $a \in S_i$, there is an edge from $v_i$ to $z_a$ with weight $w(v_i, z_a) = \tau_k$. The mechanism $f_\delta$ finds the maximum weighted matching in $G_S$ and, for each matched pair $(v_i, z_a)$, assigns item $a$ to agent $i$. If there are unmatched pairs remaining, $f_\delta$ completes the allocation arbitrarily.

**Example 1.** Let $t = 2$ and $\bar{\tau} = (\tau_1, \tau_2)$. Suppose that $S_{1,1} = \{a, c\}$, $S_{2,1} = \{d\}$, $S_{2,2} = \{c\}$, $S_{3,2} = \{a, c, d\}$, while the remaining approval sets are empty. Mechanism $f_\delta$ constructs the graph $G_S$ shown in Figure 1, computes a maximum-weight matching, and then assigns any unmatched items arbitrarily.

![Figure 1: The graph $G_S$ that is used by $f_\delta$ in Example 1.](image)

Before we bound the distortion of the mechanism, we prove two very useful technical lemmas. The first one provides us with a lower bound on the weight of the maximum-weight matching in a bipartite graph whose nodes satisfy certain properties; this will be used extensively to lower bound the social welfare of the matching computed by $f_\delta$, and also by the deterministic mechanism in Section 5.

**Lemma 5.** Consider a weighted bipartite graph $G$ with $n$ nodes on the left side $\{v_1, \ldots, v_n\}$ and $m$ nodes $\{z_1, \ldots, z_m\}$ on the right side. If $\sum_{a \in [m]} w(v_i, z_a) \geq W$ for each $v_i$, and $w(v_i, z_a) \geq L$ for each edge $(v_i, z_a)$, then there is a matching in $G$ with weight at least $\min\{W, nL\}$.

The second lemma shows that $G_S$ admits a matching with weight that is relatively close to the social welfare of the optimal matching allocation for any utility profile consistent to the input profile $S$. We will use this relation in the analysis of the distortion of $f_\delta$, as well as in the analysis of our randomized matching mechanism later on.

**Lemma 6.** Let $S^*$ be the social welfare of the optimal matching allocation. There exists a matching of $G_S$ with weight at least $\delta^{-1}(s^* - n \tau_i)$.

**Proof.** For any agent $i$, let $a^*_i$ be the item that $i$ is given in the optimal matching allocation. Clearly, either there exists $j \in [t]$ such that $a^*_i \in S_{i,j}$, or $a^*_i \not\in \bigcup_{j \in [t]} S_{i,j}$. The total utility accumulated by the agents of the second type is at most $n \tau_i$. For the agents of the first type, since $a^*_i \in S_{i,j}$ there is an edge between $v_i$ and $z_{a^*_i}$ in $G_S$ of weight

$$
\tau_j = \delta^{-1} \cdot \delta^{-j+1} = \delta^{-1} \cdot \delta^{-j-1} \geq \delta^{-1} \cdot u_i(a^*_i).
$$

Hence, the intersection of $A^*$ and $G_S$ gives us a matching of weight at least $\delta^{-1}(s^* - n \tau_i)$.

**Theorem 3.** For $t \in [n]$ and $\delta = \sqrt[4]{2n}$, the distortion of the deterministic matching mechanism $f_\delta$ is $O(\sqrt[4]{n})$.

**Proof.** Let $A = f_\delta(S)$ be the matching computed by the mechanism $f_\delta$ when given as input an arbitrary input profile $S$ that is induced by some consistent utility profile $\bar{u}$. First, observe that if $G_S$ admits a matching of weight $W$ then $sw(A, \bar{u}) \geq W$. This follows by the fact that if node $v_i$ is matched to node $z_a$ in $G_S$, then agent $i$ has utility at least $w(v_i, z_a)$ for item $a$.

Second, we argue that the weight of the maximum weight matching of $G_S$ is at least $\delta^{-1}/2$. The total utility of an agent for the items in $\bigcup_{j \in [t]} S_{i,j}$ is at least $1/2$ since the utility for
each of the remaining items is at most $\tau / n!)$, each agent is matched to each item with probability at least $1/n$. Since the sum of the utilities of each agent for all items is 1, the expected social welfare from the first part is at least
\[
\frac{1}{2} \sum_{i \in N} \sum_{a \in A} \frac{1}{n} u_i(a) = \frac{1}{2} \cdot \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{2}.
\]

For the second part of the mechanism (where the deterministic mechanism $f_i$ using $\vec{v}$ is employed), since $s := n\tau_i = n\delta - \delta^* = \frac{n}{\sqrt{n}}$, by Lemma 6, there is a matching in $G_S$ of weight at least $\delta^{-1}(s^* - s)$, and thus the expected social welfare of the mechanism from the second part is at least $\frac{1}{2} \cdot \delta^{-1}(s^* - s)$, Overall, we have established that
\[
\mathbb{E}_{A \sim R_s(S)}[sw(A)] \geq \frac{1}{2} \cdot \frac{1}{2} \cdot \delta^{-1}(s^* - s) \geq \frac{1}{2} \cdot \max \left\{ 1, \delta^{-1}(s^* - s) \right\},
\]
and thus the distortion is at most
\[
\frac{2 \cdot s^*}{\max\{1, \delta^{-1}(s^* - s)\}}.
\]

If $s^* \geq 2s$, then $s^* - s \geq s^*/2$ and the distortion is at most $\frac{2s^*}{\sqrt{s^*/2}} = 2\delta = 4^{1/\sqrt{n}}$. Otherwise, if $s^* < 2s$, the distortion is at most $2s^* \leq 4s = 4^{1/\sqrt{n}}$. In any case, the distortion is $O(1/\sqrt{n})$.

5 Generalized Setting

In this section we consider a generalized setting. Similarly to before, $N$ represents a set of $n \geq 1$ agents. However, here it is not necessarily the case that we have an equal number of items; we define $M$ to be a set of $m \geq 1$ items. Each item $a \in M$ has a supply $m_a \geq 1$, and each agent $i \in N$ has a capacity $c_i \geq 1$. For simplicity, we assume that the total supply is equal to the total capacity, that is, $T := \sum_{i \in N} c_i = \sum_{a \in M} m_a$.\footnote{Our results hold even when $\sum_{i \in N} c_i = \Theta(\sum_{a \in M} m_a)$, in which case we would need to define $T$ as the maximum between these two quantities.}

Agents are allowed to receive copies of the same item, in which case their utility depends on the number of copies they receive; in other words, copies of an item are not considered independent. For each agent $i$, item $a$ and $j \in \max\{c_i, m_a\}$, we denote by $u_{i}^{+}(a, j)$ the marginal utility that agent $i$ gets when receiving his $j$-th copy of item $a \in M$, and by $u_{i}^{-}(a, j)$ his total utility when receiving $j \leq \min\{c_i, m_a\}$ copies of item $a \in M$, i.e., $u_{i}^{+}(a, j) = \sum_{k \in [j]} u_{i}(a, j)$. An allocation $X = \{x_i(a)\}_{i \in N, a \in M}$ determines the number $x_i(a)$ of copies of item $a$ that agent $i$ is assigned to, such that $\sum_{a \in M} x_i(a) \leq c_i$ for every $i \in N$ and $\sum_{a \in M} x_i(a) \leq m_a$ for every $a \in M$. Given an allocation $X$, the utility of $i$ for $X$ is $u_{i}(X) = \sum_{a \in M} u_{i}^{+}(a, x_i(a))$. We assume that the utility function of each agent $i$ satisfies the unit-sum assumption, that is, $\sum_{a \in M} u_{i}^{+}(a, \min\{c_i, m_a\}) = 1$.\footnote{Our results hold even if we replace the unit-sum assumption by the assumption that the total utilities of the agents (as if each of them is given all items he can hold) are not equal, but known.}
definition of the social welfare of an allocation is the same as before, that is, it is the total utility of the agents for the allocation. We aim to compute allocations with high social welfare that maximally assign the items to the agents; observe that the maximum possible number of items that can be allocated is \( T \) (or \( \min \{ \sum_{i \in \mathcal{N}} c_i, \sum_{e \in \mathcal{M}} m_e \} \) in the more general case), and that any allocation \( Y \) that assigns less than \( T \) items is dominated in terms of social welfare by any allocation \( X \) that assigns the items allocated by \( Y \) in the same way, but also somehow assigns the remaining items. So, in the following, we focus on such maximal allocations only.

Since the utility functions depend on the number of item copies that the agents receive, we need to appropriately redefine the elicitation method. For a threshold vector \( \tau = (\tau_1, \ldots, \tau_t) \), each agent \( i \) reports \( t \) disjoint threshold approval sets \( S_{i,1}, \ldots, S_{i,t} \), where \( S_{i,k} \) includes pairs of items and indices for which \( i \) has marginal utility in \( [\tau_{k-1}, \tau_k) \), where \( \tau_0 := 1 \). In other words, \( S_{i,k} = \{ (a \in \mathcal{M}, j \in \min([c_i, m_a]) : \tau_{k-1} \leq u_i(a,j) > \tau_k \} \). The input profile \( \mathcal{S} \) now consists of these threshold approval sets reported by all agents. The definition of distortion can also be appropriately refined by taking the worst case over utility profiles and input profiles consistent with them.

We now define a parametric min-cost flow instance that will be used by our algorithms later.

**Definition 3 (Min-cost Flow Instance).** Let \( C = (C_i)_{i \in \mathcal{N}} \) be a capacity vector of the \( n \) agents with maximum value \( c_i \), \( M = (M_a)_{a \in \mathcal{M}} \) be a supply vector of the \( m \) items, and \( V_{n \times m \times c} \) a value matrix in which \( V(i,a,j) \) is the value of agent \( i \) when receiving \( j \) copies of item \( a \). Let \( G(C, M, V) \) be the following min-cost flow instance: \( G \) has a source node \( s \) and a destination node \( t \). Furthermore, there is a node \( v_{ij} \) for each agent \( i \) and a node \( z_a \) for each item \( a \). For every \( i \), there is an edge \((s, v_{ij})\) with capacity \( C_i \) and cost 0. For every \( a \), there is an edge \((z_a, t)\) with capacity \( M_a \) and cost 0. For each agent \( i \) and item \( a \), we add a component to the graph as shown in Figure 2. Nodes \( s \) and \( t \) have supply and demand equal to \( \min(\sum_{i \in [n]} C_i, \sum_{a \in [m]} M_a) \), respectively. The goal is to find the minimum cost for satisfying this flow.

It is well-known that the minimum-cost flow problem can be solved in polynomial time via linear programming (and also using various other algorithms), and we thus have the following property.

**Lemma 7.** For capacity vector \( C \), supply vector \( M \) and value matrix \( V \), the min-cost flow instance defined in Definition 3 has an integral solution which we can find in polynomial time.

Our next lemma provides a connection between the solution of the min-cost flow instance of Definition 3 and the social welfare of the corresponding allocation for the instance of our problem.

**Lemma 8.** The absolute value of the minimum-cost flow in \( G(C, M, V) \) is equal to the maximum social welfare of an allocation of items to agents with respect to values in \( V \).

**Proof Sketch.** Let \( X^* \) be the allocation with the maximum social welfare w.r.t. \( V \), and \( F \) be the min-cost flow in \( G(C, M, V) \). We will show that \( \text{sw}(X^*, V) = -\text{Cost}(F) \) by bounding \( \text{Cost}(F) \) from above and below by \( -\text{sw}(X^*, V) \). To do so we explain how to construct a flow solution from an allocation and vice versa. Then, we compute the cost of the constructed solution and finally show the bound. \( \square \)

We are now ready to present our deterministic mechanism \( g_t \), which is a generalization of the deterministic mechanism \( f_t \) that we used for the one-sided matching setting.

**Definition 4.** For \( \delta > 1 \) and \( t \in [n] \), consider the threshold vector \( \tilde{\tau} = (\delta^{-1}, \delta^{-2}, \ldots, \delta^{-t}) \). The deterministic generalized matching mechanism \( g_t \) uses the threshold vector \( \tilde{\tau} \) and gets as input a profile \( \mathcal{S} \), constant agent capacities \( \{c_1, \ldots, c_N\} \), and constant item supplies \( \{m_1, \ldots, m_m\} \). The mechanism defines the vector \( C = \langle c_1, \ldots, c_n \rangle \), the vector \( M = \langle m_1, \ldots, m_m \rangle \), and the matrix \( V \) as follows: For every \( i \in \mathcal{N} \) and \( j \in [\min(c_i, m_a)] \), if \( (a, j) \in S_{i,k} \) for some \( k \in [t] \), then the mechanism defines \( V_{i,a,j} = \tau_k \); otherwise, if \( (a, j) \not\in \bigcup_{k=1}^t S_{i,k} \) then it defines \( V_{i,a,j} = 0 \). The mechanism computes the solution of the min-cost flow instance defined in Definition 3 with input \( C, M \) and \( V \). For each agent \( i \in \mathcal{N} \) and item \( a \in \mathcal{M} \) the flow from \( v_{i,a} \) to \( v_{i,a,1} \) in the computed solution is the number of copies of item \( a \) that agent \( i \) receives.

Before we bound the distortion of the mechanism, we prove a technical lemma similar to Lemma 6 which provides us with a lower bound on the social welfare of the allocation computed by the mechanism in relation to the optimal social welfare.

**Lemma 9.** If there is an allocation with social welfare \( s^* \), then \( g_t \) outputs an allocation with social welfare at least \( \delta^{-\frac{1}{2}}(s^* - T \cdot \tau_t) \).

**Proof.** Let \( X^* \) be an optimal allocation. We have

\[
\text{sw}(X^*) \leq \sum_{i \in \mathcal{N}} \sum_{k \in [t]} \sum_{(a,j) \in S_{i,k}} u_i(a,j) + T \cdot \tau_t.
\]

Recall that, if \( (a,j) \in S_{i,k} \) for some \( k \), then \( u_i(a,j) \leq \delta \cdot V_{i,a,j} \). This implies

\[
\text{sw}(X^*) \leq \delta \cdot \sum_{i \in \mathcal{N}} \sum_{k \in [t]} \sum_{(a,j) \in S_{i,k}} V_{i,a,j} + T \cdot \tau_t.
\]

Consequently, with respect to \( V \), \( X^* \) has a social welfare of at least \( \delta^{-\frac{1}{2}}(\text{sw}(X^*) - T \cdot \tau_t) \). This is a lower bound on the social welfare of the allocation computed by \( g_t \), since, by Lemma 8, this is at least the social welfare of the allocation with maximum social welfare with respect to \( V \). \( \square \)

We are now ready to show the upper bound on the distortion of \( g_t \).

**Theorem 5.** For \( t \in [T] \) and \( \delta = \sqrt{2T} \), the distortion of the deterministic generalized matching mechanism \( g_t \) is \( O(c \cdot \sqrt{T}) \), where \( c = \max_{i \in \mathcal{N}} \{c_i\} \).

**Proof Sketch.** The structure of the proof is very similar to that of Theorem 3. Let \( X \) be the allocation computed by \( g_t \) when given as input an arbitrary input profile. For any \( i \in \mathcal{N} \),
When the capacities and supplies are constant, Corollary 2. The distortion is at most $2\sqrt{\frac{s}{n}}$ since the total capacity is of the same magnitude as the total supply. Hence, the parameter $\delta$ is at most $2\sqrt{\frac{s}{n}}$, where $s$ is the optimal social welfare, by appropriately using Lemma 9. So, the distortion is at most $\frac{2\delta \cdot s^*}{\max \{1/c, 2s^* - 1\}}$. If $s^* \geq 1$, then $2s^* - 1 \geq s^*$, and hence the distortion is at most $2\delta \in O(\sqrt{T})$. Otherwise, if $s^* < 1$, then the distortion is at most $2\delta c \in O(c \cdot \sqrt{T})$.

In many applications, the capacities and the supplies are constants. In this case, we have that $T = \Theta(n) = \Theta(m)$ and the following result, which is tight given the corresponding lower bound in Section 3.

**Corollary 3.** When the capacities and supplies are constant, for any $t \in [T]$, there is a randomized mechanism with distortion $O(1/\sqrt{n})$.

**Remark 1.** In the generalized setting that we considered in this section, the agents are allowed to receive potentially all available copies of the items, up to their capacity. However, in several applications, we might want to disallow this and set a limit $\ell_{i,a}$ on the number of copies of $a \in M$ that agent $i \in N$ can get. For example, in the paper assignment problem, each agent must be given at most one copy of each item since it does not make sense for someone to review a paper more than once. Such constraints can be handled in several ways. One of them is via the utility functions of the agents which, for scenarios like these, would simply assign a marginal value of 0 for any extra copy that exceeds the limit, that is, $u_i(a, j) = 0$ for every $j > \ell_{i,a}$. If the utility function is not naturally defined this way, we can modify the min-cost flow instance by setting $V_{i,a,j} = -\infty$ for $j > \ell_{i,a}$, or by removing the corresponding edges in the graph.

### 6 Conclusion and Open Problems

In this paper, we showed tight bounds on the best possible distortion of (both deterministic and randomized) mechanisms for matching settings (that capture important applications, including the one-sided matching problem and the paper assignment problem) when the elicited information about the preferences of the agents is of the form of threshold approvals. Going forward, it would be interesting to explore whether improved tradeoffs can be achieved by using randomization not only for the decision phase but also for the definition of the threshold values, similarly to the works of Benadè et al.; Bhaskar et al. [2021; 2018]. In addition, it makes sense to consider the metric version of the problem which captures the case where the items represent chores. Finally, one could explore other settings in which the same type of elicitation method can be applied, including voting settings (both utilitarian and metric) in which the full potential of using multiple threshold approvals has not been considered before, as well as other resource allocation settings, potentially also in combination with other constraints, such as truthfulness or fairness.
References


