

Leximax Approximations and Representative Cohort Selection

Monika Henzinger  

Universität Wien, Austria

Charlotte Peale  

Stanford University, CA, USA

Omer Reingold  

Stanford University, CA, USA

Judy Hanwen Shen  

Stanford University, CA, USA

Abstract

Finding a representative cohort from a broad pool of candidates is a goal that arises in many contexts such as choosing governing committees and consumer panels. While there are many ways to define the degree to which a cohort represents a population, a very appealing solution concept is lexicographic maximality (leximax) which offers a natural (pareto-optimal like) interpretation that the utility of no population can be increased without decreasing the utility of a population that is already worse off. However, finding a leximax solution can be highly dependent on small variations in the utility of certain groups. In this work, we explore new notions of approximate leximax solutions with three distinct motivations: better algorithmic efficiency, exploiting significant utility improvements, and robustness to noise. Among other definitional contributions, we give a new notion of an approximate leximax that satisfies a similarly appealing semantic interpretation and relate it to algorithmically-feasible approximate leximax notions. When group utilities are linear over cohort candidates, we give an efficient polynomial-time algorithm for finding a leximax distribution over cohort candidates in the exact as well as in the approximate setting. Furthermore, we show that finding an integer solution to leximax cohort selection with linear utilities is NP-Hard.

2012 ACM Subject Classification Theory of computation → Theory and algorithms for application domains

Keywords and phrases fairness, cohort selection, leximin, maxmin

Digital Object Identifier 10.4230/LIPIcs.FORC.2022.2

Related Version *Full Version:* <https://arxiv.org/abs/2205.01157>

Funding *Monika Henzinger:* This work was done in part as Stanford University Distinguished Visiting Austrian Chair. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 101019564 “The Design of Modern Fully Dynamic Data Structures (MoDynStruct)”) and from the Austrian Science Fund (FWF) project “Fast Algorithms for a Reactive Network Layer (ReactNet)”, P 33775-N, with additional funding from the *netidee SCIENCE Stiftung*, 2020–2024. *Charlotte Peale:* Supported by the Simons Foundation Collaboration on the Theory of Algorithmic Fairness.

Omer Reingold: Supported by the Simons Foundation Collaboration on the Theory of Algorithmic Fairness, the Sloan Foundation Grant 2020-13941 and the Simons Foundation investigators award 689988.

Judy Hanwen Shen: Supported by the Simons Foundation Collaboration on the Theory of Algorithmic Fairness.



© Monika Henzinger, Charlotte Peale, Omer Reingold, and Judy Hanwen Shen; licensed under Creative Commons License CC-BY 4.0

3rd Symposium on Foundations of Responsible Computing (FORC 2022).

Editor: L. Elisa Celis; Article No. 2; pp. 2:1–2:22



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



1 Introduction

In many fairness-related settings, we seek to select an outcome that does not disproportionately harm any key subgroup. Speaking in terms of group utilities, a fair solution would ideally provide every key subgroup with high utility. Unfortunately, such a goal may be impossible to achieve if the utilities derived by subgroups from any potential solutions are in opposition. Moreover, other goals such as seeking to equalize utilities across groups may artificially constrain the utility of certain groups in order to match some group with uniformly low utility.

The classic *maximin* objective, which seeks to output solutions that maximize the utility of the worst-off group, has been widely studied as a goal that can circumvent these potential pitfalls by seeking to achieve the best possible outcome for the worst-off group. This results in a set of solutions that optimize the outcome for the worst-off group, but may still vary quite a bit with respect to the second-worst-off group, third-worst-off group, etc. *Lexicographically maximal* solutions strengthen the maximin objective by requiring that the utility of the second-worst-off group be maximized subject to the worst-off-group achieving its maximin value, the third-worst-off group be maximized subject to the worst-off and second-worst-off values, and so on. This goal intuitively tells us that a lexicographically maximal solution gives the best-possible utility guarantee we can give for each group without harming another group.

Lexicographic maximality (which we refer to as *leximax*, but is sometimes referred to in the literature as *leximin*) has been widely studied in the context of allocations [14, 16, 19]. Recently, Diana, Gill, Globus-Harris, Kearns, Roth, and Sharifi-Malvajerdi [10] explored applying the objective to the contemporary fairness context of loss minimization. In this paper, motivated by the goal of selecting a representative cohort from a group of candidates, we generalize the approach of [10] to the goal of selecting a solution that achieves lexicographically maximal utilities for a set of key subgroups.

Our contributions fall into two main categories: definitional, where we explore useful variants of the leximax objective and their relations in the general setting of selecting a leximax solution from a set of potential solutions, and algorithmic, in which we investigate how to efficiently find exact leximax solutions as well as different variants in the specific context of selecting representative cohorts. We provide an overview of definitional contributions in Section 1.1, followed by an overview of the cohort selection context and resulting algorithms in Section 1.2.

1.1 Approximations of Lexicographically Maximal Solutions

Diana et al. [10] define an approximate notion of lexicographic maximality for which they construct oracle-efficient algorithms. Their notion is influenced by an algorithmic approach to calculating leximax solutions in that it assumes the maximal values of the worst-off group, second-worst-off group, etc. are calculated recursively based on whatever estimates came before. The definition assumes some small amount of error when calculating the maximin utility value, and then considers how this error would propagate to the second-worst-off-group's maximum value, then considers how additional errors around the second-worst-off-group's maximum value together with errors from the worst-off group maximum value might propagate to the third-worst-off group, and so on.

One of the appealing aspects of leximax solutions is that they offer a simple semantic interpretation that explains the sort of fairness guarantees such solutions provide: given a leximax solution, any alternative solution that improves the utility of some group must also decrease the utility of some worse-off group (Proposition 4). While the approximation

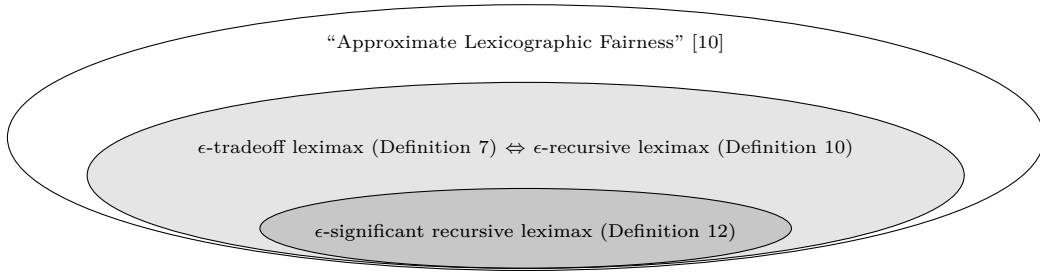
notion presented by [10] is very natural, they also show that such approximate solutions may greatly diverge from exact solutions (see Example 5 for details), meaning that they may also diverge from this appealing semantic interpretation.

Ideally, we'd like a well-defined notion of approximation that extends the semantic interpretation of leximax and relates to the algorithmically achievable notion presented by [10]. However, we find that such a definition is somewhat difficult to pin down. *Many natural relaxations of the semantic definition result in notions of approximations where either no solutions are guaranteed to satisfy the notion or the notions themselves may not imply a meaningful fairness guarantee that is analogue to that offered by leximax solutions.* Developing a meaningful notion of approximation is exactly the challenge that this paper addresses.

We provide a relaxation of the semantic definition that we term ϵ -tradeoff leximax (Definition 7) that is always guaranteed to exist, and while it is not equivalent to the notion presented in [10], in Theorem 11, we show that it is equivalent to a stronger variant of their definition that we call ϵ -recursive leximax (Definition 10). The algorithms of [10] have the potential to slightly mis-estimate the maxmin values for different groups, and therefore are only guaranteed to output approximate leximax solutions. The type of mis-estimations that may arise are actually more constrained than the full class of errors their weaker notion of approximation allows for. In particular, solutions outputted by their algorithms actually satisfy our stronger notion of ϵ -recursive leximax.

Past explorations of lexicographic maximality have mostly concentrated on finding *exact* leximax solutions. In the design of algorithms, approximations are usually viewed as alternative solutions that are “almost as good” as the exact solution and that are computed in settings where it is difficult to efficiently find exact solutions. In this paper we suggest that in some cases, we may prefer to consider an approximate notion of lexicographic maximality rather than its exact counterpart. In particular, exact leximax solutions may be highly dependent on small variations in the utility of less-well-off groups. For example, a solution where all groups receive 0.01 utility would be preferred by the exact leximax objective over a solution where one group receives 0 utility and all others receive a utility of 1, even though this second solution gets much higher utility for the majority of groups while only decreasing the utility of a single group by a tiny amount. We explore well-defined ways where approximation can benefit stakeholders and suggest a notion of approximation that is stronger than the ϵ -recursive leximax notion mentioned above that we term ϵ -significant recursive leximax approximation (Definition 12) that identifies solutions that ignore tiny variations in utility and identifies only solutions that are leximax due to significant increases in utility. In Theorem 14, we give a more formal characterization of the benefits drawn from considering ϵ -significant recursive leximax solutions rather than just any ϵ -recursive leximax solution.

A third motivator for our study of leximax approximations is how robust leximax solutions may be to small amounts of noise in the estimates of group utility. We show that when calculated in a noisy setting, our relaxed semantic notion (ϵ -tradeoff leximax) is not guaranteed to still be ϵ -tradeoff leximax, however it is guaranteed to satisfy the weaker notion of approximation defined in [10]. On the other hand, in Lemma 18 we show that we can define a stronger variant of the semantic notion that guarantees a solution will be ϵ -tradeoff leximax in the noisy setting, but it has the disadvantage that such solutions may not always exist. We also examine noise in the context of ϵ -significant recursive leximax solutions, and show that when such solutions are calculated in the presence of noise, they are somewhat robust to noise as they imply a slightly weakened variant of ϵ -significance (Lemma 19).



■ **Figure 1** Relations between the different notions of leximax approximation discussed here and in [10].

Figure 1 summarizes the various notions of approximate lexicographical maximality and how they relate to one another. All of our approximate notions are defined with respect to an arbitrary class of solutions from which we’d like to pick a leximax solution. This allows our new definitions to be applied in the deterministic setting, where each solution would represent a particular cohort, or a randomized setting, where each solution corresponds to a distribution over cohorts and utilities are given in expectation.

1.2 Algorithms for Leximax Cohort Selection

In data selection, recruiting, and civic participation settings where a representative cohort is desired, the goal of representation is juxtaposed with the constraint of selecting a small representative set. There can be tension between selecting a cohort small enough for the resources available but large enough to represent as much of the population as possible. A lexicographically maximal solution is particularly salient in a representative cohort problem because it guarantees inclusion for the worst-off-groups while optimizing for the utility of all groups. We consider a model where how well each group or individual in the population is represented by a cohort candidate is given by a utility function. While there are many different ways a cohort or committee in power might make decisions or influence outcomes, we consider a linear setting where the utility a group derives from a cohort is the sum of utilities derived from each member of the cohort. Approximate notions of lexicographical maximality are of particular interest in this setting since estimating utilities that describe representativeness is difficult and might be noisy in practice.

Diana et al. [10] give a convex formulation of approximate lexicographical fairness and an oracle-efficient algorithm to solve general leximax convex programs. For our cohort selection setting specifically, we leverage the linearity across decision variables to find a polynomial time algorithm (Algorithm 1) that can calculate both exact leximax solutions as well as the two approximate variants we consider, ϵ -tradeoff leximax and ϵ -significant recursive leximax (with no external oracle needed for the calculation).

The linearity of utilities across cohort members and the recursive definition of leximax gives us a sequence of linear programs where the number of variables is linear in the size of the candidate pool and the number of constraints is exponential in the number of groups. In each m -th linear program, we maximize the sum of utilities of all sized- m groups which gives us an exponential number of constraints; rendering the linear program too big to solve via generic LP solvers. We circumvent this difficulty by creating a separation oracle (Algorithm 2) which tests the sum of utilities of the m worst off groups efficiently, giving us a polynomial time algorithm overall (Lemma 16). We can use the same approach to efficiently find ϵ -tradeoff leximax or ϵ -significant recursive leximax solutions by modifying the lower bound constraints on the sum of group utilities.

The output of our algorithm allows a randomized approach for selecting a cohort of expected size k that guarantees leximax utilities in expectation. We can also round our algorithm output to a solution of size exactly k where the expected utility across groups is leximax¹. We focus on this distributional setting for our algorithms for a few key reasons:

- **Tractability.** If we wanted to instead find a deterministic cohort of exactly size k by finding a lexicographically maximal integer solution, the problem becomes hard. By showing that the problem of finding the exact integer lexicographically maximal cohort solves the NP-hard problem of Minimum Hitting Set, we show that finding a solution as well as approximating the number of groups with non-minimum utility within a factor of $(1 - 1/e) + o(1)$ is NP-Hard (Lemma 20).
- **Fair Arbitration between Solutions.** It is very possible that two lexicographically maximal deterministic solutions may provide wildly different utility values for a particular group. As an example, consider choosing between a cohort that provides maximum utility to Group A, but zero utility to Group B, and another cohort that provides zero utility to Group A but maximum utility to Group B. Both cohorts are a lexicographic maximum, however selecting a deterministic solution requires us to decide whether the solution should favor Group A or B. A distributional approach gets rid of this difficult decision because the randomized approach itself guarantees that we are providing both A and B a fair chance at high utility.

There are many different potential approaches to randomly selecting a cohort in the distributional setting. We choose to use a randomized approach to selection that includes or excludes each potential cohort member independently with probability outputted by the algorithm. Such an approach offers the following benefits:

- **Simple Sampling Procedure.** Rather than outputting an arbitrary and potentially complicated distribution over cohorts that is difficult to sample from, the output of our algorithm is a single vector of marginal selection probabilities for each potential candidate. Our approach still results in a cohort with expected size k , but provides an easy way to sample cohorts, and as discussed in the final bullet point, gives better guarantees about the utility groups can expect to receive in practice. We also describe a rounding approach that results in cohorts of size exactly k that are still leximax in expectation.
- **Better Concentration Guarantees for Some Natural Settings.** While a distributional leximin solution may give groups better utility guarantees in expectation, it comes with the caveat that individual runs of the randomized solution may still result in cohorts where groups receive utility that is far below their expected utility. In an extreme case, a distributional solution that guarantees all groups 0.5 utility might be achieved by choosing uniformly between solutions that provide maximum and zero utility. When the size of the cohort is large enough, our approach to randomized choice guarantees that groups receive utility near their expectation with high probability because we consider each cohort member independently, rather than outputting an arbitrary joint distribution over potential cohort members (Lemma 15).

¹ For further rounding details, see discussion in Section C.

1.3 Our Contributions

To summarize, we provide the following contributions:

1. Define a new semantic notion of leximax approximation that is always guaranteed to exist and show that it is equivalent to an algorithmically-inspired notion of approximation that is stronger but related to the one defined in [10].
2. Investigate stricter notions of approximation that identify significantly leximax solutions that can be achieved by ignoring small variations in utility.
3. Explore how our new notions of approximation behave in settings where the group utilities may be reported with some small amount of additive noise.
4. Provide polynomial time algorithms for computing exact and approximate leximax distributions over cohorts with linear utility functions.
5. Show that the alternative goal of computing deterministic cohorts in our setting is NP-hard, and moreover approximating the number of groups with non-minimum utility is also NP-hard.

► **Note 1.** Due to space constraints, all proofs except the proof of Lemma 16 (Appendix A) have been omitted, but can be found in the full version of this paper.

1.4 Related Work

Fair and diverse selection has become a prominent area of interest in algorithmic and machine learning fairness communities. In the setting of selecting representative data, prior works define metrics for diversity [24], and give algorithms for diverse data selection and summarization [7, 18]. For selecting individuals from a larger pool, prior works on cohort selection and multi-winner elections have studied individual guarantees of fairness [2] as well as group parity goals of diversity [5, 8, 27]. Other works have examined how bias and variance may affect different groups differently during a selection process and fairness amounts to remedying implicit bias and variance in the selection process for different groups of individuals [12, 17]. Parity or proportional diversity approaches to cohort selection assume the correct amount of representation for each subgroup is known and thus fairness can be achieving a predefined level of diversity.

When there is no “merit” function to guide a selection process, cohort selection can also be seen as a representation problem. Diversity is the goal of a central decision maker while representation is the objective of each group in the population when selecting a cohort. Instead of modeling overall welfare based on the number of representatives from each group, our work considers the welfare of each group based on how representative each cohort member is for that group. Since how well a cohort serves each group in a population cannot be summarized by a single value, a natural direction is to examine the utilities of all groups of a given cohort that has been selected from a general population. Lexicographical fairness emerges as a reasonable notion of fairness that guarantees Pareto optimality in this setting of multiple objectives or losses. Flanigan et al. [13] give an algorithm for recruiting “citizen’s assemblies” based on sampling from a distribution over representative panels that are generated from leximax selection probabilities over citizens in the population. Our work looks at selecting a representative cohort from a pool of candidates rather than the underlying population which allows a more general model where each member or group in the population has a utility vector describing its utility for each candidate that is being considered for the cohort. Furthermore, we optimize for leximax utilities for each group of interest rather than leximax sample probabilities for each individual in the population.

In telecommunication network design, min-max fairness (MMF) is an important solution concept to lexicographically maximize fractional flow for all parties [1, 25, 26]. An adjacent problem of lexicographically maximal flows where there are multiple sinks has also been studied and a polynomial time algorithm exists for finding fractional flow [22, 23]. The problem of finding a leximax routing for an unsplittable flow along a network is NP-Complete but finding a 2-approximation is possible [16]. An approximate solution here means that it is not possible to improve a group without decreasing the utility of another group that is more than a factor of 2 worse.

Lexicographically maximal solutions have also been studied in other domains including bottleneck combinatorial optimization problems [6, 9], sampling actions for repeated games [3], allocation of classrooms [19] as well as indivisible goods more generally [14]. It is important to note that unlike the leximax allocation problem, there is no limit on the number of groups gaining utility from the same candidate being included in a cohort or allocated set. Most recently, leximax empirical risk minimization for classification has also been studied [15, 20, 21].

2 The Leximax Objective

In this paper, we focus on approaches to selecting *lexicographically maximal* (or leximax) representative cohort solutions. We consider a setting in which we'd like to select a solution S from a set of potential solutions \mathcal{S} such that S is a good representation of some set of key (potentially overlapping) subgroups $\mathcal{G} = \{G_1, \dots, G_m\}$. We measure degree of representation via a utility function $u : \mathcal{S} \times \mathcal{G} \rightarrow [0, 1]$. Ideally, we'd like to select a cohort such that every subgroup is guaranteed to have high utility. However, this may be impossible to achieve in certain settings, such as when the utility functions of two groups are in opposition. Unlike maximizing total welfare, which may result in solutions that neglect the welfare of certain groups or seeking to equalize utilities across groups, which may artificially cap the utility some groups can achieve, lexicographically maximal solutions extend the goal of the classic maxmin objective by seeking to maximize the utility of the worst-off group, and then seeking to maximize the utility of the second-worst-off group subject to this worst-off group's value, etc. This results in a solution concept that seeks to give the best guarantee possible for every key group, rather than just the worst-off.

We now formally define the leximax objective.

► **Definition 2.** *Given two vectors u and v in \mathbb{R}^m , we say that u is lexicographically greater than v , or $v \preceq u$, if and only if there exists some i such that for all $j \leq i$ we have $v_j = u_j$, and either $i = m$ or $u_{i+1} > v_{i+1}$.*

Applying this definition to the set of sorted group utility vectors obtained from every possible solution gives us a total ordering on these vectors. A leximax solution is any vector that is maximal according to this ordering. In many portions of this paper, in order to reason about the contents of these sorted vectors, we will care about the utility that the i th worst-off group receives from a particular solution S . We denote this with the bracketed notation $u(S, G_{[i]})$.

► **Definition 3.** *Given a set of potential solutions \mathcal{S} and groups \mathcal{G} , we say that a solution $S \in \mathcal{S}$ is lexicographically maximal (leximax) if for any other solution S' , we have $\langle u(S', G_{[i]}) \rangle_{i=1}^m \preceq \langle u(S, G_{[i]}) \rangle_{i=1}^m$.*

Intuitively, when we seek to find a lexicographically maximal solution, we try to do the best we can for the worst-off group, and then within these potential solutions try to do the best we can for the second-worst-off group, etc. Note that under this definition, groups may

achieve varying utilities for different lexicographically maximal solutions, however the vector of sorted group utilities will be unique for any leximax solution. When the solution class is convex and compact and the utility function is continuous with respect to this class, a particular group receives the same utility under any leximax solution.

An attractive feature of lexicographically maximal solutions is that they have an equivalent definition that gives a semantic understanding of the solutions identified by the goal in Definition 3. We call this notion *tradeoff leximax*.

► **Proposition 4.** *Given a set of solutions \mathcal{S} and groups \mathcal{G} , $S \in \mathcal{S}$ is lexicographically maximal if and only if for any S' and $i \in [m]$ such that $u(S', G_{[i]}) > u(S, G_{[i]})$, there exists some $j < i$ such that $u(S, G_{[j]}) > u(S', G_{[j]})$.*

This equivalent definition of lexicographic maximality offers an appealing re-interpretation of this objective: a solution is optimal if increasing the utility of any particular group would result in decreasing the utility of a worse-off group.

3 Approximations of Leximax-Optimal Solutions

While the leximax objective's goal of doing the best we can for every group is attractive, one potential downside is that the set of leximax-optimal solutions can be incredibly sensitive to small variations in the utility received by certain groups. We consider the following example that illustrates this phenomenon:

► **Example 5 (Sensitivity of leximax-optimal solutions).** Consider a simple setting as in Figure 2 in which we have two groups, $\mathcal{G} = \{G_1, G_2\}$, and would like to decide between two potential solutions $\mathcal{S} = \{S_1, S_2\}$. The utilities for each group and each solution are defined as $u(S_i, G_j) = U_{ij}$ where $U \in [0, 1]^{S \times \mathcal{G}}$ is defined as follows:

$$U = \begin{bmatrix} 0 & 1 \\ 0.01 & 0.01 \end{bmatrix}.$$

Clearly the only leximax solution is S_2 (with sorted utility vector $(0.01, 0.01)$), because the worst-off group has value 0.01 rather than receiving 0 utility as it does in S_1 (which has a sorted utility vector of $(0, 1)$).

However, if we allow for the possibility that the utility estimates are off by even a tiny amount such as 0.01, suddenly S_1 is also a plausibly leximax solution despite having a completely different value for the second-worst-off group.

Example 5 is notable in that it demonstrates how small variations in the utilities of groups can lead to drastic changes with respect to the types of leximax solutions that are considered optimal. In settings where utilities may be reported with some estimation error, it is therefore incredibly important to consider how these errors might affect how the output optimal solution compares to the true leximax solution that would have been produced given completely accurate utilities.

Moreover, even when the utilities are believed to be accurate, it may be useful to consider solutions that are not exactly leximax, but are leximax when small variations in the utility are ignored. Example 5 is a situation where the exact leximax offers a tiny improvement in the worst-off group at the cost of a huge decrease in the utility of the second-worst-off group. A practitioner who views utility differences of less than 0.05 as insignificant might prefer S_1 as the only *significantly* lexicographically maximal solution because the worst-off groups between S_2 and S_1 receive comparable utility while the second-worst-off group is significantly better off under S_1 .

The search for plausibly exact lexicographic solutions given the potential for some amount of estimation error as well as the need for significantly maximal lexicographic solutions even when working with exact utility values motivates our study of new approximate leximax notions. In this section, we introduce two such notions: first, we introduce a semantic notion of approximate leximax that relaxes the standard leximax definition to consider additional solutions that may be plausibly leximax. The second notion we introduce here seeks solutions that are leximax if only “significant” improvements are considered (as in the discussion above). Unlike the first notion, the notion of significantly leximax solutions is not a strict relaxation of leximax and may not include the exact leximax solution in some cases.

3.1 Relaxations of the Leximax Objective

3.1.1 Elementwise Approximation

The most naive approach to approximation would be to require that the element-wise distance between the sorted utility vectors of the true lexicographically maximal solution and the approximate solution be small:

► **Definition 6** (Element-wise leximax approximation). *Given a set of m groups \mathcal{G} and a set of potential solutions \mathcal{S} , let ℓ be the sorted vector of utilities attained by any leximax solution. We say that a solution $S \in \mathcal{S}$ is an α -element-wise leximax approximation iff $\max_{i \in [m]} \{\ell_i - u(S, G_{[i]})\} \leq \alpha$.*

While attractive in its simplicity, [10] observe that in certain contexts, such a definition may be stricter than we can hope for. In particular, if the leximax solution is being computed recursively, small estimation errors in the values of the worst-off group’s utility can greatly effect the difference between the utility of better-off groups in a lexicographically maximal solution compared to a solution that maximizes group utilities based off of this incorrect value. Thus, we turn our attention to weaker notions of approximation.

3.1.2 Tradeoff Approximation

We introduce a new notion of approximation that is a natural relaxation of the semantic interpretation of leximax solutions provided by the *tradeoff leximax* objective discussed in Proposition 4.

► **Definition 7** (ϵ -tradeoff leximax). *Given a set of m groups, \mathcal{G} , and a set of potential solutions, \mathcal{S} , a solution $S \in \mathcal{S}$ is ϵ -tradeoff leximax if for any S' and i such that $u(S, G_{[i]}) < u(S', G_{[i]}) - \epsilon$, there exists a $j < i$ such that $u(S, G_{[j]}) > u(S', G_{[j]})$.*

Intuitively, this definition guarantees that if we can find some other solution that does a lot better on some particular group, then this new solution must also decrease the utility of some worse-off group.

ϵ -tradeoff leximax provides an appealingly simple relaxation of the semantic interpretation of exact leximax solutions. However, slight variations of this definition, also natural relaxations of leximax, will result in definitions where solutions are not guaranteed to exist. We explore this in the following example:

► **Example 8** (Altered versions of ϵ -tradeoff leximax may not have any solutions.). We define a class of alternative tradeoff definitions that we term (ϵ_1, ϵ_2) -significant tradeoff leximax for reasons that will become clear in Section 3.2 as follows:

► **Definition 9** ((ϵ_1, ϵ_2) -significant tradeoff leximax). *Given a set of m groups, \mathcal{G} , and a set of potential solutions, \mathcal{S} , a solution $S \in \mathcal{S}$ is (ϵ_1, ϵ_2) -significant tradeoff leximax for any $\epsilon_1, \epsilon_2 \geq 0$ if for any S' and i such that $u(S, G_{[i]}) < u(S', G_{[i]}) - \epsilon_1$, there exists a $j < i$ such that $u(S, G_{[j]}) > u(S', G_{[j]}) + \epsilon_2$.*

When $\epsilon_1 = \epsilon$ and $\epsilon_2 = 0$, this notion is equivalent to ϵ -tradeoff leximax. When $\epsilon_2 > 0$, the definition requires that any increase by more than ϵ_1 result in a decrease of more than ϵ_2 in a worse-off group.

However, we demonstrate that for $\epsilon_1, \epsilon_2 > 0$, no solution may exist. Consider $\epsilon_1 = \epsilon_2 = \epsilon$ for the setting depicted in Figure 3 where we have two groups and four potential solutions with utilities defined as $u(S_i, G_j) = U_{ij}$, for

$$U = \begin{bmatrix} 0 & 0.5 + 6\epsilon \\ \epsilon/2 & 0.5 + 4\epsilon \\ \epsilon & 0.5 + 2\epsilon \\ 3\epsilon/2 & 0.5 \end{bmatrix}.$$

Where we assume ϵ is sufficiently smaller than 0.5. Under these utilities, S_4 cannot be (ϵ, ϵ) -significant tradeoff leximax because S_3 improves by more than ϵ in G_2 while only decreasing G_1 by $\epsilon/2$. Similarly, S_3 and S_2 cannot be (ϵ, ϵ) -significant tradeoff leximax due to the existence of S_2 and S_1 , respectively. This means that S_2, S_3, S_4 all cannot be (ϵ, ϵ) -significant tradeoff leximax. However, we see that S_4 improves by more than ϵ over S_1 in G_1 , so S_1 also cannot be (ϵ, ϵ) -significant tradeoff leximax. We conclude that no potential solution satisfies this definition².

Computing tradeoff approximations recursively

The definition of ϵ -tradeoff leximax is only useful if we can compute ϵ -tradeoff leximax solutions efficiently. To show that this is possible, we relate ϵ -tradeoff leximax to a different notion of leximax approximation that arises from a natural algorithmic approach and is closely related to the notion of leximax approximations introduced in [10].

Consider the following approach to computing an exact leximax solution, which follows its definition: Compute the maximum value that can be guaranteed to the worst-off group, then calculate the maximum value that can be guaranteed to the second worst-off group subject to this value, and then recurse on the third, fourth, fifth, etc. until the values for all m groups are fixed and a solution is found.

However, what if our algorithm is not completely accurate at each step? Introducing some amount of estimation error at each step of the recursion may result in selecting a solution that isn't exact leximax, but can be considered approximately leximax because it arose from small estimation errors in our algorithm. We call such solutions ϵ -recursive leximax, and define them as follows:

► **Definition 10** (ϵ -recursive leximax). *Given a set of m groups, \mathcal{G} , a set of potential solutions \mathcal{S} , and a choice of allowable "slack" $\vec{\alpha} = (\alpha_1, \dots, \alpha_m)$ with $\alpha_i \in \mathbb{R}_{\geq 0}$, recursively define the sets of solutions $S_0^\alpha, \dots, S_m^\alpha \subseteq \mathcal{S}$ such that $S_0^\alpha := \mathcal{S}$ and for each $i = 1, \dots, m$,*

$$S_i^\alpha = \{S \in S_{i-1}^\alpha : u(S, G_{[i]}) \geq \max_{S' \in S_{i-1}^\alpha} u(S', G_{[i]}) - \alpha_i\}.$$

We say that $S \in \mathcal{S}$ is an ϵ -recursively approximate leximax solution if there exists an $\vec{\alpha}$ with $\max_{i \in [m]} \alpha_i \leq \epsilon$ such that $S \in S_m^\alpha$.

² This example was not tied to the specific choice of $\epsilon_1 = \epsilon_2 = \epsilon$. Similar examples exist for other choices.

Our definition of ϵ -recursive leximax is a stronger variant of the definition of approximation used in [10]. Most importantly, the definition presented in [10] is less strict because it allows for the choice of allowable slack to depend on each solution. However, the solutions outputted by their algorithms actually achieve the stronger notion presented here. Unlike the weaker version, which is only implied by ϵ -tradeoff leximax, we can show that ϵ -recursive leximax and ϵ -tradeoff leximax are equivalent.

In this definition, the choice of slack, $\vec{\alpha} \in [0, \epsilon]^m$, determines the amount of estimation error at each step. We use this $\vec{\alpha}$ to recursively construct the sets \mathcal{S}_i^α in the same way they would be calculated had we applied a recursive approach to calculating a leximax solution but under-estimated the maximum value by α_i at the i th step for each $i = 1, \dots, m$.

Unlike our ϵ -tradeoff leximax notion of approximation, ϵ -recursive leximax provides a natural algorithmic interpretation of approximate solutions which allows efficient approaches to computing ϵ -recursive leximax solutions with respect to a particular choice of slack, as we do in Section 4³. Fortunately, we can actually show that these two notions of approximation are equivalent, which means that we can also efficiently compute ϵ -tradeoff leximax solutions.

► **Theorem 11.** *For any set of groups, \mathcal{G} , and solutions, \mathcal{S} , the set of ϵ -tradeoff leximax solutions is equivalent to the set of ϵ -recursive leximax solutions.*

3.2 Significantly Leximax Solutions

ϵ -tradeoff leximax solutions are strict relaxations of the exact leximax objective. Any leximax-optimal solution will also be ϵ -tradeoff leximax and will also be ϵ -recursive leximax for any $\epsilon \geq 0$ (by simply selecting the allowable slack to be $\alpha_i = 0$ for all $i \in [m]$). Similarly, any ϵ -tradeoff (resp. recursively) approximate solution will also be ϵ' -tradeoff (recursively) approximate for any $\epsilon' \geq \epsilon$.

In this section, we introduce a modified notion of ϵ -recursive leximax that is not a relaxation of the exact leximax objective but rather tries to get significant improvements in the quality of solutions, using the allowed slack. This notion constrains the choices of slack so that solutions considered leximax due to only insignificant improvements in the utility of worse-off groups are ignored. Here, the only slack considered is where all allowable slack values are set to exactly ϵ , rather than some value that is at most ϵ .

► **Definition 12** (ϵ -significant recursive leximax). *Given a set of groups \mathcal{G} with $|\mathcal{G}| = m$ and a set of potential solutions \mathcal{S} , recursively define the sets of solutions $\mathcal{S}_0^\epsilon, \dots, \mathcal{S}_m^\epsilon \subseteq \mathcal{S}$ such that $\mathcal{S}_0^\epsilon := \mathcal{S}$ and for each $i = 1, \dots, m$,*

$$\mathcal{S}_i^\epsilon = \{S \in \mathcal{S}_{i-1}^\epsilon : u(S, G_{[i]}) \geq \max_{S' \in \mathcal{S}_{i-1}^\epsilon} u(S', G_{[i]}) - \epsilon\}.$$

We say that $S \in \mathcal{S}$ is ϵ -significant recursive leximax if $S \in \mathcal{S}_m^\epsilon$.

Why does this make sense as a way to identify significant solutions? Intuitively, setting every slack value to the maximum possible ϵ requires that the valid solutions be leximax with respect to the larger set of potential solutions when some error term is allowed, rather than putting a lot of weight on small differences in earlier groups. We present the following example to see this in practice:

³ [10] give algorithms that calculate ϵ -recursive leximax solutions because their approach estimates each sequential maximum value to within ϵ of its true value, though the notion of efficiency that they achieve does not exactly correspond to polynomial-time algorithms. We provide an alternative polynomial-time algorithm for the cohort selection setting that leverages linear group utilities to offer a more efficient approach.

► **Example 13** (Significantly recursive approximations). Consider two groups and two solutions as in Figure 4 with utilities

$$u(S_1, G_1) = \epsilon, u(S_2, G_1) = 0, u(S_1, G_2) = 0.5, u(S_2, G_2) = 1.$$

Both S_1 and S_2 are ϵ -recursive leximax approximations. If we set $\alpha_1 < \epsilon$, then S_1 because the only acceptable solution and thus an ϵ -recursive leximax-approximate solution. If we set $\alpha_1 = \epsilon$, S_2 becomes an ϵ -recursive leximax-approximate solution. We would expect a satisfying significant approximation notion to identify S_2 as the only ϵ -significant approximation because it's not too far below S_1 on the worst-off group, but does much better on the second-worst-off group. An ϵ -significant recursive leximax approximation does give us this separation between S_1 and S_2 , because while both S_1 and S_2 are included in the first-level of recursion, \mathcal{S}_1^ϵ , S_1 is too far below the maximum to be included in \mathcal{S}_2^ϵ , so S_2 is the only ϵ -significant recursive leximax approximation in this example.

So far, we have been rather loose in arguing about why the solutions identified as ϵ -significant recursive leximax might be preferred over exact leximax or the more general class of ϵ -recursive leximax solutions. We offer a more formal characterization here, but begin by taking a step back to reframe what the contents of the recursively defined sets from Definition 10, $\mathcal{S}_1^\alpha, \dots, \mathcal{S}_m^\alpha$ for some choice of slack $\vec{\alpha}$, can tell us about potential leximax solutions.

Intuitively, \mathcal{S}_i^α contains all solutions that, with respect to the first i groups, could feasibly be solutions that are ϵ -recursive leximax allowing for a slack of $\vec{\alpha}$, and are guaranteed to be within ϵ of the first i coordinates of any final ϵ -recursive leximax solution with respect to $\vec{\alpha}$, i.e. any $S \in \mathcal{S}_m^\alpha$.

This means that looking at the maximum utility achieved by any solution in each recursive group, $\langle \max_{S \in \mathcal{S}_i^\alpha} u(S, G_{[i]}) \rangle_{i=1}^m$ gives us a sense of the type of solution that results from allowing $\vec{\alpha}$ as slack. While there may not exist a $S' \in \mathcal{S}_m^\alpha$ such that $\langle u(S', G_{[i]}) \rangle_{i=1}^m = \langle \max_{S \in \mathcal{S}_i^\alpha} u(S, G_{[i]}) \rangle_{i=1}^m$, we are guaranteed that any $S' \in \mathcal{S}_m^\alpha$ will be elementwise within ϵ of this vector of maximums.

We can show that out of all possible choices of slack, the one used by the definition of ϵ -significant recursive leximax, $\vec{\alpha} = (\epsilon, \dots, \epsilon)$ results in the best-possible sequence of maximum set values (i.e. it will be lexicographically greater than the maximums attained via any other choice of slack). In other words, this backs up the motivation behind our definition of ϵ -significant recursive leximax in that it promises us that any ϵ -significant recursive leximax solution will be elementwise within ϵ of the lexicographically best solution we could possibly hope for under an optimal choice of slack.

► **Theorem 14** (Leximax properties of ϵ -significant recursive leximax). *Given a set of groups, \mathcal{G} , and solutions, \mathcal{S} , let $\mathcal{S}_1^\epsilon, \dots, \mathcal{S}_m^\epsilon$ be the recursively defined sets constructed with a slack of ϵ at each step, as used in the definition of ϵ -significant recursive leximax, and for any $\vec{\alpha} \in \mathbb{R}_{\geq 0}^m$, let $\mathcal{S}_1^\alpha, \dots, \mathcal{S}_m^\alpha$ be the sets that arise when the amount of allowable slack at each level is set according to $\vec{\alpha}$. Then, for any $\vec{\alpha} \in \mathbb{R}_{\geq 0}^m$, we have*

$$\langle \max_{S \in \mathcal{S}_i^\epsilon} u(S, G_{[i]}) \rangle_{i=1}^m \succeq \langle \max_{S \in \mathcal{S}_i^\alpha} u(S, G_{[i]}) \rangle_{i=1}^m.$$

In other words, the vector of maximums attained in each \mathcal{S}_i^ϵ is lexicographically maximal compared to any other choice of slack of size at most ϵ .

Theorem 14 tells us that out of all the ways we could identify approximate leximax solutions that ignore variations of less than ϵ , an ϵ -significant solution is guaranteed to be element-wise within ϵ on the lexicographically maximal best-possible guarantee we can give for each group at each level of recursion.

Ideally, we could obtain a similar notion to ϵ -significant recursive leximax with a satisfying semantic meaning as for ϵ -recursive leximax by modifying our definition of ϵ -tradeoff leximax so that any solution that improves the i th group by more than ϵ must also decrease some worse-off group by more than ϵ . However, as we saw in Example 8, modifying the original tradeoff definition in this way surprisingly results in an overly strict notion due to some instability arising from the pairwise comparisons that tradeoff approximations rely on. In particular, solutions that satisfy this notion may not exist. Note that in Example 8, no solution satisfied (ϵ, ϵ) -significant tradeoff leximax, which is equivalent to the modified definition suggested here, but S_2 is an ϵ -significant recursive leximax approximation and S_2, S_3, S_4 are all valid ϵ -recursive leximax solutions.

4 Solutions via Linear Programming

As discussed in Section 1.2, we provide efficient algorithms for a particular natural choice of cohort selection setting. In particular, we consider modeling utility as the sum of the utilities that a subgroup draws from each individual member of the selected cohort, and rather than outputting a lexicographically maximal cohort, we output a lexicographically maximal vector of marginal selection probabilities that provides leximax utility in expectation.

4.1 Problem Setting

We begin by discussing our choice of utility function and randomized selection approach in more detail.

4.1.1 Linear Utility Function

Let \mathcal{C} be a set of potential committee members of size n . We assume that each subgroup $G_j \in \mathcal{G}$ has a value for each individual committee member $c_i \in \mathcal{C}$, denoted by $v_{ij} \in [0, 1]$.

When we choose our set of solutions to be $\mathcal{C}^{(k)}$, the set of all cohorts of size k , these values can now be combined to give a group's utility for any particular cohort as the sum of its values for the cohort members. Given a cohort $C = \{c_1, \dots, c_k\} \in \mathcal{C}^{(k)}$ and subgroup $G_j \in \mathcal{G}$, this utility function can be written formally as

$$u(C, G_j) = \sum_{i=1}^k v_{ij}$$

This linear utility can easily be extended to the randomized case. Assuming \mathcal{C} has size n , any vector of individual assignment probabilities $D = \{x_1, \dots, x_n\} \in \mathcal{D} := [0, 1]^n$, called *marginal (selection) probabilities*, provides an approach to randomly selecting a cohort of candidates from \mathcal{C} where each c_i is included in the cohort with probability x_i , independent of the other candidates. The expected utility of a particular group G_j over a distribution $D \in [0, 1]^n$ is then

$$u(D, G_j) = \sum_{i=1}^n x_i v_{ij}.$$

We will restrict our search to marginal distributions that output a cohort with expected size k ($\sum_{i=1}^n x_i = k$).

4.1.2 Randomized Selection Approach

Our algorithms output a vector of marginal selection probabilities $D = \{x_1, \dots, x_n\} \in \mathcal{D} := [0, 1]^n$, such that when each cohort member c_i is independently included in the cohort with probability x_i , we get a cohort of size k in expectation such that the vector of expected utilities is lexicographically maximal. This independent sampling procedure provides a simple way to randomly select a cohort.

This distributional approach to selection renders the problem tractable, while we show in Appendix D that finding deterministic leximax solutions is NP-hard. Moreover, it provides a fair way to get around the issue that in a deterministic setting, there may be multiple leximax cohorts that each favor a different subgroup.

It's worth noting that this approach to cohort selection only gives a cohort with *expected* size k . While such a selection procedure may be fine in situations where the desired size of the final cohort is somewhat flexible, sometimes it may be critical to get a cohort of size exactly k . In Appendix C, we discuss a dependent rounding scheme that can be used to sample a cohort of size exactly k with utilities that are still leximax in expectation.

In general, cohorts sampled from arbitrary leximax distributions are not guaranteed to provide groups with utility near their expected value. However, our choice of selection procedure guarantees that groups receive near-expected utility with high probability when the size of the selected cohort is large enough.

► **Lemma 15.** *Consider an arbitrary group G_j and a lexicographically maximal vector of marginal selection probabilities $D \in \mathcal{D}$ (with respect to the linear utility function defined above and with expected size k).*

Then, for any $\delta > 0$, we have

$$\Pr_{C \sim D} [U(C, G_j) < U(D, G_j) - \delta] < e^{-2\delta^2/n}.$$

(Where $n := |\mathcal{C}|$ is the number of potential cohort members.)

To contextualize this result, if we consider some group G_j that is expected to get about half of their maximum possible utility for a leximax solution when $k = 50$ and $n = 100$ (so G_j is expected to get 25 utility), then they are guaranteed to get at least half their expected utility more than 95% of the time. In comparison, an arbitrary leximax distribution can potentially only guarantee that G_j gets more than half their expected utility with probability $1/3$. These concentration guarantees also hold for cohorts of size exactly k outputted by our suggested rounding approach. More details can be found in Section C.

Having explained and justified our choice of utility function as well as randomized selection approach, we now present our algorithms that calculate exact and approximate leximax solutions in this setting.

4.2 Leximax distribution over committee members

To find a marginal distribution over each potential committee member in \mathcal{C} , we break up the problem into multiple, recursively-defined subproblems to uncover the ranking of subgroup utilities in the leximax optimal solution as well as their optimal values.

Balan et. al [3] approach this problem by reducing the domain of solutions in each level of optimization. They choose the $(i + 1)$ -th subgroup to be the subgroup that least-constrains the domain of potential leximax solutions. Overall, their approach finds a leximax-optimal marginal distribution over potential committee members that requires $O(|\mathcal{G}|)$ calls to a linear program at each of the $|\mathcal{G}|$ iterations, giving us $O(|\mathcal{G}|^2)$ total calls. However, this approach

of limiting the domain of the possible solutions requires fixing an order of worst off groups in every iteration. When approximate notions of leximax are introduced, there can be multiple possible orderings of groups to consider.

We suggest finding the leximax distribution over individuals as a series of linear programs with a linear number of variables and a number of constraints that increases from linear to exponential as the series progresses. In the first LP, we are finding the maxmin utility γ_1 using the values v_{ij} that each group has for individual cohort candidates:

$$\begin{aligned} & \text{maximize}_{x, \gamma_1} && \gamma_1 \\ & \text{subject to} && \sum_{i=1}^n x_i = k \\ & && 0 \leq x_i \leq 1 \\ & && \sum_{i=1}^n x_i v_{ij} \geq \gamma_1 \quad j = 1, \dots, m. \end{aligned}$$

Once the optimal lower bound for the worst off group, γ_1^* , is found, we can then maximize the utility of the second-worse-off-group. Ogryczak et al. [26] observed that maximizing the $\gamma = (\gamma_1, \dots, \gamma_m)$ vector is equivalent to maximizing for the cumulative sum of γ_i 's from $i = 1, \dots, m$. Thus, to find the leximax distribution of individuals, we optimize a series of m linear programs using the cumulative leximax values as a constraint. The m -th last LP will be as follows:

$$\begin{aligned} & \text{maximize}_{x, \gamma_m} && \gamma_m \\ & \text{subject to} && \sum_{i=1}^n x_i = k \\ & && 0 \leq x_i \leq 1 \\ & && \sum_{i=1}^n \sum_{G_j \in S} v_{ij} x_i \geq \sum_{s=1}^l \gamma_s^* \quad \forall l = 1, \dots, m, \forall S \subseteq \mathcal{G} \text{ s.t. } |S| = l. \end{aligned}$$

Since we must ensure that the sum of utilities is above the minimum utility for all subgroups, the last constraint requires that the sum of utilities over all sized- l subsets of groups be greater than the sum of the l optimal γ^* -s (i.e. $\sum_{i=1}^l \gamma_i^*$) from previous iterations. This creates $\binom{m}{l}$ constraints for the l -th LP. Algorithm 1 describes the iterative process of finding a leximax distribution where in each successive problem we add additional constraints on the minimum value of the sum of utilities. In our setting of linear utilities, we can solve each linear program in polynomial time with the ellipsoid method using a polynomial-time separation oracle.

■ **Algorithm 1** LEXIMAXCANDIDATES. Finding the leximax distribution over candidates.

Input: $v \in \mathbb{R}_{\geq 0}^{n \times m}$ values of each group for each candidate.

Output: $\{x_1, \dots, x_n\}$ leximax distribution over candidates.

Constraints = $\{\sum_{i=1}^n x_i v_{ij} \geq \gamma_1 \quad j = 1, \dots, m; \quad 0 \leq x_i \leq 1; \quad \sum_{i=1}^n x_i = k\}$;

$\gamma_1^* \leftarrow \max_{x, \gamma_1} \gamma_1$ s.t. Constraints ;

for $l \in 2, \dots, m$ **do**

Constraints = Constraints $\cup \{\sum_{i=1}^n \sum_{G_j \in S} v_{ij} x_i \geq \sum_{s=1}^l \gamma_s^* \quad \forall S \subseteq \mathcal{G} \text{ s.t. } |S| = l\}$;

$\gamma_l^* \leftarrow \max_{x, \gamma_l} \gamma_l$ s.t. Constraints given $\gamma_1^*, \dots, \gamma_{l-1}^*$ (previously computed);

► **Lemma 16.** For n candidates and m groups, the running time of Algorithm 1 is polynomial in n and m .

Approximate Leximax distribution over candidates

When finding approximate leximax distributions over candidates, the approach of Balan et al. [3] can no longer be applied since choosing the subgroup that least constrains the domain of potential solutions may yield multiple subgroups when the leximax objective is approximate. Thus, there is no single ordering of worst-off-groups to rely on when considering group utility. However, we can easily modify our recursive linear program (Algorithm 1) to find an ϵ -recursive leximax solution (Definition 10) for a given “slack” vector $\vec{\alpha} = (\alpha_1, \dots, \alpha_m)$. While the first LP is the same as the exact case, we can loosen the constraints in the m -th LP as follows:

$$\begin{aligned} & \text{maximize}_{x, \gamma_m} && \gamma_m \\ & \text{subject to} && \sum_{i=1}^n x_i = k \\ & && 0 \leq x_i \leq 1 \\ & && \sum_{i=1}^n \sum_{G_j \in S} v_{ij} x_i > \sum_{s=1}^l (\gamma_s^* - \alpha_s) \quad \forall S \subseteq \mathcal{G} \text{ s.t. } |S| = l, \quad l = 1, \dots, m-1 \\ & && \sum_{i=1}^n \sum_{G_j \in \mathcal{G}} v_{ij} x_i > \sum_{s=1}^{m-1} (\gamma_s^* - \alpha_s) + \gamma_m. \end{aligned}$$

For a ϵ -significant recursive leximax approximate solution, we can set all the α_i 's equal to ϵ and apply algorithm 1 with modified constraints as described above.

5 Discussion and Future Work

Motivated by the problem of selecting representative cohorts, we turned to a lexicographically maximal definition of optimal representation. We investigated existing approximations of leximax fairness and introduced new definitions which consider semantic notions of noise and tradeoffs. In settings where utilities or objectives are roughly estimated and leximax fairness is desirable, the approximate notions of leximax in this paper may be useful as alternatives to exact leximax.

While we gave a polynomial time algorithm which computes a leximax distribution over a pool of candidates that is effective for both exact and approximate notions of leximax, finding an algorithm for approximation notions of leximax that is more efficient than exact algorithms remains an open problem. Furthermore, our setting of linear utilities is a natural assumption but can be extended to sub-modular or other classes of utility functions.

In another direction, our approximation notions all reason about allowing for additive amounts of error. However, considering what notions, especially those in line with ϵ -significant recursive leximax, might arise from multiplicative error could be a useful direction to explore.

Finally, we only considered how the presence of additive noise might affect our definitions, but other models of noise specific to different domains may also be considered. Noise can appear not just based on entire cohorts or distributions but also for candidates individually. Modeling how noise from individual candidates accumulate over over cohorts and distributions of candidates will vary depending on the utility function but is a promising direction to explore.

References

- 1 Miriam Allalouf and Yuval Shavitt. Centralized and Distributed Algorithms for Routing and Weighted Max-Min Fair Bandwidth Allocation. *IEEE/ACM Transactions on Networking*, 16(5):1015–1024, October 2008. doi:10.1109/TNET.2007.905605.
- 2 Konstantina Bairaktari, Huy Le Nguyen, and Jonathan Ullman. Fair and optimal cohort selection for linear utilities. *arXiv preprint*, 2021. arXiv:2102.07684.
- 3 Gabriel Balan, Dana Richards, and Sean Luke. Algorithms for leximin-optimal fair policies in repeated games. Technical Report GMU-CS-TR-2008-1, George Mason University, 2008.

- 4 Robert G Bland, Donald Goldfarb, and Michael J Todd. The ellipsoid method: A survey. *Operations research*, 29(6):1039–1091, 1981.
- 5 Robert Bredereck, Piotr Faliszewski, Ayumi Igarashi, Martin Lackner, and Piotr Skowron. Multiwinner elections with diversity constraints. *arXiv preprint*, 2017. [arXiv:1711.06527](https://arxiv.org/abs/1711.06527).
- 6 Rainer E Burkard and Franz Rendl. Lexicographic bottleneck problems. *Operations Research Letters*, 10(5):303–308, 1991.
- 7 Elisa Celis, Vijay Keswani, Damian Straszak, Amit Deshpande, Tarun Kathuria, and Nisheeth Vishnoi. Fair and diverse dpp-based data summarization. In *International Conference on Machine Learning*, pages 716–725. PMLR, 2018.
- 8 L Elisa Celis, Lingxiao Huang, and Nisheeth K Vishnoi. Multiwinner voting with fairness constraints. *arXiv preprint*, 2017. [arXiv:1710.10057](https://arxiv.org/abs/1710.10057).
- 9 Federico Della Croce, Vangelis Th Paschos, and Alexis Tsoukias. An improved general procedure for lexicographic bottleneck problems. *Operations research letters*, 24(4):187–194, 1999.
- 10 Emily Diana, Wesley Gill, Ira Globus-Harris, Michael Kearns, Aaron Roth, and Saeed Sharifi-Malvajardi. Lexicographically Fair Learning: Algorithms and Generalization. *arXiv:2102.08454 [cs, stat]*, February 2021. [arXiv:2102.08454](https://arxiv.org/abs/2102.08454).
- 11 Benjamin Doerr. *Probabilistic Tools for the Analysis of Randomized Optimization Heuristics*, pages 1–87. Springer, January 2020. [doi:10.1007/978-3-030-29414-4_1](https://doi.org/10.1007/978-3-030-29414-4_1).
- 12 Vitalii Emelianov, Nicolas Gast, Krishna P Gummadi, and Patrick Loiseau. On fair selection in the presence of implicit variance. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 649–675, 2020.
- 13 Bailey Flanigan, Paul Gözl, Anupam Gupta, Brett Hennig, and Ariel D Procaccia. Fair algorithms for selecting citizens’ assemblies. *Nature*, 596(7873):548–552, 2021.
- 14 Rupert Freeman, Sujoy Sikdar, Rohit Vaish, and Lirong Xia. Equitable allocations of indivisible goods. In Sarit Kraus, editor, *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 280–286. ijcai.org, 2019. [doi:10.24963/ijcai.2019/40](https://doi.org/10.24963/ijcai.2019/40).
- 15 Mohammad Mahdi Kamani, Rana Forsati, James Z Wang, and Mehrdad Mahdavi. Pareto efficient fairness in supervised learning: From extraction to tracing. *arXiv preprint*, 2021. [arXiv:2104.01634](https://arxiv.org/abs/2104.01634).
- 16 Jon Kleinberg, Yuval Rabani, and Éva Tardos. Fairness in routing and load balancing. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 568–578. IEEE, 1999.
- 17 Jon Kleinberg and Manish Raghavan. Selection problems in the presence of implicit bias. In *9th Innovations in Theoretical Computer Science Conference (ITCS 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.
- 18 Matthäus Kleindessner, Pranjal Awasthi, and Jamie Morgenstern. Fair k-center clustering for data summarization. In *International Conference on Machine Learning*, pages 3448–3457. PMLR, 2019.
- 19 David Kurokawa, Ariel D. Procaccia, and Nisarg Shah. Leximin Allocations in the Real World. *ACM Transactions on Economics and Computation*, 6(3-4):11:1–11:24, October 2018. [doi:10.1145/3274641](https://doi.org/10.1145/3274641).
- 20 Natalia Martinez, Martin Bertran, and Guillermo Sapiro. Minimax pareto fairness: A multi objective perspective. In *International Conference on Machine Learning*, pages 6755–6764. PMLR, 2020.
- 21 Natalia L Martinez, Martin A Bertran, Afroditi Papadaki, Miguel Rodrigues, and Guillermo Sapiro. Blind pareto fairness and subgroup robustness. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7492–7501. PMLR, 18–24 July 2021. URL: <https://proceedings.mlr.press/v139/martinez21a.html>.

- 22 Nimrod Megiddo. Optimal flows in networks with multiple sources and sinks. *Mathematical Programming: Series A and B*, 7(1):97–107, December 1974. doi:10.1007/BF01585506.
- 23 Nimrod Megiddo. A good algorithm for lexicographically optimal flows in multi-terminal networks. *Bulletin of the American Mathematical Society*, 83(3):407–409, 1977.
- 24 Margaret Mitchell, Dylan Baker, Nyalleng Moorosi, Emily Denton, Ben Hutchinson, Alex Hanna, Timnit Gebru, and Jamie Morgenstern. Diversity and inclusion metrics in subset selection. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 117–123, 2020.
- 25 Dritan Nace and Michal Pióro. Max-min fairness and its applications to routing and load-balancing in communication networks: A tutorial. *IEEE Communications Surveys & Tutorials*, 10(4):5–17, 2008.
- 26 Włodzimir Ogryczak, Michal Pióro, and Artur Tomaszewski. Telecommunications network design and max-min optimization problem. *Journal of telecommunications and information technology*, pages 43–56, 2005.
- 27 Candice Schumann, Samsara N Counts, Jeffrey S Foster, and John P Dickerson. The diverse cohort selection problem. *arXiv preprint*, 2017. arXiv:1709.03441.
- 28 Aravind Srinivasan. Distributions on level-sets with applications to approximation algorithms. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 588–597. IEEE, 2001.

A Proof of Lemma 16

Proof. We run m LPs in total. For each LP, the running time is the number of steps the ellipsoid algorithm takes multiplied by the time per iteration. For an efficient implementation, the ellipsoid algorithm needs (1) a feasible initial solution and (2) a polynomial-time separation oracle.

(1) For an initially feasible solution, set $\gamma_1 = 0$ and all $x_i = k/n$ for the first LP. It is easy to check that this gives a feasible solution. In the m -th LP use the x values of the solution to the previous LP and $\gamma_m = 0$ as the initial solution. This solution is feasible as all but the last constraint are identical to the previous LP and, thus, the x -values of the previous solution fulfill them. For the last constraint, note that the right side of the inequality equals the next-to-last constraint. As all utility values are non-negative, summing over a larger set G on the left side only increases the value of the left side in comparison to the value of the next-to-last constraint. Thus, the last constraint is fulfilled as well for $\gamma_m = 0$.

(2) Given a vector of x -values and a vector of minimum utilities γ_i the goal of a separation oracle is to decide whether these values fulfill the LP and, if they do not, find a constraint that is violated by them. The time of the separation oracle dominates the running time per iteration of the ellipsoid algorithm. Thus, it suffices to give a polynomial-time separation oracle. We present our separation oracle in Algorithm 2. It first checks whether all x -values fall into the correct range and add up to k . Then it computes the utility y_j of each subgroup G_j and sorts them in non-decreasing order of y -value. Instead of checking all $\binom{m}{l}$ constraints for each set of l subgroups, it uses the following observation: it suffices to check that, for each l , the sum of the utilities of the l groups with *smallest* utilities is at least $\sum_{s=1}^l \gamma_s$. The reason is that every other set of l subgroups must have cumulative utility at least as large. If, however, the set of l subgroups with minimum utility does not have high enough cumulative utility, then a violating constraint has been found.

Summing up utilities across n candidates takes $O(n)$ time, sorting the resulting utility vector y takes $O(m \log m)$ time. In total, this separation oracle checks if all the constraints are satisfied in $O(m \log m + n)$ time.

■ **Algorithm 2** SEPARATION ORACLE. Checking if a constraint has been violated by a given solution x and γ .

Input: $v \in \mathbb{R}_{\geq 0}^{n \times m}$, values of each group for each candidate, $\{x_1, \dots, x_n\}$ candidate solution, $\{\gamma_1, \dots, \gamma_l\}$ minimum utilities for the l -th LP

Output: {TRUE or a violated constraint}

$S \leftarrow 0$;

for $i = 1, \dots, n$ **do**

| **if** $x_i > 1$ or $x_i < 0$ **then**

| | **return** $\{0 \leq x_i \leq 1\}$

| $S \leftarrow S + x_i$;

if $S \neq k$ **then**

| **return** $\{\sum_{i=1}^n x_i \leq k\}$

$y_j \leftarrow \sum_{i=1}^n v_{ij} x_i \quad \forall j = 1, \dots, m$;

$\tilde{y} \leftarrow \text{SORT}(y)$;

$U_{min} \leftarrow 0$;

for $l = 1, \dots, m$ **do**

| $U_{min} \leftarrow U_{min} + \tilde{y}_l$;

| **if** $U_{min} < \sum_{s=1}^l \gamma_s$ **then**

| | **return** FALSE as this constraint does not hold:

| | $\{\sum_{i=1}^n \sum_{G_j \in S} v_{ij} x_i \geq \sum_{s=1}^l \gamma_s \quad \forall S \subseteq \mathcal{G} \text{ s.t. } |S| = l\}$

return TRUE

For the ellipsoid method, we are guaranteed convergence in k steps where $k \leq 2n^2 \log(\frac{R}{r})$ where R is the initial radius and r is the final radius of the feasible region [4]. For our feasibility region, R is exponential with respect to the input size (i.e. $O(2^n)$) which means $\log(\frac{R}{r})$ is linear with respect to n . Since the separation oracle and centroid method at each step runs in polynomial time and there are at most $\tilde{O}(n^2)$ steps, Algorithm 1 also runs in polynomial time. ◀

B Approximations in the Presence of Noise

So far, we have considered approximate leximax solutions with the assumption that the utilities used to calculate these solutions are known to be correct. However, a natural question is how such approximations behave if the reported utilities contain some small amount of noise.

In the case of ϵ -tradeoff leximax solutions, assuming a small amount of additive noise for each utility has the potential for resulting in solutions that do not satisfy tradeoff guarantees. In particular, noise that is solution-specific can cause individual solutions to be “kicked out” of the recursively defined sets, even though all solutions near them are included. We demonstrate this behavior in the following example:

► **Example 17** (ϵ -tradeoff leximax solutions are not robust to noise.). We consider a setting in which we have two groups, $\mathcal{G} = \{G_1, G_2\}$ and three potential solutions $\mathcal{S} = \{S_1, S_2, S_3\}$. The utilities each group derives are defined as $u(S_i, G_j) = U_{ij}$ where U is defined as follows (assume $\epsilon \ll 0.1$):

$$U = \begin{bmatrix} 0.1 & 0.2 \\ 0.1 + \epsilon/100 & 0.8 \\ 0.1 + \epsilon & 0.2 \end{bmatrix}.$$

Furthermore, assume we have a slightly noisy version of utilities in which $u(S_2, G_1)$ changes from $0.1 + \epsilon/100$ to $0.1 - \epsilon/100$. Figure 5 provides a visual representation of this instance, where the noisy version of S_2 is shown in red.

In the non-noisy version, S_1 can never be considered ϵ -tradeoff leximax because S_2 does much better than S_1 on G_2 , and is still above S_1 on G_1 .

However, in the noisy version, which introduces only a tiny amount of noise ($\epsilon/50$), much smaller than the allowed approximation threshold (ϵ), results in a setting where S_1 can be considered ϵ -tradeoff leximax.

By making the distance between S_2 and S_1 arbitrarily small, we can construct examples where even when the amount of noise is negligible compared to the allowed approximation factor, S_1 can still potentially be incorrectly classified as ϵ -tradeoff leximax.

We note that we can define a stricter notion of tradeoff approximation that guarantees a solution will be ϵ -tradeoff leximax even if calculated with noisy utilities, but for the same reasons as demonstrated in Example 8, such solutions may not always exist, making it difficult to find solutions that are guaranteed to be ϵ -tradeoff leximax in a noisy setting.

► **Lemma 18.** *Recall the notion of (ϵ_1, ϵ_2) -significant tradeoff leximax as presented in Definition 9. Any $(\epsilon - 2\delta, 2\delta)$ -significant tradeoff leximax solution when calculated using noisy utilities within an additive δ of their true values is guaranteed to be ϵ -tradeoff leximax with respect to the true utilities.*

Having considered how noise may affect ϵ -tradeoff leximax approximations, we now turn to ϵ -significant recursive leximax approximations. Here, we find that ϵ -significant recursive leximax solutions are somewhat robust to noise, in that they satisfy a slightly relaxed definition of significance.

First, we note that in Example 17, S_1 is also ϵ -significant recursive leximax in the noisy setting, but not in the non-noisy setting, and so this example also demonstrates how the standard definition of ϵ -significant recursive leximax may not be robust to noise. However, we can offer the following guarantee with respect to a modified notion:

► **Lemma 19.** *Say that a solution S is (α_1, α_2) -significant recursive leximax if there exists some choice of slack $\vec{\beta} = (\beta_1, \dots, \beta_m)$ with $\beta_i : \mathcal{S} \rightarrow [\alpha_1, \alpha_2]$ such that $S \in \mathcal{S}_m^\beta$, where $\mathcal{S}_0^\beta = \mathcal{S}$ and*

$$\mathcal{S}_i^\beta = \{S \in \mathcal{S}_{i-1}^\beta : u(S, G_{[i]}) \geq \max_{S' \in \mathcal{S}_{i-1}^\beta} u(S', G_{[i]}) - \beta_i(S)\}.$$

Then, any ϵ -tradeoff leximax solution calculated in the presence of δ additive noise is guaranteed to be $(\epsilon - 2\delta, \epsilon + 2\delta)$ -significant recursive leximax.

Thus, we conclude that while noisy ϵ -significant recursive leximax solutions are not guaranteed to be ϵ -significant recursive leximax with respect to the true utilities, they will still satisfy a slightly relaxed notion of significance that allows for slack to vary within an interval of size 4δ around the constant ϵ slack used in standard significance. When δ is tiny compared to ϵ , this is only a tiny change in the allowed slack values.

C Rounding Distributions Over Candidates

Once we obtain a distribution over cohort candidates from Algorithm 1, we can sample each individual i with probability x_i independently. The total size of the committee follows a Poisson Binomial distribution which will be size- k in expectation where $k = \sum_{i=1}^n x_i$ according to our constraints.

If a committee of size k is a hard constraint, we can instead take a rounding approach similar to previous work in cohort selection [2]. For finite samples in our cohort selection setting, we can employ a dependent rounding scheme that guarantees that the utilities for each subgroup is leximax in expectation while the size of the cohort is exactly k [28].

The rounding scheme described in [28] results in a distribution over cohorts of size exactly k such that the marginal inclusion probability for each potential cohort member is still satisfied, giving us the leximin utility values in expectation when the utility function is linear over cohort members. The scheme has the added benefit that the events corresponding to the inclusion/exclusion of each cohort member are negatively correlated. Because Chernoff bounds such as the one used in our proof of Lemma 15 have been shown to also hold in settings where random variables are not independent but are negatively correlated (See [11], Theorem 1.10.24), our concentration guarantees also apply to solutions outputted by the rounding scheme.

D Hardness of Integer Solutions

Although the focus of this work is providing distributions over candidates and cohorts, we also touch briefly on the problem of finding integer leximax cohorts. An exact integer leximax solution removes the randomness inherent in rounding from a distributional solution. However, we show such an integer solution is NP-hard to find. Moreover, the weaker maxmin version of the problem (see below) is NP-hard to compute.

Given a set of candidates $\mathcal{C} = \{c_1, \dots, c_n\}$, a set of groups $\mathcal{G} = \{G_1, \dots, G_m\}$, and the values $v \in \mathbb{R}_{\geq 0}^{n \times m}$ of each group for each candidate such that the utility of a group for a cohort is its average value over the cohort's candidates, the integer leximax cohort selection problem is:

$$\begin{aligned} & \text{maximize} && \gamma_1, \dots, \gamma_m \\ & \text{subject to} && \sum_{i=1}^n x_i = k \\ & && x_i \in \{0, 1\} \\ & && \sum_{i=1}^n \sum_{G_j \in \mathcal{G}} v_{ij} x_i \geq \sum_{s=1}^l \gamma_s \quad \forall G \subseteq \mathcal{G} \text{ s.t. } |G| = l, l = 1, \dots, m. \end{aligned}$$

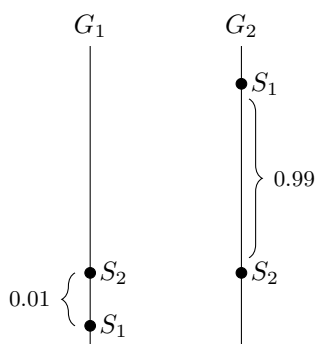
The simpler integer *maxmin cohort selection problem with cardinality k* determines a set of candidates defined by x_i 's such that the minimum utility of any group is maximized:

$$\begin{aligned} & \text{maximize} && \gamma \\ & \text{subject to} && \sum_{i=1}^n x_i = k \\ & && x_i \in \{0, 1\} \\ & && \sum_{i=1}^n x_i v_{ij} \geq \gamma \quad \forall j = 1, \dots, m. \end{aligned}$$

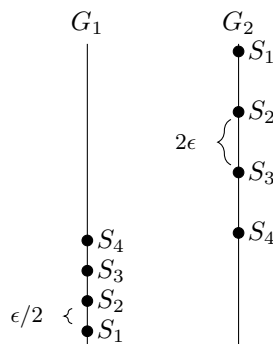
Next we show the hardness of the maxmin cohort selection problem and even of the following *integer ϵ -approximate maxmin cohort selection problem with cardinality k* , where $0 \leq \epsilon$ is a constant: Determine a set of candidates defined by x_i 's such that the minimum utility of any group is within an additive error of ϵ of γ , the maximum minimum utility possible.

► **Lemma 20.** *For $\epsilon < 0.5$ the integer ϵ -approximate maxmin cohort selection problem is NP-hard. It is also NP-hard to determine the number of groups with non-minimum utility to within a factor of $(e - 1)/e + o(1)$.*

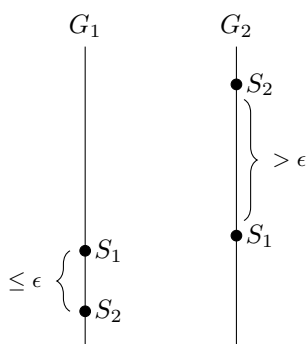
E Figures and Diagrams



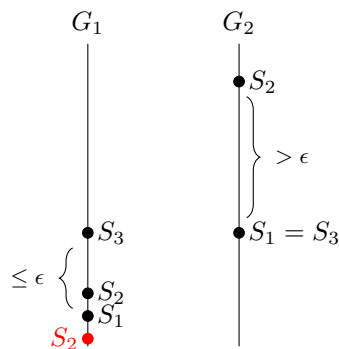
■ **Figure 2** Visual representation of the setting in Example 5 showing how exact leximax solutions are very sensitive to small changes in utility for less-well-off groups.



■ **Figure 3** Visual representation of the setting in Example 8 demonstrating how no solutions may exist under small alterations to the definition of ϵ -tradeoff leximax.



■ **Figure 4** Visual representation of the setting in Example 13 demonstrating how Definition 12 identifies significantly leximax solutions.



■ **Figure 5** Visual representation of the setting in Example 17 showing that when computed in the presence of noise, ϵ -tradeoff leximax solutions may break down. The noisy version consists of updating S_2 to the location highlighted in red.