A Just Approach Balancing Rawlsian Leximax Fairness and Utilitarianism

Violet (Xinying) Chen
xinyingc@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, USA

J.N. Hooker
jh38@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, USA

ABSTRACT

Numerous AI-assisted resource allocation decisions need to balance the conflicting goals of fairness and efficiency. Our paper studies the challenging task of defining and modeling a proper fairness-efficiency trade off. We define fairness with Rawlsian leximax fairness, which views the lexicographic maximum among all feasible outcomes as the most equitable; and define efficiency with Utilitarianism, which seeks to maximize the sum of utilities received by entities regardless of individual differences. Motivated by a justice-driven trade off principle: prioritize fairness to benefit the less advantaged unless too much efficiency is sacrificed, we propose a sequential optimization procedure to balance leximax fairness and utilitarianism in decision-making. Each iteration of our approach maximizes a social welfare function, and we provide a practical mixed integer/linear programming (MILP) formulation for each maximization problem. We illustrate our method on a budget allocation example. Compared with existing approaches of balancing equity and efficiency, our method is more interpretable in terms of parameter selection, and incorporates a strong equity criterion with a thoroughly balanced perspective.

CCS CONCEPTS

• **Computing methodologies** → *Modeling methodologies*; • **Applied computing** → *Multi-criterion optimization and decision-making*.

KEYWORDS

Fairness, Utilitarianism, Distributive Justice, Trade off

ACM Reference Format:
Violet (Xinying) Chen and J.N. Hooker. 2020. A Just Approach Balancing Rawlsian Leximax Fairness and Utilitarianism. In *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES ’20), February 7–8, 2020, New York, NY, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3375627.3375844

1 INTRODUCTION

Fairness is a justifiably major concern in the design of AI systems. However, the most equitable solution may not be the most efficient. For example, awarding mortgage loans in a distressed neighborhood may promote equity but result in more foreclosures and evictions than awarding the loans elsewhere. Business investment in a failed state may serve justice but may also deny funds to less troubled areas where startup capital could create much more benefit. The issue can be especially acute in health care. Very expensive treatments that prolong the life of a few terminally ill patients may divert funds from preventive health measures that would spare millions the suffering brought by less serious diseases.

The intuition behind these examples is that the least advantaged should have priority, but not at unlimited cost to everyone else. Since resource allocation decisions are increasingly made with the assistance of AI technology, it is imperative to find a principled and systematic way to balance fairness and efficiency in such cases. In this paper, we propose one approach to this issue: we define a mathematical model that balances Rawlsian leximax fairness and utilitarianism, and we formulate the model as a practical mixed integer/linear programming problem. The trade-off between fairness and efficiency is regulated by a single parameter that we argue has a more natural and intuitive interpretation that alternative approaches to balancing fairness and efficiency.

We use the notions of utility and social welfare widely studied in the economics literature. For a given outcome, each entity, which represents an individual person or object in this paper, evaluates how good the outcome is using its utility function. The more an entity desires an outcome, the higher its utility will be. With a well defined utility function, each outcome has a vector representation where the elements quantify individual preferences. A convenient tool to compare outcomes is social welfare function (SWF), which aggregates individual utilities into a scalar measuring how preferable an outcome is for all entities as a unity.

Fairness and efficiency objectives correspond to different social welfare functions. In our discussion, we use a utilitarian social welfare function as the efficiency objective. Utilitarianism was advocated in the 18th and 19th century: it considers the optimal outcome to be one that maximizes the sum of utilities enjoyed by all entities. We choose Rawlsian Leximax Fairness as the fairness objective. This fairness definition was given in [7], and has been studied by an extensive line of work [9, 10]. The key justification is based on a social contract with veil of ignorance. The simplest version of Rawlsian fairness principle is known as maximin fairness, which states that fairness is attainable by maximizing the utility of the worst off entity. By recursively applying the rule of prioritizing the worst-off, we reach the more refined definition we choose: Rawlsian leximax fairness considers the lexicographic maximum among all feasible outcomes to be the most equitable. On top of maximizing the worst off utility, leximax fairness tries to maximize the utility

ACM ISBN 978-1-4503-7110-0/20/02…$15.00
https://doi.org/10.1145/3375627.3375844

of the second worst off, the third worst off, and so on. The Rawlsian perspective of defining fairness, while not uncontroversial, has been defended by closely reasoned philosophical arguments in a vast literature [2, 8].

1.1 Related Works

One common technique of modeling an fairness-efficiency trade off is to combine separate fairness and efficiency objectives into a single social welfare measure. A well-studied approach is multiobjective programming [6]: instead of optimizing the pure fairness measure or efficiency measure, the decision problem has a weighted sum of fairness and efficiency measures as the objective function. Another method is to use an inequity-averse social welfare function known as α-fairness. Previous works including [1, 4, 13] have discussed that α-fairness captures a balance of fairness and efficiency. The well-known Nash social welfare function, where the social welfare of an outcome $\mathbf{u}$ is defined the product of individual elements $\Pi_{i=1}^{n} u_i$, is a special case of α-fairness. These two combination methods share the same shortcoming: the parameters lack intuitive explanation in the context of utility and social welfare, hence are difficult to choose or interpret.

Motivated by this shortcoming, [3] proposed a novel social welfare function combining Rawlsian maximin fairness and utilitarianism, which formalizes a trade off principle advocated in [11]: be equitable until it becomes too expensive for efficiency. The social welfare function in [3] uses a single parameter to split a feasible utility region into fair region and utilitarian region, then maximin fairness is active in the fair region and utilitarianism is active in the utilitarian region. [5] applied this new approach in kidney exchange and gave an algorithm to combine leximax fairness and utilitarianism. The algorithm first maximizes the Hooker-Williams social welfare function, then selects the lexicographic maximum among optimal solutions found.

1.2 Our Contributions

Our paper is motivated by the recent development on balancing Rawlsian fairness and utilitarianism. We observe that the social welfare function proposed by [3] does not impose a strong enough fairness criterion: by their definition, not all entities impact social welfare value. The fairness-efficiency trade off algorithm in [5] seeks stronger fairness by adding the extra step of choosing lexicographic maximum, but is not a direct combination of leximax fairness and utilitarianism: only the first step uses a balanced objective.

We propose a sequential optimization procedure to balance leximax fairness and utilitarianism. Compared with [3], our method incorporates a stronger fairness criterion: all entities in the fair region of an outcome have impacts on social welfare values. Compared with [5], our method takes a balanced perspective from beginning to end. Each iteration of our approach maximizes a social welfare function, which we define through extending the justification used for the Hooker-Williams social welfare function.

In this paper, we first review the Hooker-Williams social welfare function and define additional functions needed by our balancing approach. We then describe our approach, which consists of a sequence of social welfare maximization problems. Each maximization problem has a practical mixed integer/linear program

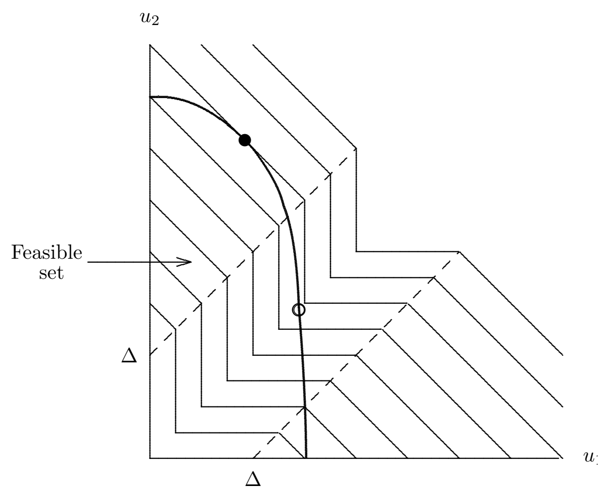

Figure 1: Piecewise linear SWF contours for 2 persons.

representation. We next apply our approach on a simple numerical example of allocating budget to projects. We conclude by discussing additional results not covered in this paper, ongoing work and future directions.

1.3 Preliminary: Maximin Fairness and Utilitarianism

A social welfare function $F(\mathbf{u})$ assigns a numerical value to a distribution of utilities $\mathbf{u} = (u_1, \ldots, u_n)$ across individuals $1, \ldots, n$. Distributions with larger functional values are regarded as more desirable. Recall from Introduction, a utilitarian social welfare function is

$$F^{uti}(\mathbf{u}) = \sum_{i=1}^{n} u_i.$$

Rawlsian maximin fairness has an explicit functional form,

$$F^{maximin}(\mathbf{u}) = \min_{i=1,\ldots,n} u_i.$$

Rawlsian leximax fairness can also be expressed as a social welfare function, using analytical form of lexicographic maximum given in [12]. We do not introduce the function here because it is rarely used in practice, due to potential numerical challenges rising from large parameters.

[11] suggested a social welfare function combining utilitarianism and maximin fairness for two persons, illustrated in Fig. 1, which prioritizes the worse-off individual without taking an excessive amount of resources from the other person. The contours of the function are shown: they are utilitarian when u_1 and u_2 differ by more than Δ and maximin otherwise; specifically,

$$F(u_1, u_2) = \begin{cases} u_1 + u_2, & \text{if } |u_1 - u_2| \geq \Delta \\ 2\min\{u_1, u_2\} + \Delta, & \text{otherwise} \end{cases}$$

The maximin function would ordinarily be $\min\{u_1, u_2\}$, but it is modified here to obtain continuous contours as one moves from a utilitarian region to the maximin region.

The feasible set in Fig. 1 (the portion of the non-negative quadrant under the curve) represents all utility outcomes that are within

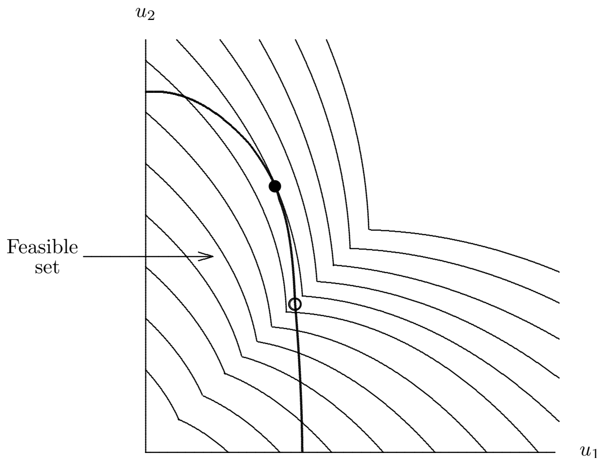

Figure 2: Nonlinear SWF contours for 2 persons.

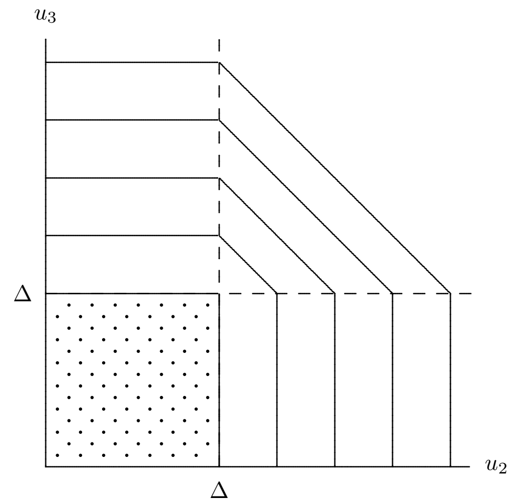

Figure 3: Contours of $F(0, u_2, u_3)$. The function is constant in the shaded region.

the resource capacity. Its shape reflects the fact that after a certain point, further improvement of entity 1's utility requires extraordinary sacrifice by entity 2 due to the transfer of resources. The utilitarian solution (black dot in the figure) might therefore be viewed as preferable to the maximin solution (small open circle) and in fact yields slightly more social welfare as indicated by the contours.

[3] extended this social welfare function to n persons as follows. We adopt the convention that $(u_{\langle 1\rangle}, \ldots, u_{\langle n\rangle})$ is the tuple $(u_1, \ldots, u_n)$ arranged in non-increasing order, that is, $u_{\langle i\rangle} \le u_{\langle i+1\rangle}$ for $i = 1, \cdots, n-1$.

$$F_1(\mathbf{u}) = nu_{\langle 1\rangle} + (n-1)\Delta + \sum_{i=1}^{n} \max\{u_i - u_{\langle 1\rangle} - \Delta, 0\} \qquad (1)$$

We refer to (1) as Hooker-Williams Social Welfare Function. Part of the intuition underlying (1) is that it should be utilitarian along the walls of the nonnegative orthant, where the utilities are widely spread, and Rawlsian near the center diagonal where the utilities are similar. This effect can be obtained by smooth contours that transition gradually from a utilitarian to a Rawlsian criterion, as in Fig. 2, which is perhaps preferable to an abrupt shift from one regime to the other. However, it is far from obvious how to specify and parameterize a nonlinear function of this kind, and it poses a difficult nonconvex continuous optimization problem if one wishes to maximize social welfare. By contrast, the function (1) is parameterized by a single quantity Δ, no matter how many individuals are involved. In addition, [3] showed that the problem of maximizing (1) can be formulated as a mixed integer/linear programming problem that is readily solved by existing software.

2 DEFINING THE SOCIAL WELFARE FUNCTIONS

For ease of discussion, we define $t(\mathbf{u})$ so that $u_{\langle 1\rangle}, \ldots, u_{\langle t(\mathbf{u})\rangle}$ are within Δ of $u_{\langle 1\rangle}$; that is, $u_{\langle i\rangle} - u_{\langle 1\rangle} \le \Delta$ if and only if $i \le t(\mathbf{u})$. We will refer to utilities $u_{\langle 1\rangle}, \ldots, u_{\langle t(\mathbf{u})\rangle}$ as being *in the fair region* and utilities $u_{\langle t(\mathbf{u})+1\rangle}, \ldots, u_{\langle n\rangle}$ as being *in the utilitarian region.*

A serious drawback of (1) is that the utilities in the fair region, other than the smallest, have absolutely no effect on social welfare. This is illustrated in the 3-person example of Fig. 3, which shows the contours of $F(u_1, u_2, u_3)$ with $\Delta = 3$ and u_1 fixed to zero. $F(0, u_2, u_3)$ is constant in the shaded region, meaning that the utilities allocated to persons 2 and 3 do not affect the social welfare as measured by $F(\mathbf{u})$, so long as they remain in the fair region. As a result, there are infinitely many utility vectors that maximize social welfare, some of which differ greatly in the fair region. One can add a tie-breaking term $\epsilon(u_2 + u_3)$ to the social welfare function, where $\epsilon > 0$ is small, to maximize utility as a secondary objective. Yet this still does not account for equity considerations within the fair region.

We address this issue by combining the utilitarian criterion with leximax fairness rather than maximin fairness. A leximax objective takes into account the second lowest utility, the third lowest, and so forth, rather than only the lowest. Formally, a utility vector $\mathbf{u}$ is lexicographically greater than or equal to $\mathbf{u}'$ when $u_{\langle k\rangle} \ge u'_{\langle k\rangle}$ and $(u_{\langle 1\rangle}, \ldots, u_{\langle k-1\rangle}) = (u'_{\langle 1\rangle}, \ldots, u'_{\langle k-1\rangle})$ for some $k \in \{1, \ldots, n\}$. A vector is a lexicographic maximum in a set if it is lexicographically greater than or equal to all other utility vectors in the set.

To combine the leximax and utilitarian criteria, we propose optimizing a *sequence* of social welfare functions $F_1(\mathbf{u}), \ldots, F_n(\mathbf{u})$, each of which balances maximin and utilitarian criteria. The first function $F_1(\mathbf{u})$ is the Hooker-Williams SWF (1) and is maximized over all feasible outcomes to obtain a value for $u_{\langle 1\rangle}$. Each subsequent function $F_k(\mathbf{u})$ is maximized over feasible outcomes while fixing utilities $u_{\langle 1\rangle}, \ldots, u_{\langle k-1\rangle}$ to the values already obtained, to determine a value for $u_{\langle k\rangle}$. The process terminates when maximizing $F_k(\mathbf{u})$ yields a value of $u_{\langle k\rangle}$ that lies outside the fair region. At this point, $F_k(\mathbf{u})$ is utilitarian, and utilities $u_{\langle k\rangle}, \ldots, u_{\langle n\rangle}$ are determined simultaneously by maximizing $F_k(\mathbf{u})$ while maintaining the fixed coordinates. We first define the functions $F_k(\mathbf{u})$ and establish their properties, then describe our sequential optimization procedure more precisely. To build intuition, we rewrite (1) as

$$F_1(\mathbf{u}) = u_{\langle 1\rangle} + (n-1)\big(u_{\langle 1\rangle} + \Delta\big) + \sum_{i=t(\mathbf{u})+1}^{n} (u_{\langle i\rangle} - u_{\langle 1\rangle} - \Delta)$$

Thus the smallest utility $u_{\langle 1\rangle}$ represents itself in the sum, while each u_i in the utilitarian region is discounted by $u_{\langle i\rangle} + \Delta$. That is, the SWF $F_1(\mathbf{u})$ counts only the portion of utility that lies above the threshold $u_{\langle 1\rangle} + \Delta$ of the utilitarian region. This, in effect, gives the lowest utility $u_{\langle 1\rangle}$ more weight and therefore more priority than u_is in the utilitarian region. If $t(\mathbf{u}) = 1$, the function becomes purely utilitarian. This can be generalized to $F_k(\mathbf{u})$ as follows:

$$F_k(\mathbf{u}) = \begin{cases} \sum_{i=1}^{k} u_{\langle i\rangle} + (n-k)(u_{\langle 1\rangle} + \Delta) \\ + \sum_{i=t(\mathbf{u})+1}^{n} (u_{\langle i\rangle} - u_{\langle 1\rangle} - \Delta) \text{ if } t(\mathbf{u}) > k \\ \\ \sum_{i=1}^{n} u_{\langle i\rangle} \text{ if } t(\mathbf{u}) \le k \end{cases} \tag{2}$$

Here, utilities $u_{\langle 1\rangle}, \ldots, u_{\langle k\rangle}$ represent themselves, while each u_i in the utilitarian region is again discounted by $u_{\langle 1\rangle} + \Delta$. When $k \ge t(\mathbf{u})$, the function becomes purely utilitarian. The fair region and utilitarian region are always defined with respect to the lowest utility $u_{\langle 1\rangle}$. The intuition here is that in stage k of the sequential optimization procedure, (2) gives utilities $u_{\langle 1\rangle}, \ldots, u_{\langle k\rangle}$ more priority than those in the utilitarian region in the same way that $F_1(\mathbf{u})$ gives priority to $u_{\langle 1\rangle}$. But $u_{\langle 1\rangle}, \ldots, u_{\langle k-1\rangle}$ have already been fixed in previous stages. Thus utility $u_{\langle k\rangle}$ is given priority, and the value obtained for $u_{\langle k\rangle}$ by maximizing $F_k(\mathbf{u})$ is taken to be the utility allocated to that party in the socially optimal distribution. As k increases, increasingly advantaged parties in the fair region are given priority, until all parties in the fair region are considered. As a result, relatively disadvantaged parties are treated in a leximax fashion, but while taking into account total utility at each stage.

Figure 4 illustrates how maximizing $F_1(\mathbf{u}), \ldots, F_n(\mathbf{u})$ sequentially is more sensitive to equity than maximizing $F_1(\mathbf{u})$, which has the flat region shown in Fig. 3, as noted earlier. Suppose we determine a value for u_1 by maximizing $F_1(\mathbf{u})$, say $u_1 = 0$. Then the function $F_2(\mathbf{u})$ has no flat regions, as is evident in Fig. 4, and the solutions in the flat region of Fig. 3 are now distinguished.

To prove continuity of the SWF $F_k(\mathbf{u})$, it is convenient to write (2) as

$$\begin{aligned} F_k(\mathbf{u}) = & \sum_{i=1}^{k} u_{\langle i\rangle} + (n-k)(u_{\langle 1\rangle} + \Delta) \\ & + \sum_{i=k+1}^{n} \max\{u_{\langle i\rangle} - u_{\langle 1\rangle} - \Delta, 0\}. \end{aligned} \tag{3}$$

Theorem 1. *The functions $F_k(\mathbf{u})$ are continuous for $k = 1, \ldots, n$.*

Proof. It suffices to show that each term of (3) is continuous, because a sum of continuous functions is continuous. Because order statistics are continuous, $u_{\langle i\rangle}$ and $(n-k)(u_{\langle 1\rangle} + \Delta)$ are continuous functions of $\mathbf{u}$. Also $\max\{u_{\langle i\rangle} - u_{\langle 1\rangle} - \Delta, 0\}$ is continuous because it is the maximum of continuous functions. □

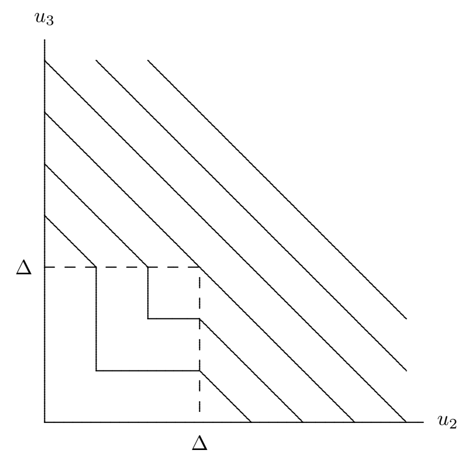

Figure 4: Contours of $F_2(0, u_2, u_3)$.

3 THE OPTIMIZATION PROBLEM

We wish to optimize a sequence of social welfare functions $F_1(\mathbf{u})$, $\ldots, F_n(\mathbf{u})$ subject to resource constraints. Maximizing $F_k(\mathbf{u})$ determines the value of the k-th smallest u_i in the solution of the social welfare problem. We therefore maximize $F_k(\mathbf{u})$ subject to the condition that (a) the u_is already determined are fixed to the values obtained for them, and (b) the remaining u_is must be at least as large as the largest u_i already determined. The unfixed u_i with the smallest value in the solution becomes the utility determined by maximizing $F_k(\mathbf{u})$. We indicate resource limits by writing $\mathbf{u} \in C$, where C is a set of feasible outcomes. The main motivation for our sequential procedure is that the leximax fairness criterion is equivalent to sequential applications of the maximin criterion, where the k-th iteration fixes the k-th smallest value in the final leximax solution.

To state the optimization procedure more precisely, we define a sequence of maximization problems $P_1, \ldots, P_k$, where P_1 maximizes $F_1(\mathbf{u})$ subject to $\mathbf{u} \in C$, and P_k for $k = 2, \ldots, n$ is

$$\begin{aligned} & \max \ F_k(\mathbf{u}) \\ & u_{i_j} = \bar{u}_{i_j}, \quad j = 1, \ldots, k-1 \\ & u_i \ge \bar{u}_{i_{k-1}}, \quad i \in I_k \\ & \mathbf{u} \in C \end{aligned} \tag{4}$$

The indices i_j are defined so that u_{i_j} is the utility determined by solving P_j. In particular, u_{i_j} is the utility with the smallest value among those obtained by solving P_j. Thus

$$i_j = \operatorname*{argmin}_{i \in I_j} \{u_i^{[j]}\}$$

where $\mathbf{u}^{[j]}$ is an optimal solution of P_j and $I_j = \{1, \ldots, n\} \backslash \{i_1, \ldots, i_{j-1}\}$ is the set of undetermined indices in the feasible region of P_j. We denote by $\bar{u}_{i_j} = u_{i_j}^{[j]}$ the solution value obtained for u_{i_j} in P_j, namely, $\bar{u}_{i_j}$ is equal to the smallest value among the $\mathbf{u}^{[j]}$ indices which are not fixed in constraints of P_j.

We need only solve P_k for $k = 1, \ldots, K+1$, where K is the largest k for which $\bar{u}_{i_k} \le \bar{u}_{i_1} + \Delta$. We note that P_{K+1} maximizes a utilitarian objective subject to resource constraints and fixed-value constraints. The solution of our sequential procedure is then

$$u_i^* = \begin{cases} \bar{u}_i & \text{for } i = i_1, \dots, i_{K-1} \\ u_i^{[K]} & \text{for } i \in I_K \end{cases}$$

As mentioned, [3] formulated problem P_1 as a MILP problem. For our new problems $P_2, \dots, P_K$, we can also formulate them as MILP problems. We first need to introduce additional constraints to make P_k (4) MILP-representable. [3] has discussed the case for $k = 1$, so we focus on $k \geq 2$. As shown, we add constraint (5c) to have a MILP-representable problem P_k'.

$$\begin{aligned} & \max \; F_k(\mathbf{u}) \\ & z_k \leq \textstyle\sum_{j=1}^{k-1} \bar{u}_{i_j} + u_{\langle k \rangle} + (n-k)(\bar{u}_{i_1} + \Delta) \\ & + \textstyle\sum_{i \in I_{k+1}} (u_i - \bar{u}_{i_1} - \Delta)^+ && (a) \\ & u_i \geq \bar{u}_{i_{k-1}}, \; i \in I_k && (b) \\ & u_i - \bar{u}_{i_1} \leq M, \; i \in I_k && (c) \\ & \mathbf{u} \in C \end{aligned} \tag{5}$$

We give the following MILP reformulation of P_k'; detailed proof of Theorem 2 is given in appendix.

Theorem 2. *The MILP model for P_k' when $k = 2, \dots, n$ is*

$$\begin{aligned} & \max \; z_k \\ & z_k \leq \sum_{j=1}^{k-1} \bar{u}_{i_j} + (n-k+1)\Delta + \sum_{i \in I_k} v_i && (a) \\ & u_i - \Delta \leq v_i \leq u_i - \Delta\delta_i, \; i \in I_k && (b) \\ & u_i - \Delta \leq v_i \leq u_i - \Delta\epsilon_i, \; i \in I_k && (c) \\ & \bar{u}_{i_1} - M(\delta_i + \epsilon_i) \leq v_i, \; i \in I_k && (d) \\ & v_i \leq \bar{u}_{i_1} + (M - \Delta)(\delta_i + \epsilon_i), \; i \in I_k && (e) \\ & w \leq u_i, \; i \in I_k && (f) \\ & u_i \leq w + M(1 - \epsilon_i), \; i \in I_k && (g) \\ & \textstyle\sum_{i \in I_k} \epsilon_i = 1 && (h) \\ & u_i \geq \bar{u}_{i_{k-1}}, \; i \in I_k && (i) \\ & u_i - \bar{u}_{i_1} \leq M, \; i \in I_k && (j) \\ & \delta_i, \epsilon_i \in \{0, 1\}, \; i \in I_k \\ & \mathbf{u} \in C \end{aligned} \tag{6}$$

4 NUMERICAL EXPERIMENTS

We apply our leximax fairness and utilitarianism balancing approach on an example of budget allocation. Suppose the decision maker (DM) wishes to allocate a total budget of 7000 dollars to 20 projects indexed as $\{1, \dots, 20\}$. For each project i, it has a fixed budget requirement b_i, and the DM decides whether to allocate to a project its full required budget or nothing. Project i's utility is measured as

$$u_i = p_i + r_i y_i, \tag{7}$$

where p_i is the base performance of project i, r_i is the performance increase from receiving its required budget, and $y_i \in \{0, 1\}$ is a binary variable denoting whether project i receives budget. Specifically, $y_i = 1$ means the DM assigns b_i out of \$7000 to project i. Suppose the DM values all projects and wishes to assign budgets fairly and efficiently without exceeding the available amount.

Table 1: Data for Budget Allocation Example

Project i	p_i	r_i	b_i (\$)	i	p_i	r_i	b_i (\$)
[1]	22	95	600	[11]	9	42	400
[2]	12	110	800	[12]	15	50	550
[3]	22	130	950	[13]	5	18	200
[4]	16	120	900	[14]	5	30	400
[5]	20	85	660	[15]	11	35	500
[6]	28	140	1100	[16]	3	20	300
[7]	25	125	1000	[17]	3	40	600
[8]	35	150	1200	[18]	7	22	500
[9]	25	100	850	[19]	10	25	600
[10]	18	80	700	[20]	8	27	700

Table 2: Optimal Decision Balancing Maximin Fairness and Utilitarianism: Projects in Fair Region Shown in Boldface

Δ range	Projects receiving budget	$u^*_{\langle 1 \rangle}$; $\bar{u^*}$
0 − 51	1,2,3,4,5,7,8,9	3; 60.7
52 − 90	2,3,4,6,7,8,9,**13**	3; 59.6
91 − 97	3,4,6,7,8,**13,14,16,17**	7; 53.6
98 − 102	3,6,7,8,**13,14,16,17,18,20**	9; 50.05
103 − 105	2,8,**11,12,13,14,15,16,17,18,19,20**	16; 43.4
106 − 129	**2,8,11,12,13,14,15,16,17,18,19,20**	16; 43.4
130−up	**2,4,11,12,13,14,15,16,17,18,19,20**	18; 41.9

We implement the optimization procedure described in previous sections using the respective MILP representation in each iteration. Since the single parameter Δ affects whether the trade-off is more equitable or more efficient, we experiment with a broad range of integer values $\Delta \in [0, 150]$ to compare the optimal outcomes and provide insights of the corresponding trade-off. The MILP instances are solved with GUROBI 8.1.1. Table 1 contains data in this example: our data are artificial. We index the projects so that the cost over performance-increase ratio b_i/r_i increases with i. To compare our procedure with the Hooker-Williams scheme, we implement both methods: Table 2 and 3 summarize the respective experiment results. We use $\mathbf{u}^*$ to denote the optimal outcome: in result tables, $u^*_{\langle 1 \rangle}$ is the smallest utility of $\mathbf{u}^*$ and $\bar{u^*}$ is the average utility of $\mathbf{u}^*$. For conciseness, we give the list of projects receiving their required budgets in outcome corresponding to $\mathbf{u}^*$, instead of giving the complete $\mathbf{u}^*$. To further facilitate interpretation, we highlight the projects whose utilities are contained in the fair region $[u^*_{\langle 1 \rangle}, u^*_{\langle 1 \rangle} + \Delta]$ in boldface.

As shown in Tables 2 and 3, with Δ increasing, both methods return optimal decision where the worst off entity's utility $u^*_{\langle 1 \rangle}$ is non-decreasing and the average utility decreases. This fits the expected balancing performance of our approach: as Δ increases, the combined criterion captured by our method prioritizes more fairness and less efficiency. Despite the monotonic trends observed in this simple example, it is less obvious whether theoretical results about monotonic behaviors on the optimal outcome from our sequential procedure can be stated. We also observe that the

Table 3: Optimal Decision Balancing Leximax Fairness and Utilitarianism: Projects in Fair Region Shown in Boldface

Δ range	Projects receiving budget	$u^*_{\langle 1 \rangle}; \bar{u}^*$
0 − 90	1,2,3,4,5,7,8,9	3; 60.7
91 − 97	1,2,3,4,7,8,**13,14,16,17**	7; 56.85
98 − 102	1,2,3,4,7,**13,14,16,17,18,20**	9; 51.8
103 − 105	**1**,2,9,**11,12,13,14,15,16,17,18,19,20**	16; 45.65
106 − 108	**1,2**,9,**11,12,13,14,15,16,17,18,19,20**	16; 45.65
109 − 129	**1,2,9,11,12,13,14,15,16,17,18,19,20**	16; 45.65
130−up	**2,4,11,12,13,14,15,16,17,18,19,20**	18; 41.9

size of fair region grows with Δ, implying both methods are more fairness-oriented at larger Δ values.

Recall our index rule, the projects with larger indices have higher cost over performance increase ratio, namely these projects are worse off in terms of utility distribution. We expect the worse off projects to receive better treatment in a fairer balancing scheme, which is exactly what we observe from the experiment results. Another interesting observation from this example is that for the same Δ range, the optimal solution found by our leximax fairness balancing scheme is both fairer and more efficient than the optimal solution found by the Hooker-Williams scheme.

5 CONCLUSION AND FUTURE WORK

Motivated by the justice-driven trade-off principle between fairness and efficiency, that is, one should prioritize fair treatment of the less advantaged until too much sacrifice from efficiency is needed, we design a sequential optimization approach to balance Rawlsian leximax fairness and utilitarianism. We take the novel combination idea presented in [3] to full potentials. Using a parameter Δ, we divide the feasible utility region into fair region and utilitarian region. In all social welfare functions, the utilitarian region is associated with a utilitarian measurement; each function has a different fair region measurement, and all together they characterize leximax fairness.

Our approach maximizes the social welfare function $F_i(\mathbf{u})$ in iteration i. Using optimal solution returned in the i-th iteration, we fix the i-th smallest utility value in the final optimal outcome. Each optimization problem has a practical MILP formulation. We illustrate the practical potentials of our approach with a budget allocation example.

For exposition convenience, we focus on the case where an entity is an individual person or object in this paper. In practice, it is common to have an entity represent a group of people or objects sharing certain characteristics. We have generalized (4) to define slightly different social welfare functions to model group-entities. The optimization procedure is easily extended to fit the general group case. Our ongoing experiments on healthcare provision and paper-reviewer assignment examples have generated interesting results.

On a more theoretical perspective, we have been studying mathematical properties of our balancing approach. One property of special interest to fairness research is the Pigou-Dalton (PD) condition, which states any utility-invariant wealth transfer from a better-off entity to a worse-off entity should not decrease the social welfare value. Many popular fairness measures including α-fairness satisfy the PD condition. Given the sequential nature of our method, we first adapt the conventional definition to specify what it means for our method to violate or satisfy Pigou-Dalton condition. One way to view our approach is that each iteration eliminates some outcomes by fixing one utility coordinate, hence the sequential procedure generates a weak ranking (some outcomes share the same rank) of all feasible outcomes. We suppose a more preferable outcome has a higher rank in the order, then we state that our method violates Pigou-Dalton condition if there exists feasible outcomes $\mathbf{u}, \mathbf{u'}$ such that $\mathbf{u}$ receives a higher rank than $\mathbf{u'}$ but $\mathbf{u}$ can be obtained from $\mathbf{u'}$ via poor-to-rich wealth transfer. An ongoing direction of ours is to provide a definitive answer to whether the Pigou-Dalton condition is violated or satisfied.

Another potentially useful question to study is whether the choice of Δ can be formulated as a well-defined problem. Since Δ directly affects the nature of the underlying trade-off, practical potentials of our approach will be greatly enhanced if we can provide a well-justified process for choosing its value, rather than leaving the choice to trial-and-error.

REFERENCES

[1] Dimitris Bertsimas, Vivek F. Farias, and Nikolaos Trichakis. 2012. On the Efficiency-Fairness Trade-off. *Manage. Sci.* 58, 12 (Dec. 2012), 2234–2250. https://doi.org/10.1287/mnsc.1120.1549

[2] S. Freeman (Ed.). 2003. *The Cambridge Companion to Rawls.* Cambridge University Press.

[3] J. N. Hooker and H. P. Williams. 2012. Combining Equity and Utilitarianism in a Mathematical Programming Model. *Management Science* 58, 9 (2012), 1682–1693. https://doi.org/10.1287/mnsc.1120.1515

[4] T. Lan, D. Kao, and M. Chiang. 2010. An Axiomatic Theory of Fairness in Network Resource Allocation. In *IEEE INFOCOM.*

[5] C.D. McElfresh and J. Dickerson. 2018. Balancing Lexicographic Fairness and a Utilitarian Objective with Application to Kidney Exchange. *AAAI* (02 2018).

[6] W. Ogryczak, A. Wierzbicki, and M. Milewski. 2008. A multi-criteria approach to fair and efficient bandwidth allocation. *Omega* 36, 3 (2008), 451 – 463. https://doi.org/10.1016/j.omega.2005.12.005 Special Issue on Multiple Criteria Decision Making for Engineering.

[7] J. Rawls. 1972. *A Theory of Justice.* Oxford University Press.

[8] H. S. Richardson and P. J. Weithman (Eds.). 1999. *The Philosophy of Rawls* (5 volumes). Garland.

[9] Amartya Sen. 1976. Welfare inequalities and Rawlsian axiomatics. *Theory and decision* 7, 4 (1976), 243–262.

[10] Amartya Sen. 1977. Social choice theory: A re-examination. *Econometrica: journal of the Econometric Society* (1977), 53–89.

[11] A. Williams and R. Cookson. 2000. Equity in Health. *Culyer, A.J. and Newhouse, J.P. editors, Handbook of Health Economics* (Aug. 2000).

[12] R.R. Yager. 1997. On the analytic representation of the Leximin ordering and its application to flexible constraint propagation. *European Journal of Operational Research* 102, 1 (1997), 176 – 192. https://doi.org/10.1016/S0377-2217(96)00217-2

[13] Özlem Karsu and Alec Morton. 2015. Inequity averse optimization in operational research. *European Journal of Operational Research* 245, 2 (2015), 343 – 359. https://doi.org/10.1016/j.ejor.2015.02.035

A PROOF OF THEOREM 2

Proof. We first show that given any $(\mathbf{u}, z_k)$ that is feasible for (5), where $u_{i_j} = \bar{u}_{i_j}$ for $j = 1, \ldots, k-1$, there exist $\mathbf{v}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, w$ for which $(\mathbf{u}, z_k, \mathbf{v}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, w)$ is feasible for (6). Constraints (6i) and (6j) follow directly from (5b) and (5c) respectively. To satisfy the remaining constraints in (6), we set

$$w = u_\kappa,$$

where κ is an arbitrarily chosen index in I_k such that $u_\kappa = u_{\langle k \rangle}$. For $i \in I_k$, we set

$$(v_i, \delta_i, \epsilon_i) = \left\{ \begin{array}{ll} (\bar{u}_{i_1}, 0, 0), & \text{if } u_i - \bar{u}_{i_1} \leq \Delta \text{ and } i \neq \kappa \\ (u_i - \Delta, 0, 1), & \text{if } u_i - \bar{u}_{i_1} \leq \Delta \text{ and } i = \kappa \\ (u_i - \Delta, 1, 0), & \text{if } u_i - \bar{u}_{i_1} > \Delta \text{ and } i \neq \kappa \\ (u_i - \Delta, 1, 1), & \text{if } u_i - \bar{u}_{i_1} > \Delta \text{ and } i = \kappa \end{array} \right\}$$

It is easily checked that these assignments satisfy constraints (6b)–(6h). Constraint (6a) becomes

$$\begin{aligned} z_k \leq & \sum_{j=1}^{k-1} \bar{u}_{i_j} + (n-k+1)\Delta + (u_\kappa - \Delta) \\ & + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ u_i - \bar{u}_{i_1} \leq \Delta}} \bar{u}_{i_1} + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ u_i - \bar{u}_{i_1} > \Delta}} (u_i - \Delta) \end{aligned}$$

This can be written

$$\begin{aligned} z_k \leq & \sum_{j=1}^{k-1} \bar{u}_{i_j} + (n-k+1)(\bar{u}_{i_1} + \Delta) + (u_\kappa - \bar{u}_{i_1} - \Delta) \\ & + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ u_i - \bar{u}_{i_1} \leq \Delta}} (\bar{u}_{i_1} - \bar{u}_{i_1}) + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ u_i - \bar{u}_{i_1} > \Delta}} (u_i - \bar{u}_{i_1} - \Delta) \end{aligned}$$

which simplifies to

$$z_k \leq \sum_{j=1}^{k-1} \bar{u}_{i_j} + u_\kappa + (n-k)(\bar{u}_{i_1} + \Delta) + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ u_i - \bar{u}_{i_1} > \Delta}} (u_i - \bar{u}_{i_1} - \Delta)$$

This is equivalent to (5a), and we conclude that (6a) is satisfied by our assignments above.

For the converse, we show that for any $(\mathbf{u}, z_k, \mathbf{v}, \boldsymbol{\delta}, \boldsymbol{\epsilon}, w)$ that satisfies (6), $(\mathbf{u}, z_k)$ satisfies (5). Constraint (5b) and (5c) are obviously satisfied, due to (6i) and (6j). To verify that (5a) is satisfied, we first let κ be the index for which $\epsilon_\kappa = 1$, which is unique due to (6h). We then write (6a) as follows:

$$z_k \leq \sum_{j=1}^{k-1} \bar{u}_{i_j} + (n-k+1)\Delta + v_\kappa + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ \delta_i = 0}} v_i + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ \delta_i = 1}} v_i$$

This can be written

$$\begin{aligned} z_k \leq & \sum_{j=1}^{k-1} \bar{u}_{i_j} + v_\kappa + (n-k)(\bar{u}_{i_1} + \Delta) \\ & + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ \delta_i = 0}} (v_i - \bar{u}_{i_1}) + \sum_{\substack{i \in I_k \setminus \{\kappa\} \\ \delta_i = 1}} (v_i - \bar{u}_{i_1}) \end{aligned} \tag{8}$$

We show that (8) implies (5a) by observing the following. First, $v_\kappa = u_\kappa = u_{\langle k \rangle}$, due to (6c). The second summation of (8) vanishes because $v_i = \bar{u}_{i_1}$ due to (6d), (6e) and the fact that $\epsilon_i = 0$ for $i \neq \kappa$. In the third summation, we have

$$v_i - \bar{u}_{i_1} = u_i - \bar{u}_{i_1} - \Delta \leq \left(u_i - \bar{u}_{i_1} - \Delta\right)^+$$

where the equality is due to (6b). We conclude that (5a) is satisfied. □