

# Maxmin-Fair Ranking: Individual Fairness under Group-Fairness Constraints

David García-Soriano  
ISI Foundation, Turin, Italy  
d.garcia.soriano@isi.it

Francesco Bonchi  
ISI Foundation, Turin, Italy  
Eurecat, Barcelona, Spain  
francesco.bonchi@isi.it

## ABSTRACT

We study a novel problem of fairness in ranking aimed at minimizing the amount of individual unfairness introduced when enforcing group-fairness constraints. Our proposal is rooted in the *distributional maxmin fairness* theory, which uses randomization to maximize the expected satisfaction of the worst-off individuals. We devise an exact polynomial-time algorithm to find maxmin-fair distributions of general search problems (including, but not limited to, ranking), and show that our algorithm can produce rankings which, while satisfying the given group-fairness constraints, ensure that the maximum possible value is brought to individuals.

## CCS CONCEPTS

• Computing methodologies → Machine learning.

## KEYWORDS

fairness, ranking, max-min fairness

## ACM Reference Format:

David García-Soriano and Francesco Bonchi. 2021. Maxmin-Fair Ranking: Individual Fairness under Group-Fairness Constraints. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '21), August 14–18, 2021, Virtual Event, Singapore*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3447548.3467349>

## 1 INTRODUCTION

As the position in a ranking influences to a great extent the amount of attention that an item receives, biases in ranking can lead to unfair distribution of exposure, thus producing substantial economic impact. If this is important when ranking items (e.g., web pages, movies, hotels, books), it raises even more crucial concerns when ranking people. In fact, ranking is at the core of many decision-making processes in spheres such as health (e.g., triage in pandemic), education (e.g., university admission), or employment (e.g., selection for a job), which can have a direct tangible impact on people's life. These concerns have captured the attention of researchers, which have thus started devising ranking systems which are *fair* for the items being ranked [3, 7, 13, 26, 30, 31, 34].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

KDD '21, August 14–18, 2021, Virtual Event, Singapore

© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-8332-5/21/08...\$15.00

<https://doi.org/10.1145/3447548.3467349>

The bulk of the algorithmic fairness literature deals with *group fairness* along the lines of *demographic parity* [9] or *equal opportunity* [16]: this is typically expressed by means of some fairness constraint requiring that the top- $k$  positions (for any  $k$ ) in the ranking contain enough elements from some groups that are protected from discrimination based on sex, race, age, etc. In fact, [6] shows that in a certain model, *group-fairness constraints can eliminate the bias implicit in the ranking scores*. Besides, some legal norms enforce these constraints [1, 2]. For these reasons we will consider a ranking *valid* if it satisfies a given set of group-fairness constraints of this type, as detailed in Section 3.

More formally, consider a set of elements (items or individuals) to be ranked  $\mathcal{U} = \{u_1, \dots, u_n\}$ , a partition of  $\mathcal{U}$  into groups defined by some protected attributes, and a relevance score  $R : \mathcal{U} \rightarrow \mathbb{R}^{\geq 0}$  for each element. For instance,  $\mathcal{U}$  could be the result of a query while  $R$  represents the relevance of each item for the query, or  $\mathcal{U}$  could be the set of applicants for a job while  $R$  their fitness for the job. Let  $\mathcal{R}$  denote all possible rankings of  $\mathcal{U}$  (bijections from  $\mathcal{U}$  to  $[n]$ ), where  $r(u) \in [n]$  denotes the position of element  $u$  in a ranking  $r \in \mathcal{R}$  and let  $\mathcal{S} \subseteq \mathcal{R}$  denote the subset of valid rankings satisfying the agreed-upon constraints. Let  $W(r, u)$  denote the *utility* that placing  $u$  at position  $r(u)$  brings to the overall ranking: this is typically a function of the relevance score  $R$ , so that having higher relevance elements at top positions is rewarded. In other words,  $W$  is such that, if  $r^*$  denotes the ranking by decreasing  $R$ , then  $r^*$  is also the ranking maximizing the total utility (the so-called *Probability Ranking Principle* [24] in Information Retrieval). As the maximum-utility ranking  $r^*$  might not satisfy the given group-fairness constraint, the problem typically addressed in the literature is to find a valid ranking which maximizes the global utility, i.e.,

$$\tilde{r} \in \arg \max_{r \in \mathcal{S}} \sum_{u \in \mathcal{U}} W(r, u). \quad (1)$$

**Table 1: Example instance. Top row: identifiers and protected attribute (gender). Bottom row: relevance score  $R$ .**

$u_1, \sigma$	$u_2, \sigma$	$u_3, \varphi$	$u_4, \sigma$	$u_5, \sigma$	$u_6, \varphi$	$u_7, \varphi$	$u_8, \varphi$
0.97	0.93	0.89	0.81	0.73	0.72	0.64	0.62

EXAMPLE 1. Consider the case described in Table 1 and suppose that the group-fairness constraint requires to have at least  $\lfloor k/2 \rfloor$  individuals of each gender in the top- $k$  positions starting from  $k \geq 3$ .

The ranking by decreasing relevance  $r^* = \langle u_1, u_2, \dots, u_8 \rangle$  is not a valid ranking in this case, as  $\varphi$  is underrepresented in the top- $k$  positions for  $k = 4, 5, 6$ . A valid ranking which is as close as possible to  $r^*$  would be  $r' = \langle u_1, u_2, u_3, u_6, u_4, u_7, u_5, u_8 \rangle$ .

This approach stems from an information retrieval standpoint: the set of items to be ranked is the result of a query, and as long as the given group-fairness constraint is satisfied, it suffices for the application at hand to maximize the global utility. While at first sight this setting might seem adequate to rank people, maximizing global utility provides no guarantee to individuals, who care little about global utility. In Example 1, individuals  $u_4$  and  $u_5$  have been uniquely penalized from a *meritocratic fairness* point of view: They may accept the group-fairness constraints and agree with the fact that the produced ranking  $r$  is as close as possible to  $r^*$ , but nevertheless feel discriminated against, for being the only ones in a worse position in  $r$  than in  $r^*$  despite other solutions being possible. For example,  $\langle u_4, u_1, u_3, u_6, u_2, u_7, u_5, u_8 \rangle$  is valid and more favourable to  $u_4$ . In other words, while the use of group-fairness constraints is often desirable and may be required by law, certain individuals in a such a valid ranking might feel unfairly penalized, even when comparing only to individuals within the same group. As soon as a group-fairness constraint is enforced in ranking problems, some individual-level unfairness is inevitably introduced<sup>1</sup>.

In this paper we study the *problem of minimizing the amount of individual unfairness introduced when enforcing a group-fairness constraint*. While much of the literature for ranking attempts to maximize global utility, global quality metrics generally fail to adequately capture the treatment of *individuals*. Thus, differently from the literature which tries to maximize the global utility, we adopt Rawls's theory of justice [23], which advocates arranging social and financial inequalities to the benefit of the worst-off. Following this precept, a natural task is to find a ranking that, while satisfying the group-fairness constraint, maximizes the utility of the least-advantaged individual:

$$r' \in \arg \max_{r \in S} \min_{u \in \mathcal{U}} V(r, u). \quad (2)$$

Here  $V(r, u)$  represents the value (utility) that placing  $u$  at position  $r(u)$  brings to the individual  $u$ , *relative to  $u$ 's quality  $R(u)$* .

In Section 5 we provide an exact optimal solution for (2). This, however, is not the main focus of our paper. In fact, we can improve individual treatment even further through *randomization*.

**Randomization for individual fairness.** We next show how, by means of randomization, we can improve individual treatment over the best deterministic solution of (2). In particular, we show that there exists a *probability distribution over valid rankings*, where the minimum expected value that any individual gets is higher than is possible with any single ranking.

**EXAMPLE 2.** Consider the value function  $V(r, u) = r^*(u) - r(u)$ , i.e., the difference between the meritocratic ranking by relevance and the ranking produced. This is positive for individuals who are in a better (lower-ranked) position in  $r$  w.r.t.  $r^*$  and negative for others. It is easy to see that the ranking  $r'$  in Example 1 maximizes the minimum value of  $V(r, u)$ : in fact in order to have 3 ♀ in the first 6 positions, some ♂ has to give up at least 2 positions w.r.t.  $r^*$ .

Even when optimizing for (2), individual  $u_5$  in Example 2 might have concerns for being the one receiving the largest part of the burden of satisfying the group-fairness constraint. The only way

to improve on this situation is to introduce randomization into the process. This means producing a probability distribution over possible valid rankings instead of a single deterministic ranking.

**EXAMPLE 3.** Consider the same instance of Example 1. The following distribution over four rankings  $r_{1-4}$  maximizes the minimum expected value of  $V(r, u) = r^*(u) - r(u)$  among all individuals in  $\mathcal{U}$ :

$$\begin{aligned} \Pr(\langle u_1, u_4, u_3, u_7, u_2, u_6, u_5, u_8 \rangle) &= 1/4 \\ \Pr(\langle u_2, u_1, u_3, u_6, u_4, u_8, u_7, u_5 \rangle) &= 1/2 \\ \Pr(\langle u_2, u_1, u_3, u_7, u_5, u_6, u_4, u_8 \rangle) &= 1/16 \\ \Pr(\langle u_5, u_1, u_3, u_7, u_2, u_6, u_4, u_8 \rangle) &= 3/16 \end{aligned}$$

It is easy to check that, under this distribution, everyone has expected value at least  $-0.75$  (which is achieved by the four ♂), while under the best deterministic solution (Example 2) we had  $V(r, u_5) = -2 < -0.75$ .

While in Example 2 the burden required for ensuring the group-fairness constraint was all on  $u_4$  and  $u_5$ , in Example 3 it has been equally distributed among the four ♂. Notice that all four rankings in the distribution above satisfy the group-fairness constraint in Example 1. However, by combining these four rankings probabilistically, we have succeeded in achieving a higher minimum expected value than is possible via any single deterministic ranking. In fact, we have also minimized the disparity in the expected value that each individual receives: whereas requiring all expected values to be the same is not mathematically possible when satisfying group constraints, the solution above comes as close as possible by minimizing the maximum gap. A complete problem definition formalizing these ideas is given in Section 3.

**Implications and practical deployment.** In order to guarantee the maximum possible value is brought to each individual, in this paper we embrace randomization and produce a probability distribution over possible valid rankings. This distributional fairness approach is very well suited for a search context in which the same query can be served many times for different users of a platform (e.g., headhunters searching for a specific type of professional on a career-oriented social networking platform such as LinkedIn or XING). Notice also that *amortized fairness* in the sense of [4, 26] is an immediate application of this distributional approach: if there are several rankings to be made, we can draw them independently from a fair distribution of rankings, so that the empirical properties of the sample approach those of the fair distribution.

However, the usefulness of randomization extends to settings with a single, non-repeated trial (as in, e.g., university admissions). In this case it is an essential tool to secure “ex-ante” (procedural) individual fairness, i.e., fairness of the procedure by which the outcome is selected, as opposed to “ex-post” fairness, which is based on the final outcome alone (see, e.g., [5]).

Regarding implementation and transparency issues, notice that instead of treating the algorithm as a black box outputting a single ranking, one can make the entire distribution public. For instance, we can publish the distribution described in Example 3 above, letting all the individuals verify the expected value, as well as the fact that this distribution is optimal under the maxmin-fair criterion (see Section 3). Then one of the four rankings  $r_{1-4}$  can be picked at random, via any fair and transparent lottery mechanism or coin-tossing protocol. Moreover, our algorithms guarantee that the

<sup>1</sup>This situation resembles some cases in fair classification in which enforcing statistical parity constraints cause a form of unfairness from an individual viewpoint [9].

optimal distribution found is supported on a small (polynomial-size) set of rankings, even if the space of all valid rankings is exponential.

**Paper contributions and roadmap.** In the rest of this paper, following the randomized *maxmin-fairness framework* [12], we study how to efficiently and accurately compute this type of distributions over the rankings satisfying a given set of group-fairness constraints. We achieve the following contributions:

- We introduce the distributional maxmin-fair ranking framework and provide the formal problem statement (Section 3.1). We show that maxmin-fair ranking distribution maintains within-group meritocracy and, in certain cases, it has the desirable properties of being generalized Lorenz-dominant, and minimizing social inequality (Section 3.2).
- Our main result is an exact polynomial-time algorithm to find maxmin-fair distributions of many problems, including ranking (Section 4). A quicker method to find maxmin-fair distributions approximately is explained in Appendix A.4.
- We also provide an exact optimal solution (Section 5) for the deterministic version of the problem as in (2). This is achieved by a means of a variant of Celis et al. [7]. We use this as a baseline allowing us to quantify the advantage of probabilistic rankings over the optimal deterministic ranking.
- Our experiments on two real-world datasets confirm empirically the advantage of probabilistic rankings over deterministic rankings in terms of minimizing the inequality for the worst-off individuals (Section 6).

To the best of our knowledge, this is the first work studying the problem of minimizing the amount of individual unfairness introduced when enforcing group-fairness constraints in ranking. A major contribution is showing how randomization can be a key tool in reconciling individual and group fairness: we believe that this might hold for other problems, besides ranking.

## 2 RELATED WORK

There are some works on algorithmic fairness focused on individual fairness, but none of them considers them in conjunction with group fairness. Dwork et al [9] introduce a notion of individual fairness in classification problems. Roughly speaking, their definition requires that all pairs of similar individuals should be treated similarly. This is impossible to satisfy with a deterministic classifier so, similarly to ours, their definition of fairness requires randomized algorithms. The individual similarity metric is assumed given, while they base their notion of 'similar treatment' on the difference between the probabilities of a favourable classification. Kearns et al. [19] introduce the notion of *meritocratic fairness* in the context of selecting a group of individuals from incomparable populations (with no group-fairness constraint). Their notion intuitively requires that less qualified candidates do not have a higher chance of getting selected than more qualified ones. Another work focusing on individual fairness is that of Biega et al. [4], which aims at achieving *equity-of-attention fairness* amortized across many rankings, by requiring that exposure be proportional to relevance.

Our previous work [12] presents a very general framework to deal with individual fairness, based on randomized maxmin-fairness: the idea is to use a distribution of solutions in order to

maximize the expected value for the worst-off individual. In particular, [12] analyzes the case of unweighted matching with no group-fairness constraint: presents efficient algorithms and shows that these maxmin-fair matching distributions minimize inequality. While the techniques from [12] are combinatorial and can only deal with unrestricted matchings, we greatly generalize the algorithmic results therein via convex optimization techniques, showing that for a wide class of problems (including weighted matching and ranking with constraints), a maxmin-fair distribution may be found in polynomial time; we only require the existence of a *weighted optimization oracle* (see Section 4).

The bulk of recent literature on fairness in ranking [3, 7, 13, 26, 30, 31, 34] and learning-to-rank [10, 21, 27] deals with group fairness. Singh and Joachims [26] propose an algorithm computing a fair probabilistic ranking maximizing expected global utility. The fairness constraints expressible in their framework apply to the ranking distribution and not to each single ranking, as required by the group-fairness constraints we use. Celis et al. [7] also investigate fair ranking with group-fairness constraints with an objective function of the form (1), assuming the values  $W(r, u)$  satisfy the Monge condition. They give a polynomial-time algorithm for disjoint protected groups, and a faster greedy algorithm that works when only upper bound constraints are given. When the protected groups are allowed to overlap the problem becomes NP-hard and a polynomial-time approximation algorithm is provided in [7].

## 3 MAXMIN-FAIR RANKING

We are given a set of  $n$  individuals to be ranked  $\mathcal{U}$ , a partition of  $\mathcal{U}$  into groups  $C_1, \dots, C_t$ , and a relevance function  $R : \mathcal{U} \rightarrow \mathbb{R}$ . For the sake of simplicity we assume that ties are broken so that all  $R(u)$  are distinct. Moreover, we are given group-fairness constraints as in [7], defined by setting, for each  $i \in [n]$  and  $k \in [t]$ , a lower bound  $l_i^k \in \mathbb{N}$  and an upper bound  $u_i^k \in \mathbb{N}$  on the number of individuals from class  $k$  in the first  $i$  positions. We denote by  $\mathcal{R}$  the set of all possible rankings of  $\mathcal{U}$  (bijections from  $\mathcal{U}$  to  $[n]$ ), and by  $\mathcal{S} \subseteq \mathcal{R}$  the set of all valid rankings:

$$\mathcal{S} = \left\{ r \in \mathcal{R} \mid l_i^k \leq |\{u \in C_k \mid r(u) \leq i\}| \leq u_i^k \quad \forall i \in [n], k \in [t] \right\}. \quad (3)$$

Finally, we consider a value function  $V : \mathcal{S} \times \mathcal{U} \rightarrow \mathbb{R}$  such that  $V(r, u)$  represents the value (utility) that placing  $u$  at position  $r(u)$  brings to the individual  $u$ , relative to  $u$ 's quality  $R(u)$ . As we are interested in modeling meritocratic fairness, our value function must take into consideration the input relevance score  $R(u)$  and the produced ranking  $r(u)$ . We consider value functions of the form:

$$V(r, u) = f(r(u)) - g(u), \quad (4)$$

where  $f : [n] \rightarrow \mathbb{R}$  is a decreasing function and  $g : \mathcal{U} \rightarrow \mathbb{R}$  is increasing in  $R(u)$ .

The intuition is the following: suppose that being assigned at position  $i$  carries intrinsic utility  $f(i)$ , while  $u$ 's merit for the ranking problem is  $g(u)$  (which may depend on  $u$  and hence also on  $R(u)$ ); then  $V(r, u)$  measures the net difference between  $f(r(u))$  and  $g(u)$ , i.e., how much  $u$  has gained in  $r$  w.r.t.  $u$ 's actual merit. In typical applications we can take any decreasing function  $p : [n] \rightarrow \mathbb{R}^{\geq 0}$  encoding *position bias* or *exposure* (see [8] for common models) and set  $f = p$  and  $g = p \circ r^*$ . As simple examples, by setting

$p(i) = n - i$  and  $p(i) = \log(n/i)$ , we can get  $V(r, u) = r^*(u) - r(u)$  and  $V(r, u) = \log(\frac{r^*(u)}{r(u)})$ . When the ranking is a selection process where  $k \in \mathbb{N}$  individuals are selected and there is no advantage to being ranked first over  $k_{th}$  as long as one is selected, we may use

$$V(r, u) = \begin{cases} 1, & \text{if } r^*(u) > k \text{ and } r(u) \leq k \\ -1, & \text{if } r^*(u) \leq k \text{ and } r(u) > k \\ 0, & \text{otherwise.} \end{cases}$$

These are but a few examples. Determining which value function  $V$  is best from a psychological or economical standpoint is beyond the scope of this work. Instead we take  $V$  as given and design algorithms which can efficiently deal with any function of the form (4).

### 3.1 Maxmin-fairness framework

Consider an input instance  $\mathcal{T}$  of a general search problem which defines implicitly a set  $\mathcal{S} = \mathcal{S}(\mathcal{T})$  of feasible solutions, assumed to be finite and non-empty. Let  $\mathcal{U}$  denote a finite set of individuals and let us associate with each solution  $S \in \mathcal{S}$  and each individual  $u \in \mathcal{U}$  a real-valued satisfaction  $A(S, u) \in \mathbb{R}$  (which in [12] takes binary values). Consider a *randomized algorithm*  $\mathcal{A}$  that, for any given problem instance  $\mathcal{T}$ , always halts and selects a solution  $\mathcal{A}(\mathcal{T}) \in \mathcal{S}$ . Then  $\mathcal{A}$  induces a probability distribution  $D$  over  $\mathcal{S}$ :  $\Pr_D[S] = \Pr[\mathcal{A}(\mathcal{T}) = S] \forall S \in \mathcal{S}$ . Denote the *expected satisfaction* of each  $u \in \mathcal{U}$  under  $D$  by  $D[u] \triangleq \mathbb{E}_{S \sim D}[A(S, u)]$ . A distribution  $F$  over  $\mathcal{S}$  is *maxmin-fair* for  $(\mathcal{U}, A)$  if it is impossible to improve the expected satisfaction of any individual without decreasing it for some other individual which is no better off, i.e., if for all distributions  $D$  over  $\mathcal{S}$  and all  $u \in \mathcal{U}$ ,

$$D[u] > F[u] \implies \exists v \in \mathcal{U} \mid D[v] < F[v] \leq F[u]. \quad (5)$$

Maxmin-fair distributions always exist [12]. Due to the convexity of the set of feasible probability distributions, an equivalent definition can be given based on the sorted vectors of expected satisfactions. Given a distribution  $D$  over  $\mathcal{S}$ , let  $D \uparrow = (\lambda_1, \dots, \lambda_n)$  be the vector of expected satisfactions  $(D[u])_{u \in \mathcal{U}}$  sorted in increasing order. Let  $>$  denote the lexicographical order of vectors: i.e.,  $(v_1, \dots, v_n) > (w_1, \dots, w_n)$  iff there is some index  $i \in [n]$  such that  $v_i > w_i$  and  $v_j = w_j$  for all  $j < i$ . Write  $v \geq w$  if  $v = w$  or  $v > w$ . Then a distribution  $F$  over  $\mathcal{S}$  is *maxmin-fair* if and only if  $F \uparrow \geq D \uparrow$  for all distributions  $D$  over  $\mathcal{S}$  [12].

**PROBLEM 1 (MAXMIN-FAIRNESS IN COMBINATORIAL SEARCH).** *Given a fixed search problem, a set  $\mathcal{U}$  of individuals, and a satisfaction function  $A$ , design a randomized algorithm  $\mathcal{A}$  which always terminates and such that, for each instance  $\mathcal{T}$ , the distribution of  $\mathcal{A}(\mathcal{T})$  is maxmin-fair for  $(\mathcal{U}, A)$ .*

Problem 1 is a general formulation of maxmin-fairness in search problems. Different choices for the set of feasible solutions  $\mathcal{S}$  and the satisfaction function  $A$  lead to different algorithmic problems. The problem involves continuous optimization over infinitely many distributions, each defined over the set  $\mathcal{S}$  of valid solutions (which is exponential-size). Despite these difficulties, we will show that Problem 1 is tractable under mild conditions (Section 4).

García-Soriano and Bonchi [12] instantiate Problem 1 with the case of matching. The main problem studied in the rest of this paper is obtained by instantiating Problem 1 with the case of *ranking under*

*group-fairness constraints with individual-level value function*: in our setting  $\mathcal{S}$  is the set of rankings  $r$  over  $\mathcal{U}$  satisfying the group-fairness constraints, and  $A(S, u)$  is our value function  $V(r, u)$ .

**PROBLEM 2 (MAXMIN-FAIR RANKING WITH GROUP-FAIRNESS CONSTRAINTS).** *Given a set of individuals to be ranked  $\mathcal{U}$ , a partition of  $\mathcal{U}$  into groups, a set  $\mathcal{S}$  of rankings satisfying a given set of group-fairness constraints as defined in (3), and a value function  $V$  as defined in (4), design a randomized algorithm which outputs rankings in  $\mathcal{S}$ , such that its output distribution over  $\mathcal{S}$  is maxmin-fair.*

### 3.2 Properties of maxmin-fair rankings

We next state some important properties of maxmin-fair rankings. For the sake of readability, the proofs can be found in the Appendix. The first property states the maintenance of the meritocratic order within each group of individuals (e.g., gender).

**THEOREM 3.1 (INTRA-GROUP MERITOCRACY).** *For any two individuals  $u_1, u_2 \in \mathcal{U}$  belonging to the same group and such that  $R(u_1) \geq R(u_2)$ , it holds that, if a distribution  $F$  over valid rankings  $\mathcal{S}$  is maxmin-fair, then  $\mathbb{E}_{r \sim F}[f(r(u_1))] \geq \mathbb{E}_{r \sim F}[f(r(u_2))]$ .*

Our second property employs the notion of (*generalized*) *Lorenz dominance* from [25], a property indicating a superior distribution of net incomes. Consider two ranking distributions  $A$  and  $B$ . Let  $A_{(k)} = A \uparrow[k]$  denote the  $k_{th}$  element of the expected satisfaction values sorted in increasing order. Then  $A$  *dominates*  $B$  if  $\sum_{i=1}^k A_{(i)} \geq \sum_{i=1}^k B_{(i)} \forall k \in [n]$ , i.e., the expected cumulative satisfaction of the bottom individuals is always higher or equal in  $A$ .

A distribution is *generalized Lorenz-dominant* if it dominates every other distribution. When it exists, such a distribution has a strong claim to being superior to all others, in terms of equity and efficiency [20, 25, 28]. A generalized Lorenz-dominant distribution must also be maxmin-fair. We show that a dominant distribution does exist for rankings, in the important case where only upper bound constraints are given in (3). Notice that in the case of two groups (e.g., a binary protected attribute), lower bound constraints may be replaced with an equivalent set of upper bound constraints.

**THEOREM 3.2.** *The maxmin-fair ranking with upper bounds is generalized Lorenz-dominant.*

Since  $\sum_{u \in \mathcal{U}} V(r, u)$  is a constant independent of  $r$ , an easy consequence of Theorem 3.2 is that the maxmin-fair distribution also *minimizes social inequality* in the sense of [12]: i.e., the maximum difference between the expected satisfactions of two users.

**COROLLARY 3.1.** *The maxmin-fair ranking with upper bounds minimizes  $\max_{u \in \mathcal{U}} D[u] - \min_{v \in \mathcal{U}} D[v]$  over all ranking distributions  $D$ , as well as any other quantile range.*

Moreover, by the majorization inequality [18], it also maximizes any social welfare function that is additively separable, concave and symmetric w.r.t.  $\mathcal{U}$ :

**COROLLARY 3.2.** *Suppose  $f : \mathbb{R} \rightarrow \mathbb{R}$  is concave. When only upper bound constraints are present, the maxmin-fair distribution maximizes  $\sum_{u \in \mathcal{U}} f(D[u])$  over all ranking distributions  $D$ .*

In particular, in this case the maxmin-fair distribution minimizes the variance of  $D \uparrow$ , and when the values  $V(r, u)$  are positive, it also

maximizes, for instance, the Nash social welfare [22] (geometric mean) of expected satisfactions. It must also minimize the Gini inequality index when it is well-defined [25].

## 4 ALGORITHM

We show that our fair ranking problem (Problem 2) is efficiently solvable. Notice that the set  $\mathcal{S}$  of valid solutions can be exponential-size, so enumerating  $\mathcal{S}$  is out of the question in an efficient algorithm. Instead, we need a method to quickly single out the best solutions to combine for a maxmin-fair distribution. To show how this can be done, we abstract away from the specifics of the problem and show how to find a maxmin-fair distributions of general search problems (Problem 1). The following notion is key:

**DEFINITION 1.** A weighted optimization oracle for  $A : \mathcal{S} \times \mathcal{U} \rightarrow \mathbb{R}$  is an algorithm that, given  $w : \mathcal{U} \rightarrow \mathbb{R}^{\geq 0}$ , returns  $S^*(w)$  and  $A(S^*(w), u)$  for all  $u \in \mathcal{U}$ , where

$$S^*(w) \in \arg \max_{S \in \mathcal{S}} \sum_{u \in \mathcal{U}} w(u) \cdot A(S, u). \quad (6)$$

Roughly speaking, the intuition why these oracles are important is the following. Suppose we have constructed a distribution  $D$  which is not maxmin-fair. By putting more weight on the individuals less satisfied by  $D$ , we can use the weighted optimization oracle to find a new solution  $S$  placing more emphasis on them, which can be added to “push  $D$  towards maxmin-fairness”.

Designing an efficient weighted optimization oracle is a problem-dependent task. Our first algorithmic result reveals that their existence suffices to solve Problem 1 efficiently.

**THEOREM 4.1.** *Given a weighted optimization oracle, Problem 1 is solvable in polynomial time.*

We emphasize that Theorem 4.1 is very general and its applicability is in no way limited to ranking problems, or to value functions of a certain form. They apply to an arbitrary search problem  $P$  (e.g., searching for a ranking, a matching, a clustering...) and an arbitrary set of individuals. As long as an efficient weighted optimization algorithm exists for  $P$ , it yields efficient algorithms for maxmin-fair- $P$  (Problem 1). The wide applicability of this condition implies that maxmin-fair distributions may be efficiently solved in a great many cases of interest: most polynomial-time solvable problems studied in combinatorial optimization (e.g., shortest paths, matchings, polymatroid intersection...) admit a polynomial-time weighted optimization oracle. Thus, Theorem 4.1 extends in a new direction the results of [12]: as efficient weighted matching algorithms exist, the main result of [12] becomes a corollary to Theorem 4.1 (up to the loss of a polynomial factor in runtime).

More importantly for us, the same holds for constrained ranking.

**THEOREM 4.2.** *Ranking with group-fairness constraints as in (3) and a value function of the form (4) admits a polynomial-time weighted optimization oracle.*

**COROLLARY 4.1.** *Maxmin-fair ranking with group-fairness constraints (Problem 2) is solvable in polynomial-time.*

In Section 4.1 we prove Theorem 4.1 by solving a sequence suitably designed linear programs. Each of these programs requires exponentially many constraints to be written down explicitly, but

can nonetheless be solved efficiently via the ellipsoid method using a weighted optimization oracle. (As explained in Appendix A.4, if we settle for some approximation error, these LPs can also be solved approximately using techniques to solve zero-sum games and packing/covering LPs [11, 32, 33].) Finally, in Section 4.2 we show the existence of weighted optimization oracles for ranking (Theorem 4.2).

### 4.1 Proof of Theorem 4.1

We start by showing a weaker result concerning the computation of the optimal expected satisfaction values, rather than the actual distribution of solutions.

**LEMMA 4.1.** *Given a weighted optimization oracle, the expected satisfactions of a maxmin-fair distribution can be computed in polynomial time.*

**PROOF.** Let  $F$  be a maxmin-fair distribution. We maintain the invariant that we know the expected satisfaction  $\alpha_v$  of  $F$  for all  $v$  in a subset  $K \subseteq \mathcal{U}$ :

$$F[v] = \alpha_v \text{ for all } v \in K, \quad (7)$$

$$K \neq \emptyset \implies F[v] \geq \max_{w \in K} \alpha_w \text{ for all } v \notin K. \quad (8)$$

Initially  $K = \emptyset$ . We show how to augment  $K$  in polynomial time while maintaining (7) and (8), which gives the result since  $K = \mathcal{U}$  will be reached after at most  $|\mathcal{U}|$  iterations. We need to find the largest minimum expected satisfaction possible outside  $K$  for a distribution  $D$  subject to the constraints that the expected satisfaction inside  $K$  must be equal to  $\alpha_v$ . By (7), (8) and the lexicographical definition of maxmin-fairness, for any distribution  $D$  the constraints  $D[v] = \alpha_v$  for all  $v \in K$  are equivalent to the constraints  $D[v] \geq \alpha_v$  for all  $v \in K$ . We can write our optimization problem as the following (primal) linear program:

$$\begin{aligned} \max \quad & \lambda \\ \text{s.t.} \quad & \sum_{S \ni v} -p_S \cdot A(S, v) \leq -\alpha_v \quad \forall v \in K \\ & \lambda + \sum_{S \ni v} -p_S \cdot A(S, v) \leq 0 \quad \forall v \notin K \\ & \sum_{S \in \mathcal{S}} p_S = 1 \\ & p_S \geq 0, \end{aligned} \quad (9)$$

whose dual is

$$\begin{aligned} \min \quad & \mu - \sum_{v \in K} \alpha_v w_v \\ \text{s.t.} \quad & \mu - \sum_{v \in \mathcal{S}} w_v \cdot A(S, v) \geq 0 \quad \forall S \in \mathcal{S} \\ & \sum_{v \in \mathcal{U}} w_v = 1 \\ & w_v \geq 0. \end{aligned} \quad (10)$$

The dual (10) has  $|\mathcal{U}|$  variables but a possibly exponential number of constraints (one for each candidate solution  $S$ ). To get around

this difficulty, observe that it can be written in the equivalent form

$$\begin{aligned} \min \quad & \left[ \max_{S \in \mathcal{S}} \sum_{v \in \mathcal{S}} w_v \cdot A(S, v) \right] - \sum_{v \in K} \alpha_v w_v \\ \text{s.t.} \quad & \sum_{v \in \mathcal{U}} w_v = 1 \\ & w_v \geq 0. \end{aligned} \quad (11)$$

This formulation makes it apparent that, given a weighted optimization oracle, we can construct a *separation oracle* for the dual, i.e., an algorithm that given a candidate solution to (10) and a parameter  $\lambda$ , returns “yes” if it is a feasible solution of value at most  $\lambda$ , and otherwise returns “no” along with some violated constraint or reports the fact that the value of the candidate solution is larger than  $\lambda$ . Indeed, given  $\{w_v\}_{v \in \mathcal{U}}$  and  $\lambda \in \mathbb{R}$ , we can determine if the optimum of (11) is at most  $\lambda$  by using the weight optimization oracle and answering “yes” if the weight of the solution is no larger than  $\lambda + \sum_{v \in K} \alpha_v w_v$ . Otherwise the separation oracle answers “no” and reports a violated constraint, given either by the constraint  $\sum_v w_v = 1$ , which can be checked separately, or by the constraint  $\sum_{v \in S^*} w_v \cdot A(S^*, v) \leq \mu$ , where  $S^*$  is the solution found by the weighted optimization oracle.

The existence of a separation oracle for a linear program implies its polynomial-time solvability via the ellipsoid algorithm [15]. Hence (11) can be solved exactly in polynomial time, and we can find an optimal solution to (11). Let us denote the optimal primal and dual solutions by  $\{p_S^*\}_{S \in \mathcal{S}}$  and  $\{w_v^*\}_{v \in \mathcal{U}}$ . Suppose now that the optimum value of (9) is  $\lambda^*$ ; notice that if  $K \neq \emptyset$ , we must have  $\lambda^* \geq \max_{v \in K} \alpha_v$  by our assumptions (7) and (8). Let  $K' = \text{support}(w^*) \setminus K = \{v \notin K \mid w_v^* > 0\}$ . By complementary slackness, for every  $v \in K'$  its corresponding primal constraint in (9) is tight, hence  $\sum_{S \ni v} p_S^* A(S, v) = \lambda^*$ . From the lexicographical definition of maxmin-fairness we infer that  $F[v] = \lambda^*$  for all  $v \in K'$  and  $F[v] \geq \lambda^*$  for all  $v \notin K$ . Therefore adding  $K'$  to  $K$  maintains the invariants (7) and (8) if we set  $\alpha_v = \lambda^*$  for all  $v \in K'$ . This allows us to augment  $K$  as long as  $K' \neq \emptyset$ .

On the other hand, if  $K' = \emptyset$ , then  $K \neq \emptyset$  (as  $\sum_v w_v = 1$ ) and  $\lambda^* = \max_{v \in K} \alpha_v$ , since the objective function did not increase since the last iteration. In this case we simply add the constraint  $\mu - \sum_{v \in K} \alpha_v w_v = \lambda^*$  to (10) and change the objective function to minimize  $\sum_{v \in K} w_v$ . This also yields an optimal solution to (11). But in this case the new solution  $w^{**}$  must satisfy  $\text{support}(w^{**}) \setminus K \neq \emptyset$ , so we are back to the previous case.

We repeat this process until  $K = \mathcal{U}$ . The number of iterations is at most  $|\mathcal{U}|$ , and each iteration runs in polynomial time.  $\square$

**Proof of Theorem 4.1.** Consider the last pair of LP programs used in the proof of Lemma 4.1 (i.e., when  $K \cup K' = \mathcal{U}$ ). We used the separation oracle and the ellipsoid algorithm to solve the dual LP (11); it remains to show that we can also find a solution to the primal problem (9), whose variables define the maxmin-fair distribution. If all numbers  $A(S, u)$  are rationals whose numerators and denominators are specified with  $b$  bits, then the number  $T$  of calls to the separation oracle during the run of the ellipsoid algorithm can be bounded by a polynomial in  $|\mathcal{U}|$  and  $b$  (see [15]). Consider the subprogram  $P$  of (11) formed by using only these  $T$  constraints, along with  $\sum_v w_v = 1$  and the non-negativity constraints. If we run the ellipsoid algorithm (which

is deterministic) on the new subprogram  $P$  instead, we will find the same solution, because the separation oracle will return the exact same sequence of solutions. Since the ellipsoid algorithm is guaranteed to find an optimal solution, it follows that the reduced set of constraints is enough by itself to guarantee that the optimum of LP (11) is at least  $\lambda^*$  (hence exactly  $\lambda^*$ ); all other constraints in (11) are redundant (their inclusion does not further increase the minimum objective value). The dual of this subprogram  $P$  is a subprogram  $Q$  of the primal LP (9) using only  $T$  of the variables  $p_S$  and having the same optimal value as (11) and (9). Since  $Q$  has a polynomial number of variables and constraints, it can be solved explicitly in polynomial time; any solution to this reduced primal subprogram  $Q$  gives the desired distribution.  $\square$

Pseudocode for the maxmin-fair algorithm is given below.

#### Algorithm 1: Maxmin-fair solver

```

input : User set  $\mathcal{U}$ ; weighted optimization oracleA for
         $A : \mathcal{S} \times \mathcal{U} \rightarrow \mathbb{R}$ 
output : A maxmin-fair distribution for  $\mathcal{S}$ 
1  $K \leftarrow \emptyset$ 
2  $\alpha_u \leftarrow -\infty$  for all  $u \in \mathcal{U}$ 
3 while  $K \neq \mathcal{U}$  do
4   Solve (11) using oracleA as separation oracle.
5   Let  $\lambda^*$  be the optimal value.
6   Let  $\{w_v^*\}_{v \in \mathcal{U}}$  be a solution with value  $\lambda^*$  and
   support( $w^*$ )  $\setminus K \neq \emptyset$ .
7    $K' = \text{support}(w^*) \setminus K = \{v \notin K \mid w_v^* > 0\}$ .
8    $\alpha_v \leftarrow \lambda^*$  for all  $v \in K'$ 
9    $K = K \cup K'$ 
10  $V \leftarrow$  violated constraints found by the separation oracle in
   previous calls.
11 Solve the subprogram  $Q$  of (9) comprising the constraints in  $V$  and
   the simplex constraints  $p_S \geq 0$ ,  $\sum_S p_S = 1$ .
12 return an optimal solution  $\{p_S^*\}_{S \in \mathcal{S}}$  to  $Q$ .
```

## 4.2 Proof of Theorem 4.2

**PROOF.** Sort  $\mathcal{U} = \{u_1, \dots, u_n\}$  by decreasing order of  $w$  so that

$$w(u_1) \geq w(u_2) \geq \dots \geq w(u_n), \quad (12)$$

and let us identify  $\mathcal{U}$  with the set  $[n]$  for ease of notation, so that  $u_i = i$ . Recall that the positions  $[n]$  are sorted by decreasing  $f$ :

$$f(1) \geq f(2) \geq \dots \geq f(n). \quad (13)$$

Define  $B(i, u) = f(i) - g(u)$  and  $W_{iu} = w(u) \cdot B(i, u)$ . Observe that, because of the orderings defined by (12) and (13), the matrix  $W$  satisfies the following “Monge property”: if  $i < j$  and  $u < v$ , then  $W_{iu} + W_{jv} \geq W_{iv} + W_{ju}$ . Indeed,

$$\begin{aligned} W_{iu} + W_{jv} - (W_{iv} + W_{ju}) &= w(u)(B(i, u) - B(j, u)) \\ &\quad + w(v)(B(j, v) - B(i, v)) \\ &= (w(u) - w(v))(f(i) - f(j)) \geq 0. \end{aligned}$$

Thus we may apply the algorithm<sup>2</sup> from [7] to find a valid ranking  $r$  maximizing  $\sum_{u \in \mathcal{U}} W_{r(u), u} = \sum_{u \in \mathcal{U}} w(u) \cdot V(r, u)$ , as required by

<sup>2</sup>In [7] an additional monotonicity property is assumed (that  $W_{iu}$  is decreasing with  $u$ ), but it is easy to check that it is not actually needed.

the definition of weighted optimization oracle from Section 4. Then we may compute each  $V(r, u)$  explicitly using  $f$  and  $g$ .  $\square$

An important case is where only upper bounds are given in the group constraints, i.e., when the set of valid rankings is of the form

$$S = \{r \in \mathcal{R} \mid |\{u \in C_k \mid r(u) \leq i\}| \leq u_i^k \ \forall i \in [n], k \in [t]\}. \quad (14)$$

Plugging in the algorithm from [7] into Theorem 4.2 we obtain:

**Algorithm 2:** Weighted optimization oracle for ranking with upper bounds

- input** : Set of individuals  $\mathcal{U}$ ; weight function  $w : \mathcal{U} \rightarrow \mathbb{R}$ ; value function  $V : S \times \mathcal{U} \rightarrow \mathbb{R}$   
**output**: Best response ranking  $r$  and  $V(r, u)$  for all  $u \in \mathcal{U}$
- 1 Sort individuals in  $\mathcal{U}$  in order of decreasing weight:  
 $w(u_1) \geq w(u_2) \geq \dots \geq w(u_n)$ .
  - 2 For each position  $i \in [n]$  in increasing order (as in (13)), let  $u$  be the smallest-index unassigned individual whose additional placement at position  $i$  does not violate the group upper bound constraints in the first  $i$  positions, and set  $r(u) = i$ .
  - 3 Return  $r$  and  $V(r, u)$  for all  $u \in \mathcal{U}$ .

**COROLLARY 4.2.** *Algorithm 2 is a weighted optimization oracle for ranking with upper bounds.*

## 5 DETERMINISTIC BASELINE

In this section we present an exact optimal solution for the deterministic version of the problem as formulated in (2). This is useful for our experiments (Section 6) as it allows us to quantify the advantage of probabilistic rankings over deterministic rankings in terms of the amount of individual fairness maintained.

Although Celis et al. [7] study the problem of the form (1), we devise a variant of their algorithm to deal with the problem as in (2): this variant can be shown to provide the *optimal deterministic ranking* solution to the constrained ranking problem (2) when the group-fairness constraints are expressed in terms of upper bounds on the number of elements from each class that appear in the top- $k$  positions, as in (14). As noted in Section 3 and in [7], in the case of two disjoint groups (e.g., a binary protected attribute such as gender), lower bound constraints may be replaced with an equivalent set of upper bound constraints.

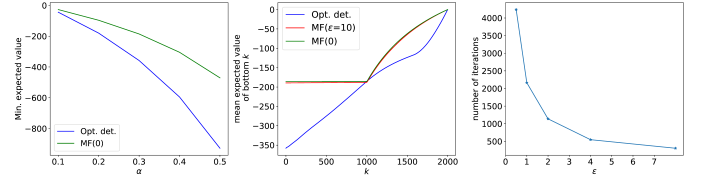
**Algorithm 3:** Deterministic baseline

- input** : Set of individuals  $\mathcal{U}$ ; relevance function  $R : \mathcal{U} \rightarrow \mathbb{R}^{\geq 0}$   
**output**: Deterministic ranking  $r$ .
- 1 Sort individuals in  $\mathcal{U}$  in order of decreasing score:  
 $R(u_1) \geq R(u_2) \geq \dots \geq R(u_n)$ .
  - 2 For each position  $i \in [n]$  in increasing order (as in (13)), let  $u$  be the smallest-index unassigned individual whose additional placement at position  $i$  does not violate the group upper bound constraints in the first  $i$  positions, and set  $r(u) = i$ .
  - 3 Return  $r$ .

At the basis of our deterministic baseline (Algorithm 3) lies the idea of using the function  $\text{softmax}(x_1, \dots, x_n) = -\ln(\sum_{i=1}^n e^{-x_i})$

**Table 2:** Minimum expected value produced by MF(0) and optimal deterministic solution, spread (maximum - minimum) of expected value, Gini inequality index (%), and discounted cumulative gain for IIT-JEE and Law-schools datasets for different values of  $\alpha$ .

	IIT-JEE			Law-schools		
	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
$\min V(\text{MF})$	-26.82	-96.44	-185.7	-0.87	-1.03	-5.48
$\min V(\text{det})$	-44	-180	-358	-1	-2	-10
$\text{sprd}(\text{MF})$	53.76	193.4	372.8	0.97	1.14	6.19
$\text{sprd}(\text{det})$	433	899	1192	6	7	32
$\text{gini}(\text{MF})$	0.6714	2.413	4.658	0.0005	0.0011	0.05
$\text{gini}(\text{det})$	1.062	3.772	7.027	0.0010	0.0020	0.08
$\text{DCG}(\text{MF})$	84847 $\pm$ 118	84444 $\pm$ 180	84062 $\pm$ 242	31379 $\pm$ 1	31379 $\pm$ 1	31379 $\pm$ 1
$\text{DCG}(\text{det})$	85123	85008	84807	31380	31380	31380



**Figure 1:** IIT-JEE: Minimum expected value produced by MF(0) and optimal deterministic solution (left); distribution of expected value  $V(r, u)$  (for  $\alpha = 0.3$ ) (center); number of iterations (calls to the optimization oracle) vs error  $\epsilon$  (right).

to force the algorithm from [7] to approximately maximize a minimum instead of a sum, and observe that the limiting behaviour of the function  $x \rightarrow \text{softmax}(Mx)/M$  must also occur in this case for finite  $M$  because the algorithm from [7] does not depend on the specific values of the matrix  $W_{ij}$ , but only on the existence of an ordering of rows/columns of  $W$  where the Monge property holds (see Section 4.2).

**THEOREM 5.1.** *When the group-fairness constraints are defined only by upper bounds, Algorithm 3 returns a ranking  $r'$  such that*

$$r' \in \arg \max_{r \in S} \min_{u \in \mathcal{U}} V(r, u).$$

## 6 EXPERIMENTS

**Datasets.** We use two real-world datasets containing gender information and one score for each individual. Our first dataset comes from the IIT Joint Entrance Exam (known as IIT-JEE 2009) [6]<sup>3</sup>, from which we select the top  $N = 1000$  scoring males and the top  $N$  scoring females. The score distribution is heavily biased at the top, with just four females making the top-100. Our second dataset is much less skewed: it contains admissions data from all of the public law schools in the United States<sup>4</sup>. We use the top  $N = 1000$  LSAT scorers, of whom 362 are female.

**Settings.** We impose the following group-fairness constraints, parameterized by  $\alpha \in [0, \frac{1}{2}]$ : at least  $\lceil \alpha \cdot k - 1 \rceil$  females should be ranked in the top  $k$ , for  $k = 1, 2, \dots, 2N$ . We employ  $V(r, u) = r^*(u) - r(u)$  as our value function, where  $r^*$  is the ranking by decreasing score.

<sup>3</sup><https://jumpshare.com/v/yRUSJrnw3bzGGNf0jL3A>

<sup>4</sup><http://www.seaphe.org/databases.php>



**Algorithms.** We implement our maxmin-fair solver for ranking, using the technique described in Appendix A.4 to solve the LPs approximately with an additive error parameter  $\epsilon$ ;  $\epsilon = 1$  corresponds to an additive error in expected ranking position of 1 (out of 2000 for IIT-JEE and out of 1000 for Law school). We denote by  $\text{MF}(\epsilon)$  the ranking distribution produced by our approximate maxmin-fair algorithm with parameter  $\epsilon$ , and by  $\text{MF}(0)$  the one obtained with the smallest  $\epsilon$  tested (0.5). Our code is available on Dropbox<sup>5</sup>.

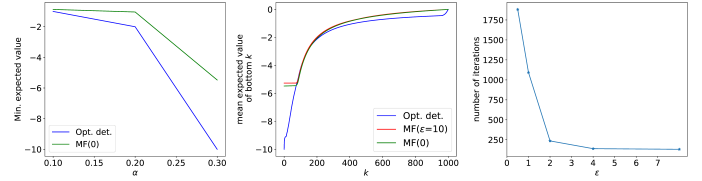
In order to quantify the advantage of probabilistic rankings over the optimal deterministic ranking, we also test the deterministic algorithm we devised (Algorithm 3) to solve the problem in (2). This provides the strongest possible deterministic competitor for our algorithm.

**Measures.** Besides comparing the minimum expected value, which is the main focus of our work, we also report other measures of inequality of the produced solution: spread (maximum - minimum) of expected value and Gini inequality index [14] (after normalizing values to the interval  $[0, 1]$  to make the index well-defined). Finally, to examine if there is a loss in global ranking quality, we use the popular *discounted cumulative gain* metric [4, 6, 7, 17, 26, 29], which can be defined as  $\text{DCG}(r) = \sum_{u \in \mathcal{U}} \text{score}(u) / \log(r(u) + 1)$ .

**Results.** The first two rows of Table 2 report the expected value (over a random ranking from the distribution) of the solution for the worst-off individual; we can observe that the maxmin-fair solution improves significantly on the optimal deterministic solution, with the gap between the two increasing with  $\alpha$  (the strength of the group-fairness constraint). The same can be observed in Figure 1 (left) and Figure 2 (left) for the two datasets. We do not report the *average* value of the solution for all individuals because it is the same for every ranking, as rankings are bijections onto  $[n]$ .

In Table 2 we can also observe that the inequality measures for the maxmin-fair solution are always smaller than the optimal deterministic one. Finally, we report the ranking-quality measure DCG. Since, unlike the three other measures in Table 2 DCG is defined for deterministic rankings, we report average and standard deviation. We see that DCG is nearly the same for  $\text{MF}(0)$  and det. Thus in this experiment *improving individual fairness with respect to a group-only fairness solution incurs a negligible loss in DCG*.

Figure 1 (center) and Figure 2 (center) depict the average expected value of the bottom  $k$  individuals in three solutions: our best solution  $\text{MF}(0)$ , an approximate solution with  $\epsilon = 10$ , and the optimal deterministic solution. The peculiar behaviour of the curve in Figure 1 (center) (constant up to roughly  $k = n/2$  for MF) is due to the skew of the input scores, which forces the maxmin-fair solution to essentially increase the ranking positions of most men by a certain minimum amount  $X$  and decrease that of most women by  $X$  with the best possible distribution. We notice that the maxmin-fair solution yields stronger cumulative value to the worst-off users than the other two do, for any  $k$ . In particular, the maxmin-fair solution found Lorenz-dominates the approximate one and the deterministic one, in accordance with Theorem 3.2. Because of the error allowed, the approximate solution  $\text{MF}(10)$  stays somewhat below  $\text{MF}(0)$  and its curve crosses that of the deterministic solution sporadically before distancing itself again. Finally, Figure 1 (right)



**Figure 2: Law-schools: Minimum expected value produced by  $\text{MF}(0)$  and optimal deterministic solution (left); distribution of expected value  $V(r, u)$  (for  $\alpha = 0.3$ ) (center); number of iterations (calls to the optimization oracle) vs error  $\epsilon$  (right).**

and Figure 2 (right) show the number of calls to the optimization oracle (which is also the size of the support of the ranking distribution) as a function of the additive error parameter  $\epsilon$ . Runtime is linear in the number of calls to the optimization oracle. The longest runtime of our Python implementation of MF (which occurred on the IIT-JEE dataset with  $\alpha = 0.3$  and  $\epsilon = 0.5$ ) was under one hour.

## 7 CONCLUSIONS

We introduced the problem of minimizing the amount of individual unfairness introduced when enforcing group-fairness constraints in ranking. We showed how a randomized approach ensures more individual fairness than the optimal solution to the deterministic formulation of the problem. We proved that our maxmin-fair ranking distributions provide strong fairness guarantees such as maintaining within-group meritocracy and, under a mild assumption (i.e., when we have only upper-bound constraints or when the protected attribute is binary), they have the desirable properties of being generalized Lorenz-dominant, and minimizing social inequality. Besides the technical contributions, our work shows how randomization is key in reconciling individual and group fairness. In our future work we plan to extend this intuition beyond ranking.

## ACKNOWLEDGEMENTS

The authors acknowledge support from Intesa Sanpaolo Innovation Center. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## REFERENCES

- [1] [n.d.]. European Commission Diversity Charters. [https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/tackling-discrimination/diversity-management/diversity-charters-eu-country\\_en](https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/tackling-discrimination/diversity-management/diversity-charters-eu-country_en).
- [2] [n.d.]. U.S. Equal Employment Opportunity Commission. <https://www.eeoc.gov/laws/index.cfm>.
- [3] Abolfazl Asudeh, H. V. Jagadish, Julia Stoyanovich, and Gautam Das. 2019. Designing Fair Ranking Schemes. In *SIGMOD*. 1259–1276.
- [4] Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. 2018. Equity of Attention: Amortizing Individual Fairness in Rankings. In *SIGIR*. 405–414.
- [5] Gary E Bolton, Jordi Brandts, and Axel Ockenfels. 2005. Fair procedures: Evidence from games involving lotteries. *The Economic Journal* 115, 506 (2005), 1054–1076.
- [6] L. Elisa Celis, Anay Mehrotra, and Nisheeth K. Vishnoi. 2020. Interventions for ranking in the presence of implicit bias. In *FAT\** '20. 369–380.
- [7] L. Elisa Celis, Damian Straszak, and Nisheeth K. Vishnoi. 2018. Ranking with Fairness Constraints. In *ICALP*.
- [8] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *WSDM*. 87–94.
- [9] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. 2012. Fairness through awareness. In *ITCS*. 214–226.
- [10] Cynthia Dwork, Michael P. Kim, Omer Reingold, Guy N. Rothblum, and Gal Yona. 2019. Learning from Outcomes: Evidence-Based Rankings. In *FOCS*. 106–125.
- [11] Yoav Freund and Robert E Schapire. 1999. Adaptive game playing using multiplicative weights. *Games and Economic Behavior* 29, 1-2 (1999), 79–103.

<sup>5</sup>[https://www.dropbox.com/sh/0kc17h36p632m0a/AACyO\\_PNPeBOJvPirEhQzFUDa?dl=0](https://www.dropbox.com/sh/0kc17h36p632m0a/AACyO_PNPeBOJvPirEhQzFUDa?dl=0)



- [12] David García-Soriano and Francesco Bonchi. 2020. Fair-by-design matching. *Data Mining and Knowledge Discovery* (2020), 1–45.
- [13] Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. 2019. Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search. In *KDD*. 2221–2231.
- [14] Corrado Gini. 1921. Measurement of inequality of incomes. *The Economic Journal* 31, 121 (1921), 124–126.
- [15] Martin Grötschel, László Lovász, and Alexander Schrijver. 1981. The ellipsoid method and its consequences in combinatorial optimization. *Combinatorica* 1, 2 (1981), 169–197.
- [16] Moritz Hardt, Eric Price, and Nati Srebro. 2016. Equality of Opportunity in Supervised Learning. In *NeurIPS*. 3315–3323.
- [17] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst. (TOIS)* 20, 4 (2002), 422–446.
- [18] Jovan Karamata. 1932. Sur une inégalité relative aux fonctions convexes. *Publications de l'Institut Mathématique* 1, 1 (1932), 145–147.
- [19] Michael J. Kearns, Aaron Roth, and Zhiwei Steven Wu. 2017. Meritocratic Fairness for Cross-Population Selection. In *ICML*. 1828–1836.
- [20] Peter J Lambert. 1992. The distribution and redistribution of income. In *Current issues in public sector economics*. Springer, 200–226.
- [21] Harikrishna Narasimhan, Andrew Cotter, Maya R. Gupta, and Serena Wang. 2020. Pairwise Fairness for Ranking and Regression. In *AAAI*.
- [22] John F Nash Jr. 1950. The bargaining problem. *Econometrica* (1950), 155–162.
- [23] John Rawls. 1971. *A theory of justice*. MA: Harvard University Press.
- [24] Stephen Robertson. 1977. The Probability Ranking Principle in IR. *Journal of Documentation* 33 (12 1977), 294–304.
- [25] Anthony F Shorrocks. 1983. Ranking income distributions. *Economica* 50, 197 (1983), 3–17.
- [26] Ashudeep Singh and Thorsten Joachims. 2018. Fairness of Exposure in Rankings. In *KDD*. 2219–2228.
- [27] Ashudeep Singh and Thorsten Joachims. 2019. Policy Learning for Fairness in Ranking. In *NeurIPS*. 5427–5437.
- [28] Paul D Thistle. 1989. Ranking distributions with generalized Lorenz curves. *Southern Economic Journal* (1989), 1–12.
- [29] Yining Wang, Liwei Wang, Yuanzhi Li, Di He, Wei Chen, and Tie-Yan Liu. 2013. A theoretical analysis of NDCG ranking measures. In *Proceedings of the 26th annual conference on learning theory (COLT 2013)*, Vol. 8. 6.
- [30] Ke Yang, Vasilis Gkatzelis, and Julia Stoyanovich. 2019. Balanced Ranking with Diversity Constraints. In *IJCAI*. 6035–6042.
- [31] Ke Yang and Julia Stoyanovich. 2017. Measuring Fairness in Ranked Outputs. In *SSDBM*. 22:1–22:6.
- [32] Neal E Young. 1995. Randomized rounding without solving the linear program. In *SODA*. 170–178.
- [33] Neal E Young. 2001. Sequential and parallel algorithms for mixed packing and covering. In *FOCS*. 538–546.
- [34] Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. 2017. FA\*IR: A Fair Top-k Ranking Algorithm. In *CIKM*. 1569–1578.

## A APPENDIX

We present here the proofs missing from the main text.

### A.1 Proof of Theorem 3.1

PROOF. Recall the form of our value function  $V(r, u) = f(r(u)) - g(u)$  and observe that for any distribution  $D$ ,

$$D[u] = \mathbb{E}_{r \sim D} [V(r, u)] = \mathbb{E}_{r \sim D} [f(r(u))] - g(u). \quad (15)$$

Let us write  $r \in F$  to mean that  $r$  occurs with non-zero probability in the maxmin-fair distribution  $F$ . We show that if  $u_1, u_2$  belong to the same group and  $R(u_1) \geq R(u_2)$ , then the following holds:

$$f(r(u_2)) > f(r(u_1)) \text{ for some } r \in F \implies F[u_1] \geq F[u_2], \quad (16)$$

$$\mathbb{E}_{r \in F} [f(r(u_1))] \geq \mathbb{E}_{r \in F} [f(r(u_2))]. \quad (17)$$

Suppose by contradiction that (16) fails, so  $f(r(u_2)) > f(r(u_1))$  but  $F[u_1] < F[u_2]$  for some  $r \in F$ . Let  $\hat{r}$  denote a ranking which is identical to  $r$  except that the positions of  $u_1$  and  $u_2$  are swapped. As  $u_1$  and  $u_2$  belong to the same group, swapping their positions will not affect the group-fairness constraints, so  $\hat{r}$  is a valid ranking too. Consider a distribution  $D$  over valid rankings  $\mathcal{S}$  obtained by

drawing  $s$  from  $D$  and returning  $s$  if  $s \neq r$  and  $\hat{r}$  if  $s = r$ . We have  $\mathbb{E}_{s \in D} [f(s(u_1))] - \mathbb{E}_{s \in F} [f(s(u_1))] = \Pr_{s \in F} [s = r] \cdot (f(r(u_2)) - f(r(u_1))) > 0$ , so  $\mathbb{E}_{s \in D} [f(s(u_1))] > \mathbb{E}_{s \in F} [f(s(u_1))]$  and therefore, by (15),  $D[u_1] > F[u_1]$ . Moreover  $\forall v \in \mathcal{U} \setminus \{u_1, u_2\}$  it holds that  $F[v] = D[v]$ . Therefore  $D$  is a distribution improving the expected satisfaction of  $u_1$  w.r.t.  $F$  and such that no  $v \in \mathcal{U}$  exists such that  $F[v] \leq F[u_1]$  and  $D[v] < F[v]$ , thus contradicting the assumption that  $F$  is maxmin-fair. This proves (16).

To prove (17), consider first the case  $F[u_1] \geq F[u_2]$ . Since  $R(u_1) \geq R(u_2)$  implies  $g(u_1) \geq g(u_2)$ , in this case substituting  $D = F$  in (15) we trivially obtain (17). If instead  $F[u_1] < F[u_2]$  then, by (16), we conclude that  $f(r(u_1)) \geq f(r(u_2))$  for all  $r \in F$ , which implies (17), as we wished to show.  $\square$

### A.2 Proof of Theorem 3.2

In this subsection we consider the case where we only have upper bounds in the group-fairness constraints.

First we need a result characterizing the minimum expected satisfaction of a maxmin-fair distribution. It has been inspired by the proof of [12][Theorem 15]. While [12] only considers matroid problems (which do not cover constrained ranking), our key insight is that this type of argument can be generalized *whenever there is a weight optimization oracle depending only on the weight order* (as opposed to the numerical values of the weights). This is true of the greedy algorithm from Corollary 4.2 (Algorithm 2).

LEMMA A.1. *Let  $\lambda : 2^{\mathcal{U}} \rightarrow \mathbb{R}$ . There is a distribution of valid rankings such that  $D[u] \geq \lambda_u$  if and only if*

$$\max_{S \in \mathcal{S}} \sum_{u \in X} A(S, u) \geq \sum_{u \in X} \lambda_u \quad \text{for all } X \subseteq \mathcal{U}. \quad (18)$$

**Proof of Lemma A.1.** Given a set  $E$ , let  $\Delta(E) = \{x : E \rightarrow \mathbb{R}^{\geq 0} \mid \sum_{e \in E} x_e = 1\}$  denote the set of distributions over  $E$ . Consider the following two-player zero-sum game: Player 1 (the maximizer) chooses a distribution of solutions  $p \in \Delta(\mathcal{S})$ , Player 2 (the minimizer) chooses a distribution of users  $w \in \Delta(\mathcal{U})$ , and the payoff for Player 1 when she plays  $S \in \mathcal{S}$  and Player 2 plays  $u \in \mathcal{U}$  is  $A(S, u) - \lambda_u$ . The value of this game is

$$v = \max_{p \in \Delta(\mathcal{S})} \min_{w \in \Delta(\mathcal{U})} \sum_{S \in \mathcal{S}} p_S (A(S, u) - \lambda_u);$$

the required distribution exists when  $v \geq 0$ . By Von Neumann's minimax theorem we have

$$v = \min_{w \in \Delta(\mathcal{U})} \max_{S \in \mathcal{S}} \sum_{u \in \mathcal{U}} w_u (A(S, u) - \lambda_u). \quad (19)$$

Thus,  $v \geq 0$  exactly when for all  $w \in \Delta(\mathcal{U})$ , it holds that

$$\max_{S \in \mathcal{S}} \sum_{u \in \mathcal{U}} w_u A(S, u) \geq \sum_{u \in \mathcal{U}} w_u \lambda_u. \quad (20)$$

The result will follow if we can show that the minimization problem (19) has an optimal solution of the form

$$w_u = \begin{cases} \frac{1}{|X|}, & \text{if } u \in X \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

for some non-empty  $X \subseteq \mathcal{U}$ , because for  $w_u$  of the form (21), (20) simplifies to (18) on multiplication by  $|X|$ . We have seen in Corollary 4.2 that for each  $w$ ,  $\max_{S \in \mathcal{S}} \sum_{u \in \mathcal{U}} w_u (A(S, u) - \lambda_u)$  can

be optimized by an oracle (Algorithm 2) that only depends on the order determined by  $w$  (observe that subtracting  $\lambda_u$  from  $A(S, u)$  amounts to adding  $\lambda$  to the function  $g$  in the definition of  $A(S, u)$ ). In other words, for any bijection  $\pi : [n] \rightarrow \mathcal{U}$  and any weight  $w \geq 0$  compatible with  $\pi$  (i.e., satisfying  $w_{\pi(1)} \geq w_{\pi(2)} \geq \dots w_{\pi(n)}$ ), we have

$$\max_{S \in \mathcal{S}} \sum_{u \in \mathcal{U}} w_u(A(S, u) - \lambda_u) = \sum_{u \in \mathcal{U}} w_u(A(G(\pi), u) - \lambda_u),$$

where  $G(\pi)$  is the solution returned by the greedy weighted optimization oracle.

Fix an order  $\pi : [n] \rightarrow \mathcal{U}$  and let  $B_u = A(G(\pi), u) - \lambda_u$ . Consider the minimization problem

$$\min \left\{ \sum_{u \in \mathcal{U}} w_u B_u \mid w \in \Delta(\mathcal{U}), w \text{ compatible with } \pi \right\}. \quad (22)$$

Let  $d_n = w_{\pi(n)}$  and  $d_i = w_{\pi(i)} - w_{\pi(i+1)}$ . The compatibility conditions for  $w$  may be rewritten as  $d_i \geq 0$  for all  $i$ , and the distributional constraint  $\sum_i w_i = 1$  becomes  $\sum_i i \cdot d_i = 1$ . If we write  $z_i = \sum_{j \leq i} B_{\pi(j)}$ , then (22) becomes

$$\min \left\{ \sum_{i \in [n]} d_i \cdot z_i \mid d_i \geq 0, \sum_{i \in [n]} i d_i = 1 \right\} = \min \left\{ \frac{z_t}{t} \mid t \in [n] \right\}; \quad (23)$$

the last equality is easily seen to hold because  $z_i \leq \lambda \cdot i$  for all  $i$  implies  $\sum_i d_i z_i \leq \lambda \sum_i i \cdot d_i = \lambda$ . Therefore for each  $\pi$ , an optimal solution to (22) is of the form (21), where  $X = \{\pi(1), \dots, \pi(t)\}$ ; hence the same is also true of an optimal solution to (19).  $\square$

**COROLLARY A.1.** *The minimum expected satisfaction in any maxmin-fair distribution of valid rankings is  $\min_{\emptyset \neq X \subseteq \mathcal{U}} \frac{\max_{S \in \mathcal{S}} \sum_{u \in \mathcal{U}} A(S, u)}{|X|}$ .*

We also need the following technical lemma concerning the behaviour of the expression in Corollary (A.1).

**LEMMA A.2.** *The following function  $H : 2^{\mathcal{U}} \rightarrow \mathbb{R}$  is submodular:*

$$H(E) = \max_{S \in \mathcal{S}} \sum_{u \in E} A(S, u). \quad (24)$$

**PROOF.** Let  $J(E) = \max_{S \in \mathcal{S}} \sum_{u \in E} f(S(u))$ . Then  $H(E) = J(E) - \sum_{u \in E} g(u)$ , i.e.,  $H$  is the difference between  $J$  and a modular function. So it suffices to show that  $J$  is submodular; let us fix  $X \subseteq Y$  and  $Z \subseteq \mathcal{U} \setminus Y$ . Recall from corollary 4.2 that  $\sum_w w_u f(S(u))$  is maximized by a greedy algorithm. By setting  $w_u = 1$  for  $u \in E$  and  $w_u = 0$  elsewhere, it can be used to compute  $J(E)$  for any  $E$ ; let us denote by  $r_E$  the ranking returned. Whenever we have two equal weights  $w_u = w_v$ , we can break ties in Algorithm 2 in favor of  $X$ , followed by  $Y \setminus X$ ,  $Z$ , and  $\mathcal{U} \setminus (Y \cup Z)$ . Then the greedy algorithm to maximize  $f(Y \cup Z)$  attempts to place the elements of  $X$  in top positions whenever possible, then elements of  $Y$ , and then elements of  $Z$ . This ensures that in  $r_X$  and  $r_{X \cup Z}$  the position of the elements of  $X$  is the same, allowing us to simplify the marginal gains:

$$\begin{aligned} J(X \cup Z) - J(X) &= \sum_{u \in X \cup Z} f(r_{X \cup Z}(u)) - \sum_{u \in X} f(r_X(u)) \\ &= \sum_{u \in Z} f(r_{X \cup Z}(u)). \end{aligned}$$

Similarly,

$$\begin{aligned} J(Y \cup Z) - J(Y) &= \sum_{u \in Y \cup Z} f(r_{Y \cup Z}(u)) - \sum_{u \in Y} f(r_Y(u)) \\ &= \sum_{u \in Z} f(r_{Y \cup Z}(u)). \end{aligned}$$

Moreover, for any  $x \in Z$ ,  $r_{X \cup Z}(x) \leq r_{Y \cup Z}(x)$  by the greedy rule in Corollary 4.2 and our tie-breaking rule. Therefore  $f(r_{X \cup Z}(x)) \geq f(r_{Y \cup Z}(x))$ , which implies  $J(X \cup Z) - J(X) \geq J(Y \cup Z) - J(Y)$ , as desired.  $\square$

The following is an analog of the “fair decompositions” of [12]:

**LEMMA A.3.** *Define a sequence of sets  $B_1, B_2, \dots, B_k$  iteratively by:  $B_i$  is a maximal non-empty set  $X \subseteq \mathcal{U} \setminus S_{i-1}$  minimizing*

$$\frac{H(X \cup S_{i-1}) - H(S_{i-1})}{|X|},$$

$$\text{where } H \text{ is given by (24) and } S_i = \bigcup_{j \leq i} B_j. \quad (25)$$

*We stop when  $S_i = L$  (i.e.,  $k$  is the first such  $i$ ); this will eventually occur as the sequence  $(S_i)$  is strictly increasing. Then for every  $i \in [k]$ , the following hold:*

- (a) *The expected satisfaction of any  $u \in B_i$  in any maxmin-fair distribution  $F$  is  $F[u] = \frac{H(B_i)}{|B_i|}$ .*
- (b) *For all  $u \in S_i, v \notin S_i$ , we have  $F[u] < F[v]$ .*
- (c) *For any  $D \in \Delta(\mathcal{S})$  and  $m \leq |\mathcal{U}|$ , it holds that  $\sum_{j=1}^m D_{(j)} \leq \sum_{j=1}^m F_{(j)}$ .*

**PROOF.** Let  $\lambda_j = \frac{H(B_j)}{|B_j|}$  for all  $j \in [k]$ . Notice that, using (25),  $H(S_i) = H(S_{i-1} \cup B_i) = \lambda_i |B_i| + H(S_{i-1})$  holds for all  $i$ , hence

$$H(S_i) = \sum_{j=1}^i \lambda_j |B_j|. \quad (26)$$

Notice also that the definition of  $H$  trivially implies that for any distribution  $D$  we must have

$$H(X) \geq \sum_{u \in X} D[u]. \quad (27)$$

We can give an alternative definition of  $S_i$  as

$$S_i \text{ is a maximal set } X \not\subseteq S_{i-1} \text{ minimizing } \frac{H(X) - H(X \cap S_{i-1})}{|X \setminus S_{i-1}|}. \quad (28)$$

Indeed, for any fixed difference  $Y = X \setminus S_{i-1}$ , the submodularity of  $H$  (Lemma A.2) implies that the minimum of the numerator in (28) is attained for a set  $X$  which properly contains  $S_{i-1}$ . In particular we have that for any  $X \subseteq \mathcal{U}$  and  $j \in [k]$ ,  $H(X) - H(X \cap S_{j-1}) \geq \lambda_j |X \setminus S_{j-1}|$ , and by replacing  $X$  with  $X \cap S_j$  above we also get  $H(X \cap S_j) - H(X \cap S_{j-1}) \geq \lambda_j |X \cap S_j|$ , implying

$$H(X) \geq \lambda_i |X \setminus S_{i-1}| + \sum_{j < i} \lambda_j |X \cap S_j| \quad \forall i \in [k]. \quad (29)$$

We show that properties (a) and (b) hold for all  $i \leq t \leq k$ , reasoning by induction on  $t$ . There is nothing to show when  $t = 0$  or  $S_{i-1} = \mathcal{U}$  (in the latter case,  $k < i$ ), so assume that  $t \geq 1$  and the claims hold for all  $i < t$ ; we show they also hold for  $i = t$ .

From property (a) in the induction hypothesis, we know that in the maxmin-fair distribution,  $F[u] = \lambda_j \forall u \in S_j, j \leq i$ . We can use Lemma A.1 to determine the minimum expected satisfaction of  $F$  outside  $S_{i-1}$ ; we conclude, by (29), that  $\min_{u \notin S_{i-1}} F[u] \geq \lambda_i$ . As (26) shows, equality in (29) is attained when  $X = S_i$ , thus by (27) we must in fact have  $\min_{u \notin S_{i-1}} F[u] = \lambda_i$  and  $F[u] = \lambda_i$  for all  $u \in B_i$ , proving (a).

To prove (b), we need to show the strict inequality  $\lambda_{i-1} < \lambda_i$ . By Lemma (A.2), the function  $J(X) = H(X \cup S_{i-1}) - H(S_{i-1})$  is submodular. A consequence of this is that, if  $X, Y$  are non-empty sets minimizing  $J(X)/|X|$ , then  $X \cup Y$  also minimizes  $J(X)/|X|$ . Indeed, suppose  $\frac{J(Y)}{|Y|} = \frac{J(X)}{|X|} \triangleq \lambda$ . By the submodularity of  $J$ ,

$$J(X \cup Y) + J(X \cap Y) \leq J(X) + J(Y) = \lambda(|X| + |Y|).$$

Notice that  $J(X \cup Y) \geq \lambda|X \cup Y|$  and  $J(X \cap Y) \geq \lambda|X \cap Y|$  by definition. If any of these two inequalities were strict we would have the contradiction

$$J(X \cup Y) + J(X \cap Y) > \lambda(|X \cup Y| + |X \cap Y|) = \lambda(|X| + |Y|).$$

Hence these inequalities are not strict, and  $J(X \cup Y) = \lambda|X \cup Y|$ .

Now, due to the maximality of  $B_i$  as defined by (25), the set  $B_i$  is the union of all non-empty sets  $X$  minimizing  $J(X)/|X|$ . This means that, when  $t > 1$ , the strict inequality  $\lambda_t > \lambda_{t-1}$  holds (otherwise  $B_{t-1}$  would not be maximal), which by (a) implies (b).

Finally we show (c). We argue by contradiction. Pick a counterexample with minimum  $m$ ; then  $m \geq 1$ . Let  $i$  be such that  $|S_{i-1}| < m \leq |S_i|$ . Then we have  $\sum_{j=1}^{m-1} D_{(j)} \leq \sum_{j=1}^{m-1} F_{(j)}$  and  $\sum_{j=1}^m D_{(j)} > \sum_{j=1}^m F_{(j)}$ , thus  $D_{(m)} > F_{(m)} = \lambda_i$  by properties (a) and (b). Now let  $X$  be the individuals with the  $m$  smallest satisfactions in  $D$ . It follows that

$$\begin{aligned} H(S_i) &\geq \sum_{u \in S_i} D[u] \geq \sum_{u \in X} D[u] + (|S_i| - m)D_{(m)} \\ &= \sum_{j=1}^m D_{(j)} + (|S_i| - m)D_{(m)} > \sum_{j=1}^m F_{(j)} + (|S_i| - m)\lambda_i \\ &= \sum_{j \leq i} \lambda_j |B_j| = H(S_i). \end{aligned}$$

This contradiction completes the proof.  $\square$

**Proof of Theorem 3.2.** Property (c) of Lemma A.3 asserts that the maxmin-fair distribution  $F$  is generalized Lorenz-dominant.  $\square$

### A.3 Proof of Theorem 5.1

PROOF. Sort  $\mathcal{U} = \{u_1, \dots, u_n\}$  by increasing order of  $g$  so that

$$g(u_1) \geq g(u_2) \geq \dots \geq g(u_n), \quad (30)$$

and let us identify  $\mathcal{U}$  with the set  $[n]$  for ease of notation, so that  $u_i = i$ . Recall that the positions  $[n]$  are sorted by decreasing order of  $f$  so that

$$f(1) \geq f(2) \geq \dots \geq f(n). \quad (31)$$

Let  $M > 0$  be a large enough number and define  $W_{iu} = -e^{-M(f(i)-g(u))}$ . Observe that, because of the orderings defined by (30) and (31), the matrix  $W$  satisfies the Monge property: if  $i < j$  and  $u < v$ , then  $W_{iu} + W_{jv} \geq W_{iv} + W_{ju}$ . Indeed,

$$W_{iu} + W_{jv} - (W_{iv} + W_{ju}) = (e^{-Mf(j)} - e^{-Mf(i)})(e^{Mg(u)} - e^{Mg(v)}) \geq 0$$

because  $g(u) \geq g(v)$  and  $f(i) \geq f(j)$ , so both factors are non-negative. Thus we may apply the algorithm from [7] to maximize  $\sum_u W_{r(u),u}$  over valid rankings  $r$ . The resulting algorithm is Algorithm 3. For any fixed  $M$ , maximizing  $\sum_u W_{r(u),u}$  is the same as maximizing  $(1/M) \cdot \text{softmax}\{M \cdot A(S, u) \mid u \in \mathcal{U}\}$ . But since the solution  $S^* = S^*(M)$  returned by this algorithm does *not* depend on  $M > 0$  and  $\lim_{M \rightarrow \infty} \frac{\text{softmax}(M \cdot z)}{M} = \min(z)$ , it follows that  $S^*$  maximizes  $\min\{A(S, u) \mid u \in \mathcal{U}\}$ , as we wished to show.  $\square$

### A.4 Solving maxmin-fairness approximately

Instead of solving the LPs in the proof of Theorem 4.1 exactly, we can use iterative methods designed to approximately solve zero-sum games and packing/covering programs, as sketched next.

Recall that the exact algorithm 1 works by solving the linear program (9) and updating  $K$  and  $\alpha_v$ . Let us apply a positive affine transformation to normalize all  $A(S, v)$  to the range  $[0, 1]$  and select an additive approximation parameter  $\epsilon > 0$ , so we want to ensure that in the final solution, the expected satisfaction for every  $v \in \mathcal{U}$  is at most an additive  $\epsilon$  below that which would have been computed by solving LP (9) exactly at the point where  $\alpha_u$  was assigned.

Rather than maximizing  $\lambda$  directly in (9), we can guess an approximation  $\tilde{\lambda}$  to the optimum, and verify if the guess is correct by eliminating the  $\lambda$  variable from this LP and replacing it with our guess  $\tilde{\lambda}$ , and then checking if the program is feasible. Denote by  $M$  the resulting LP.  $M$  is a fractional covering program, equivalent to a zero-sum game, hence the techniques from [32] apply. If  $M$  is feasible, the algorithm from [32] returns a non-negative solution with  $\sum_s p_s = 1$  and violating the remaining constraints by at most an additive term  $\delta$ , using  $O(\log n / \delta^2)$  calls to the separation oracle. This solution is sparse, having  $O(\log n / \delta^2)$  non-zero coefficients. By performing binary search on  $\tilde{\lambda}$ , we can solve (9) up to a  $\delta/2$  term in the satisfaction probabilities by approximately solving  $O(\log(1/\delta))$  packing problems. Then we augment  $K$  by adding those users where the satisfaction probabilities in the approximate solution increased by at most  $\delta/2$ ; then we know that in the exact solution, it was impossible to increase then by more than  $\delta$ . This process, however, may decrease the satisfaction probabilities of individuals already in  $K$  by up to  $\delta/2$ . If we have solved  $m$  different LPs by the time we reach  $K = \mathcal{U}$ , the total cumulative error in the satisfaction probabilities in the final solution and the optimum values of the program where they were computed is at most  $\delta m$ . Since  $m \leq O(n \log(1/\delta))$ , we can take  $\delta = O(\epsilon / (n \log(n/\epsilon)))$  to guarantee that  $\delta m \leq \epsilon$ .

Several improvements over this basic scheme can be made. First, the above bound for  $m$  is often too pessimistic, and it is more efficient to do a “doubling trick”: start with  $m = 2$  and keep doubling  $m$  and restarting again with  $K = \emptyset$  if during the execution of the algorithm sketched above we end up solving more than  $m$  programs. Second, we can use the variable-step increase technique from [33]. Third, in the case of ranking problems with upper bounds, the separation oracle only depends on the order of the weights and not their specific values, so there is no need to call it again if this order does not change; we can simply increase the probability of that solution. Finally, for a given order of weights, (23) allows us to obtain an optimal dual solution that respects that given weight order, which can be used to detect convergence of the iterative algorithm from [33] earlier.