Structural Characteristics of Equitable and Efficient Distributions

Özgün Elçi[1], J. N. Hooker[2], and Peter Zhang[3]

[1,2]Tepper School of Business, Carnegie Mellon University, USA
[3]Heinz School of Public Policy, Carnegie Mellon University, USA

July 2021

Abstract

When used to improve societal outcomes, optimization models that simply focus on utilitarian goals may produce extreme and undesirable solutions. To balance equity and efficiency, researchers have proposed various ways to incorporate both goals in the decision-making process. From the perspective of optimization, this can be done by using a social welfare function that combines equity and efficiency as the objective function. In this work, we show that the plethora of SWFs proposed in the literature can also produce extreme solutions. For example, the maximin (Rawlsian) criterion ignores less fortunate individuals except for the very worst one; alpha fairness SWF may equate an egalitarian solution with an extremely imbalanced solution; Kalai-Smorodinsky bargaining solution may favor individuals that are already privileged by proportionally allocating resources to everyone based on their utility upper bounds; and a threshold SWFs with leximax criterion may produce more moderate outcomes, at the price of increasing complexity in model formulation and analysis. We analyze these SWFs roughly in increasing order of their complexity. Such complexity arises from the consideration to prevent extreme outcomes, thus may attenuate the extremity of solutions. But it also can at the same time lead to more complex and undesirable properties related to other issues. We illustrate the latter point with a representative example throughout, focusing on a simple hierarchical resource allocation model. In conclusion, one needs to be careful in choosing a SWF if we are to avoid unacceptable outcomes.

1. Introduction. Optimization offers a powerful tool for identifying an efficient and equitable allocation of resources. By maximizing a suitable objective function subject to resource limits and other constraints, one can find the best possible allocation of resources as measured by that function. Far and away the most widely used objective is the maximization of total utility, which can take the form of minimizing cost or maximizing revenue or some other benefit. Yet a purely utilitarian criterion lacks an explicit measure of equity. When maximized subject to typical resource constraints, it can lead to extreme and unsatisfactory solutions that benefit a very few at the expense of the many. If this result does not occur for many optimization models used in practice, it is because the complexity of the constraint set excludes extreme solutions—not because they are recognized as unjust, but simply because they do not happen to satisfy all of the constraints. In effect, the constraints conceal the inherent inadequacy of the objective function.

This poses the challenge of identifying an objective function that incorporates within itself a criterion for equity as well as efficiency. We address this challenge by examining the structure of optimal solutions that result from maximizing a variety of social welfare functions (SWFs), subject to basic resource limitations. We find that not only a utilitarian objective, but some of the best-known fairness measures, can result in extreme and often unacceptable solutions. Again, if these extreme solutions are not often observed, it is because they are excluded by side constraints that reflect the exigencies of the situation rather than some underlying concept of fairness. This indicates that one must look further for an SWF that encapsulates an adequate concept of equity.

With this goal in mind, we examine a series of increasingly sophisticated social welfare functions. They include utilitarian, Rawlsian maximin, and leximax SWFs, as well as alpha fairness, proportional fairness (the Nash bargaining solution), the Kalai-Smorodinsky bargaining solution, and recently proposed threshold functions that combine utilitarian with maximin or leximax fairness. We focus primarily on SWFs that combine efficiency and equity criteria in some fashion, partly because this is an obvious strategy for avoiding the extreme outcomes, and partly because efficiency is as important as equity in most practical applications. We derive the structural properties of optimal solutions that result when these SWFs are maximized subject to simple but generic constraints that form the core of a wide variety of applications. Specifically, we suppose there is a budget limit that constrains total available resources, optional bounds on each party's utility, and a linear (or concave piececise linear) utility function that links each party's utility to the resources provided. To our knowledge, very few of these structural results appear in the literature. We find that while each SWF avoids some of the extreme solutions associated with the previous ones, it introduces anomalies of its own. Only the last criterion seems to avoid these difficulties, although it may itself require further refinement.

We also examine the structure of solutions in a hierarchical distribution network. This represents the common situation in which a national authority allocates resources to regions, which in turn combine these with their own resources for distribution to its subregions or institutions. We find that the more sophisticated SWFs are more likely to be regionally nondecomposable, as perhaps they should be. This means that the regional authorities must take into account the national picture before they can equitably allocate resources within their own territory.

The paper begins with a statement of the generic optimization problem, followed by a healthcare example that illustrates how this type of problem can occur in practice. It then states the optimization problem on a hierarchical network and defines concepts of collapsibility, monotone separability, and regional decomposability. Following this, it considers a sequence of SWFs and derives properties of the optimal solutions they deliver. Proofs may be found in the Appendix. The paper concludes by drawing lessons from these results.

2. The Optimization Problem We wish to maximize social welfare subject to a budget constraint. If $\boldsymbol{x} = (x_1, \ldots, x_n)$ are the resources allotted to individuals $1, \ldots, n$, a general distribution problem may be stated

$$\max_{\boldsymbol{x}\in\mathbb{R}^n} \left\{ F\big(\boldsymbol{U}(\boldsymbol{x})\big) \;\middle|\; \sum_j x_j \leq B,\ \bar{\boldsymbol{c}} \leq \boldsymbol{x} \leq \bar{\boldsymbol{d}} \right\} \tag{1}$$

The social welfare function $F(\boldsymbol{u})$ measures the desirability of a distribution of utilities $\boldsymbol{u} = (u_1, \ldots, u_n)$ across individuals $1, \ldots, n$. The utility function $\boldsymbol{U}$ determines the vector $\boldsymbol{u} = \boldsymbol{U}(\boldsymbol{x})$ of utilities resulting from resource allotment $\boldsymbol{x} = (x_1, \ldots, x_n)$. The budget constraint $\sum_j x_j \leq B$ limits total resource consumption to B. The bounds $\bar{\boldsymbol{c}} \leq \boldsymbol{x} \leq \bar{\boldsymbol{d}}$ constrain the resources allotted to individuals by requiring that $\bar{c}_j \leq x_j \leq \bar{d}_j$ for each j.

We will focus on linear utility functions of the form

$$\boldsymbol{U}(\boldsymbol{x}) = (x_1/a_1, \ldots, x_n/a_n) \tag{2}$$

where each $a_j > 0$. This allows us to eliminate $\boldsymbol{x}$ and write (1) as

$$\max_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ F(\boldsymbol{u}) \mid \boldsymbol{a}^\top \boldsymbol{u} \leq B,\ \boldsymbol{c} \leq \boldsymbol{u} \leq \boldsymbol{d} \right\} \tag{3}$$

where $c_j = \bar{c}_j/a_j$ and $d_j = \bar{d}_j/a_j$ for all j. Thus a large coefficient a_j indicates that it is expensive to provide for the welfare of individual j, perhaps due to a disease that is costly to treat. The lower bounds impose a floor on the welfare of each individual, or may reflect a default utility level without resources. The upper bounds reflect the fact that greater resources can yield greater utility only up to a point; once the disease is cured, there is no need to provide more medical resources.

For most SWFs we examine, we will find it revealing to analyze versions of (3) with only a subset of the constraints. In particular, we will study (3) with only a budget constraint:

$$\max_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ F(\boldsymbol{u}) \;\middle|\; \boldsymbol{a}^\mathsf{T}\boldsymbol{u} \le B,\ \boldsymbol{u} \ge \boldsymbol{0} \right\} \tag{4}$$

as well as (3) with only a budget constraint and upper bounds:

$$\max_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ F(\boldsymbol{u}) \;\middle|\; \boldsymbol{a}^\mathsf{T}\boldsymbol{u} \le B,\ \boldsymbol{0} \le \boldsymbol{u} \le \boldsymbol{d} \right\} \tag{5}$$

It will be convenient to suppose that individuals are indexed so that $a_1 \le \cdots \le a_n$. This means that individual 1's welfare is the least costly to provide. We also assume without loss of generality that $0 \le c_j \le d_j \le B/a_j$ for each j, since $0 \le u_j \le B/a_j$ is already enforced by the budget constraint and $\boldsymbol{u} \ge \boldsymbol{0}$.

The model (3) accommodates a wide variety of resource allocation scenarios, one of which is described in the next section. Yet in some cases it may be desirable to measure utility as a nonlinear function of resources, as when there are decreasing returns to scale. In the latter case, the utility function $U(\boldsymbol{x})$ is concave and can be approximated by imposing a system $\boldsymbol{Au} \le \boldsymbol{B}$ of linear budget constraints. In such cases one can solve the problem

$$\max_{\boldsymbol{u}\in\mathbb{R}^n} \left\{ F(\boldsymbol{u}) \;\middle|\; \boldsymbol{Au} \le \boldsymbol{B},\ \boldsymbol{c} \le \boldsymbol{u} \le \boldsymbol{d} \right\} \tag{6}$$

where each budget constraint has the form $A_i\boldsymbol{u} \le B_i$ for $i = 1, \ldots, m$.

3. A motivating example A health provision problem solved by Hooker and Williams (2012) illustrates how the optimization model of the previous section can occur in practice. Hooker and Williams solved the problem using their threshold SWF (described in Section 8), but any of the SWFs we survey can be used. In this section, we focus on the problem constraints.

The problem is to allocate healthcare resources in a manner that is both equitable and efficient, subject to a budget limitation. We are given m treatment groups that are distinguished by the severity and prognosis of the disease. Each group i has size n_i. We let c_i be the cost per patient of administering the treatment, q_i the average net gain in quality-adjusted life years (QALYs) for a member of group i when the treatment is administered, and α_i is the average QALYs that results from medical management without the treatment in question. The budget constraint is

$$\sum_i n_i c_i y_i \leq \tilde{B} \tag{7}$$

where

$$u_i = q_i y_i + \alpha_i \tag{8}$$

and $y_i \in \{0, 1\}$. The variables y_i are binary to require that either all or none of a group receive the treatment. Thus utility in each group (measured in QALYs) is naturally an affine function of resources, because the utility of treatment is proportional to the number of patients treated. We solve a relaxed version of the problem in which $y_i \in [0, 1]$, which allows partial funding for a group, with medical personnel making triage decisions.

Since we have from (8) that $y_i = (u_i - \alpha_i)/q_i$, we substitute this into (7) to get the budget constraint

$$\sum_i \frac{n_i c_i}{q_i} u_i \leq \tilde{B} + \sum_i \frac{n_i c_i \alpha_i}{q_i} \tag{9}$$

Also (8) and $0 \leq y_i \leq 1$ imply the lower and upper bounds

$$\alpha_i \leq u_i \leq q_i + \alpha_i \tag{10}$$

Now the problem of maximizing $F(\boldsymbol{u})$ subject to (9) and (10) has the form of our general model (3), where $a_i = n_i c_i / q_i$ and $B = \tilde{B} + \sum_i n_i c_i \alpha_i / q_i$. This example illustrates how model (3) can encompass a broad class of resource allocation problems with linear utility functions.

4. Hierarchical Distribution. We also examine distribution on a hierarchical network, a type of allocation problem that frequently arises (Simchi-Levi et al. 2019). The resulting optimization problem is a special case of (6). Each region k has an existing resource budget B_k, and the national government must decide how much resources y_k to allocate to each region. If there are r regions, the distribution problem (1) becomes

$$\max_{\boldsymbol{x},\boldsymbol{y}} \left\{ F\big(\boldsymbol{U}(\boldsymbol{x})\big) \;\middle|\; \begin{array}{l} \displaystyle\sum_{k=1}^{r} y_k \leq B,\ \bar{\boldsymbol{c}} \leq \boldsymbol{x} \leq \bar{\boldsymbol{d}},\ \boldsymbol{y} \geq \boldsymbol{0} \\ \displaystyle\sum_{j\in J_k} x_j \leq B_k + y_k,\ k = 1,\ldots,r \end{array} \right\}$$

where J_k is the index set for the subregions in region k. Again using the linear utility function (2), this problem becomes

$$\max_{\boldsymbol{y},\boldsymbol{u}} \left\{ F(\boldsymbol{u}) \;\middle|\; \begin{array}{l} \boldsymbol{e}^{\mathsf{T}}\boldsymbol{y} \leq B,\ \boldsymbol{c} \leq \boldsymbol{u} \leq \boldsymbol{d},\ \boldsymbol{y} \geq \boldsymbol{0} \\ \boldsymbol{a}^k \boldsymbol{u}^k \leq B_k + y_k,\ k = 1,\ldots,r \end{array} \right\} \tag{11}$$

where $\boldsymbol{e} = (1,\ldots,1)$ and vector $\boldsymbol{u}^k$ contains the utilities of subregions of region k.

Interestingly, if we drop the requirement $\boldsymbol{y} \geq \boldsymbol{0}$ (i.e, we allow the national government to take resources from the regions), the model collapses into a single-level problem:

$$\max_{\boldsymbol{u}} \left\{ F(\boldsymbol{u}) \;\middle|\; \boldsymbol{a}^{\mathsf{T}}\boldsymbol{u} \leq B + \sum_{k=1}^{r} B_k,\ \boldsymbol{c} \leq \boldsymbol{u} \leq \boldsymbol{d} \right\} \tag{12}$$

We will say that a hierarchical problem is *collapsible* if it can be solved by solving its collapsed version (12). A problem is collapsible if each region's allocation in the collapsed problem (12) is no less than its stock already on hand.

Proposition 1. *A hierarchical problem* (11) *is collapsible if for any optimal solution* $\bar{\boldsymbol{u}}$ *of* (12), $\boldsymbol{a}^k\bar{\boldsymbol{u}}^k \geq B_k$ *for* $k = 1,\ldots,r$.

Paradoxically, the individual regions may not compute the same distribution for their subregions as recommended by national planners, even when they use the same social welfare function. We will say that a problem is *regionally decomposable* when this issue does not arise. More precisely, (11) is regionally decomposable if $\hat{\boldsymbol{u}} = (\hat{\boldsymbol{u}}^1,\ldots,\hat{\boldsymbol{u}}^r)$ is optimal in (11) for any set of solutions $\hat{\boldsymbol{u}}^1,\ldots,\hat{\boldsymbol{u}}^r$ that are optimal in the regional distribution problems

$$\max_{\boldsymbol{u}^k} \left\{ F(\boldsymbol{u}^k) \;\middle|\; \boldsymbol{a}^k\boldsymbol{u}^k \leq \boldsymbol{a}^k\bar{\boldsymbol{u}}^k,\ \boldsymbol{c}^k \leq \boldsymbol{u}^k \leq \boldsymbol{d}^k \right\} \tag{13}$$

for $k = 1,\ldots,r$ and for any optimal solution $\bar{\boldsymbol{u}}$ of (11). A key to regional decomposability is *monotonic separability* of $F(\boldsymbol{u})$. This means that for any partition $\boldsymbol{u} = (\boldsymbol{u}^1, \boldsymbol{u}^2)$, $F(\bar{\boldsymbol{u}}^1) \geq F(\boldsymbol{u}^1)$ and $F(\bar{\boldsymbol{u}}^2) \geq F(\boldsymbol{u}^2)$ imply $F(\bar{\boldsymbol{u}}) \geq F(\boldsymbol{u})$. In particular, a separable function[1] $F(\boldsymbol{u}) = \sum_j F_j(u_j)$ is monotonically separable. Then we have

Proposition 2. *If $F(\boldsymbol{u})$ is monotonically separable, then problem (11) is regionally decomposable.*

Thus if the SWF is monotonically separable, the regions will distribute their allotment in a way consistent with a nationally optimal solution. However, if the SWF is not monotonically separable, or some regions use a different social welfare criterion, a region's distribution to its subregions may depart from the national plan. If so, the nation's allocation of resources to regions is based on the false assumption that the resources will be distributed within regions as prescribed by the national solution.

5. Utilitarian, Maximin, and Leximax Criteria.

An analysis of utilitarian, maximin and leximax social welfare functions is straightforward, but it reveals the problems inherent in pursuing efficiency or equity alone. The utilitarian SWF

$$F(\boldsymbol{u}) = \sum_j u_j$$

stems ultimately from Jeremy Bentham's idea that one should maximize the greatest good for the greatest number (Bentham 1789). While many policies are ostensibly designed to achieve this goal, a consistently utilitarian objective can lead to extreme and unanticipated distributions. This is borne out by solutions of the optimization problem (3).

Recall that $a_1 \leq \cdots \leq a_n$, so that utility is the least expensive for individual 1. Then we have

Proposition 3. *Problem (4) with a utilitarian SWF has an optimal solution that allocates all available utility to individual 1; that is,*

$$u_1 = B/a_1, \text{ and } u_j = 0, \text{ for } j = 2, \ldots, n$$

When there are upper bounds, as in (5), the available utility is allocated to individual 1 up to the corresponding bound d_1, with any remaining utility going to individual 2 up to d_2, and so forth.

[1] A common assumption made by previous researchers such as Dalton (1920) and Atkinson (1970). For more discussion on such a welfare function, we refer readers to Section 6.

While lavishing all resources on a single or a few individuals seems manifestly unjust, a pure fairness criterion can also produce questionable outcomes. Perhaps the best known is the maximin criterion, which is inspired by the Difference Principle of John Rawls (1999) and discussed in a large literature (Richardson and Weithman 1999, Freeman 2003). The maximin SWF is

$$F(\boldsymbol{u}) = \min_j \{u_j\}$$

While Rawls intended the principle only to apply to the design of social institutions and the distribution of "primary goods," it can be investigated as a possible rule for fair distribution in general.

In problem (3), the maximin criterion is maximized by splitting utility equally among individuals, subject to bounds. This can require a very unequal distribution of resources. For example, when devoting all available resources to a seriously ill patient yields only a slight improvement, nearly all resources must be devoted to that person, and the welfare of everyone else reduced to the same low level. More generally, if we let $a_J = \sum_{j \in J} a_j$ for any $J \in N = \{1, \ldots, n\}$ and let $d_k = \min_j\{d_j\}$, we have

Proposition 4. *Problem* (4) *with a maximin SWF* $F(\boldsymbol{u})$ *has an optimal solution*

$$\boldsymbol{u} = \frac{B}{a_N}\boldsymbol{e} \tag{14}$$

where $\boldsymbol{e} = (1, \ldots, 1)$. *If there are upper bounds, as in* (5), *the solution is the same unless* $B/a_N > d_k$ *where* $d_k = min_j\{d_j\}$, *in which case we let* $u_k = d_k$ *and set* u_j *for* $j \neq k$ *to any feasible value greater than or equal to* d_k.

The proposition reveals another problem with the maximin criterion. If the maximin solution is constrained by the smallest upper bound, everyone's utility can be reduced to this level without affecting social welfare, even when much greater total utility is feasible.

This last anomaly is largely avoided by moving to a *leximax* criterion (lexicographic maximization). It maximizes the utility of the worst-off, and while holding that person's utility fixed, maximizes the utility of the second worst-off, and so forth. The social contract argument with which Rawls defends the maximin criterion can reasonably be extended to a leximax criterion, even if the weaknesses of the argument are inherited as well.

A lexmax solution can be obtained by applying Proposition 4 repeatedly, each time reducing the budget as necessary. Thus we terminate with solution (14) if $B/a_N \leq d_k$. Otherwise, we set $u_k = d_k$ and re-solve the maximin problem with budget constraint $\sum_{j \neq k} a_j u_j \leq B - a_k u_k$, and so forth in recursive fashion. Lexicographic maximization therefore considers disadvantaged persons other than the very worst off, although it is in other respects very similar to the maximin solution and subject to its extreme solutions.

The hierarchical problem (11) is regionally decomposable for the utilitarian and maximin criteria, because these SWFs are monotonically separable. However, it is collapsible for the utilitarian criterion only in the degenerate case where $B = B_k = 0$ for all $k \neq 1$. The utilitarian solution is again extreme. If there are no upper bounds, for example, at most one subregion in each region receives positive utility. A straightforward argument shows that the maximin problem is collapsible when

$$B + \sum_i B_i \geq \min \left\{ a_N d_{\min},\ \max_k \left\{ (a_N / a_{J_k}) B_k \right\} \right\}$$

where $d_{\min} = \min_j \{d_j\}$. An optimal solution of the collapsed problem is

$$u_j = \min \left\{ d_{\min},\ (1/a_N) \Big(B + \sum_k B_k \Big) \right\},\ \text{all } j$$

6. Alpha Fairness. Alpha fairness is perhaps the most popular criterion for balancing equity and efficiency (Mo and Walrand 2000, Lan et al. 2010, Bertsimas et al. 2012). It has the advantage of allowing one to regulate the balance with a parameter α, where larger values of α place a greater emphasis on fairness. In particular, $\alpha = 0$ corresponds to a purely utilitarian criterion, and $\alpha = \infty$ to a maximin criterion. An important special case is proportional fairness, also known as the Nash bargaining solution (Nash 1950), which is frequently used in such engineering contexts as telecommunications and traffic signal timing (Mazumdar et al. 1991, Kelly et al. 1998).

The alpha fairness SWF is

$$F_\alpha(\boldsymbol{u}) = \begin{cases} \dfrac{1}{1-\alpha} \displaystyle\sum_j u_j^{1-\alpha}, & \text{if } \alpha \geq 0 \text{ and } \alpha \neq 1 \\ \displaystyle\sum_j \log(u_j), & \text{if } \alpha = 1 \end{cases}$$

The special case of $\alpha = 1$ corresponds to the Nash bargaining solution. Nash provided an axiomatic justification for this solution, but it rests on a strong axiom of interpersonal noncomparability that arguably rules out the possibility of assessing distributive justice. The Nash solution is also the outcome of certain "rational" bargaining procedures, but they, too, rely on strong assumptions.

Since the alpha fairness SWF is concave (strictly concave for $\alpha > 0$), classical optimality conditions yield a simple closed-form solution for the optimization problem without utility bounds.

Proposition 5. *Problem (4) with an alpha fairness SWF has an optimal solution in which*

$$u_i = \frac{B}{a_i^{1/\alpha} \sum_j a_j^{1-1/\alpha}}, \; i = 1, \ldots, n$$

A solution can be derived for upper bounds as well, but it is complicated to state. It is evident from Proposition 5 that alpha fairness gives some priority to individuals with a smaller budget coefficient a_j, but without giving everything to one individual (if $\alpha > 0$) as in a utilitarian context. The solutions also transform smoothly from utilitarian to maximin as α increases.

Yet alpha fairness is capable of extreme solutions in a nonconvex feasible set, because it can assign equality the same social welfare as arbitrarily extreme inequality. In a 2-player situation, for example, the distribution $\boldsymbol{u} = (s, s)$ has the same social welfare value as (t, T), where

$$t = \begin{cases} s^2/T & \text{if } \alpha = 1 \\ \left(2s^{1-\alpha} - T^{1-\alpha}\right)^{1/(1-\alpha)} & \text{if } \alpha > 1 \text{ and } 2s^{1-\alpha} > T^{1-\alpha} \end{cases}$$

Thus for $\alpha = 1$, we have $t \to 0$ has $T \to \infty$, and for $\alpha > 1$, $t \to 2^{1/(1-\alpha)}s$ as $T \to \infty$, even when social welfare is held fixed. If the feasible set is the union of the box $[0, s] \times [0, s]$ with the box $[0, t] \times [0, T]$, both $[s, s]$ and $[t, T]$ are optimal. Alpha fairness judges an egalitarian solution to be no better than a solution in which one party has arbitrarily more wealth than the other. Oddly, this anomaly does not arise when $0 \leq \alpha < 1$.

An advantage of alpha fairness is that since the SWF is separable and therefore monotonically separable, the hierarchical problem is regionally decomposable (Proposition 2). It is collapsible under the condition stated in Proposition 1.

A difficulty with alpha fairness, however, is that the parameter α is difficult to interpret in practice. It is somewhat helpful to characterize mathematically a welfare-preserving transfer of utility from one individual to another. If $u_k > u_j$, then individual k's utility must be reduced by $(u_k/u_j)^\alpha$ to compensate for a one-unit increase in individual j's utility, if total social welfare is to remain constant. Thus equality is a stronger imperative for larger α, but it is not obvious what particular value of α is appropriate in a given context.

7. Kalai–Smorodinsky Bargaining. The Kalai–Smorodinsky (1975) bargaining solution minimizes each person's relative concession. It is defined as the feasible vector $\boldsymbol{u}$ of utilities that maximizes a scalar β subject to $\boldsymbol{u} = \beta \boldsymbol{u}^{\max}$, where $u_j^{\max}$ is the "ideal" utility for individual j (i.e., the maximum of u_j over all feasible utility distributions). The K–S solution therefore maximizes each individual's fraction of his or her ideal utility, subject to the condition that this fraction is the same for all individuals. This can be interpreted geometrically as the furthest feasible point from the origin on the line segment connecting the origin and $\boldsymbol{u}^{\max}$. A curious feature of the K–S criterion is that it proposes no SWF in the usual sense. The social welfare of a distribution $\boldsymbol{u}$ that lies even slightly off this line segment is undefined.

The K–S solution is easily derived for a budget constraint, with or without upper bounds. Recall that we assume (without loss of generality) that the upper bounds d_j satisfy $d_j \leq B/a_j$ for all j, which means that $\boldsymbol{u}^{\max} = \boldsymbol{d}$. We assume in this section that $\boldsymbol{a}^\top \boldsymbol{d} \geq B$, since otherwise the budget constraint plays no role, and the K–S solution simply sets each utility equal to its upper bound ($\boldsymbol{u} = \boldsymbol{d}$).

Proposition 6. *Problem (5) with a Kalai-Smorodinsky SWF has optimal solution $\boldsymbol{u} = B\boldsymbol{d}/\boldsymbol{a}^\top \boldsymbol{d}$. If there are no upper bounds, as in (4), the solution is $u_j = (1/n)(B/a_j)$ for all j.*

Thus, in the absence of upper bounds, each individual j receives $1/n$ of his or her ideal utility B/a_j. This avoids the extreme solutions of a purely utilitarian or maximin criterion. Yet it may allocate far more utility to an individual whose welfare is easily improved than to one who is less fortunate. For example, it may divert treatment resources from cancer patients to persons suffering from the common cold to provide them the same fraction of their maximum health potential. The K-S model offers no means to prevent this kind of outcome by adjusting the trade-off between equity and efficiency, as is possible with alpha fairness.

The hierarchical problem may or not be collapsible, and a sufficient condition is given in the following proposition. A collapsible problem is regionally decomposable.

Proposition 7. *The hierarchical problem* (11) *with a Kalai-Smorodinsky SWF is collapsible if*

$$\frac{\boldsymbol{a}^k \boldsymbol{d}^k}{\boldsymbol{a}^\top \boldsymbol{d}} \left(B + \sum_i B_i \right) \geq B_k \tag{15}$$

for all k*. Furthermore, the problem is regionally decomposable if it is collapsible.*

A final observation is in order. Since the K–S criterion proposes no social welfare function, an equitable solution is determined entirely by the feasible set. One consequence is that an individual whose potential utility is large receives proportionately more utility. Individuals who happen to enjoy favorable circumstances, perhaps through no merit of their own, are automatically favored. This rules out any notion that justice should compensate for the capriciousness of fate.

8. Threshold Criteria. Williams and Cookson (2000) proposed a pair of 2-person social welfare criteria based on thresholds. One uses an equity threshold: it employs a utilitarian criterion until inequality becomes excessive, at which point it switches to a maximin criterion. The other employs a maximin criterion until the utility cost becomes excessive, at which point it switches to a utilitarian criterion. Hooker and Williams (2012) generalized the latter, utility-based threshold criterion to n persons, formulated a mixed integer programming model for it, and applied it to a health resources problem. Their approach is further generalized by Chen and Hooker (2020a, 2020b) to incorporate a leximax rather than a maximin criterion of fairness. We study the solution structure of the Hooker–Williams approach in this section, and that of its generalization in the next section.

Contours of the 2-person utility-based threshold SWF are illustrated in Fig. 1. They are based on a maximin criterion but switch to a utilitarian criterion when $|u_1 - u_2| > \Delta$. Given the feasible region shown, a maximin solution (small open circle) requires great sacrifice from individual 2. It may therefore be desirable to use a utilitarian solution (solid dot), whose social welfare is slightly greater than that of the maximin solution.

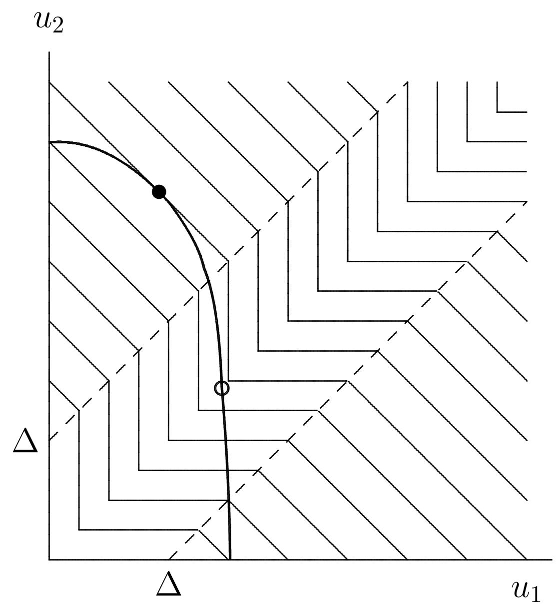

Figure 1: Contours for a 2-person utility threshold SWF.

The n-person SWF of Hooker and Willams is

$$F_\Delta(\boldsymbol{u}) = (n-1)\Delta + nu_{\min} + \sum_{j=1}^{n} (u_j - u_{\min} - \Delta)^+ . \tag{16}$$

where $u_{\min} = \min_j\{u_j\}$. They propose a practically meaningful interpretation of the parameter Δ that goes as follows. Utilities within Δ of the lowest utility are regarded as belonging to the *fair region*, and the corresponding individuals receive special priority. The remaining individuals belong to the *utilitarian region*. The SWF treats utilities in the fair region as though they were equal to the smallest utility, which therefore receives weight in (16) equal to the number of utilities in the fair region. Utilities in the utilitarian region receive unit weight. The function becomes purely utilitarian when $\Delta = 0$ and maximin as $\Delta \to \infty$. The parameter Δ is chosen so as to locate utilities in the fair region when the corresponding individuals should be seen as disadvantaged enough to deserve higher priority. Larger values of Δ therefore place a greater emphasis on fairness as measured by the Rawlsian maximin criterion.

The threshold function $F_\Delta(\boldsymbol{u})$ escapes an anomaly that, as noted earlier, characterizes alpha fairness. It cannot assign equality the same social value as arbitrarily extreme inequality. In a 2-person context, for example, an egalitarian distribution $\boldsymbol{u} = (s, s)$ can have the same social value as a distribution in which one party has no utility and the other $\Delta + 2s$, but the gap can be no greater than this.

When the utility-based threshold SWF is maximized subject only to a single budget constraint as in (4), the solution is either purely utilitarian or purely maximin, depending on the value of Δ. In particular,

Proposition 8. *Problem (4) has a purely utilitarian optimal solution* $\boldsymbol{u} = (B/a_1)\boldsymbol{e}_1$ *when*

$$\Delta \leq B\left(\frac{1}{a_1} - \frac{n}{a_N}\right) \tag{17}$$

and otherwise a purely maximin solution $\boldsymbol{u} = (B/a_N)\boldsymbol{e}$.

When there are lower and upper bounds on individual utilities, as often occurs in practice, the solutions are less extreme and perhaps more useful. They have an interesting structure as well, particularly when there is a single budget constraint. We prove two propositions for m budget constraints and then specialize them to a single budget constraint in a corollary. None of these results rely on our previous assumption that $\boldsymbol{a} > \boldsymbol{0}$.

Proposition 9. *Problem (6) with a utility-based threshold function* $F_\Delta(\boldsymbol{u})$ *has an optimal solution in which at most* m *utilities* u_i *are strictly between* $u_{\min}$ *and* d_i.

Proposition 10. *If a given solution of (6) with a utility-based threshold function* $F_\Delta(\boldsymbol{u})$ *contains exactly* m *utilities strictly between* $u_{\min}$ *and their upper bounds, then some utility* u_j *that is equal to* $u_{\min}$ *is at its lower or upper bound.*

Corollary 11. *Problem (3) with a utility-based threshold function* $F_\Delta(\boldsymbol{u})$ *has an optimal solution in which at most one utility* u_i *is strictly between* $u_{\min}$ *and* d_i. *Furthermore, if there is such a utility, then some other utility* u_j *that is equal to* $u_{\min}$ *is at its lower or upper bound.*

These results say that nearly all utilities will be at their upper bound or equal to the lowest utility. This can simplify implementation and provide managerial insight. In the healthcare example described earlier, for example, it tells us that nearly all patients (all but those suffering from one particular disease) will either receive their maximum possible utility or else end up as one of the worst-off patients, who are given the highest priority. We can also specialize Proposition 9 to a hierarchical network with r regions, each of which has subregions.

Proposition 12. *Any optimal solution of* (11) *with a utility-based threshold function* $F_{\Delta}(\boldsymbol{u})$ *satisfies the following:*

(i) At most r *utilities lie strictly between* $u_{\min}$ *and their upper bounds.*

(ii) At most one utility in any region lies strictly between $u_{\min}$ *and its upper bound.*

The threshold function $F_{\Delta}(\boldsymbol{u})$ is not monotonically separable, and so there is no assurance that a given instance of the problem is regionally decomposable. As a simple example, suppose there are two regions, one with subregional utilities u_1, u_2 and the other with a single utility u_3, and let $\Delta = 1$. Then if $\boldsymbol{a} = (1, 1, 4)$ and $(B, B_1, B_2) = (1, 1, 0)$, Proposition 8 tells us that the solution of the collapsed problem (12) is $\boldsymbol{u} = (2, 0, 0)$. This instance of the problem is, in fact, collapsible by Proposition 1 because $(B_1, B_2) \leq (2, 0)$, and so $\boldsymbol{u} = (2, 0, 0)$ solves the original problem (11). However, the regionally optimal solution for (u_1, u_2) is $(1, 1)$, which is suboptimal in the national problem. This instance of the problem is therefore not regionally decomposable.

9. A Threshold Criterion with Leximax Fairness While a utility-based threshold criterion with maximin fairness tends to avoid extreme solutions, at least in the presence of utility bounds, the maximin component continues to ignore all but the lowest utility value in the fair region. This can result in solutions that are insensitive to the plight of disadvantaged individuals other than the very worst off. Chen and Hooker (2020a, 2020b) avoid this problem by combining utilitarianism with a leximax rather than maximin criterion. An added benefit of this approach is that it tends to avoid extreme solutions even when there is a single budget constraint with no upper bounds on utilities.

The Chen–Hooker approach sequentially maximizes social welfare functions $F_1, \ldots, F_n$, where F_1 is the Hooker–Williams SWF. The first maximization problem P_1 is (3). The remaining maximization problems P_k for

$k = 2, \ldots, n$ are

$$\text{maximize} \quad F_k(\boldsymbol{u}^k) = (n-k+1)u_{\min} + \sum_{i \in I_k} \big(u_i - \bar{u}_{i_1} - \Delta\big)^+ \tag{18a}$$

$$u_i \geq \bar{u}_{i_{k-1}},\ i \in I_k \tag{18b}$$

$$\sum_{i=I_k} a_i u_i \leq B_k \tag{18c}$$

$$\max\{c_i,\ \bar{u}_{i_k}\} \leq u_i \leq d_i,\ i \in I_k \tag{18d}$$

where $u_{\min} = \min_{i \in I_k}\{u_i\}$ and

$$B_k = B - \sum_{j=1}^{k-1} a_{i_j} \bar{u}_{i_j}$$

Problem P_k is solved over the variable set $\{u_i \mid i \in I_k\}$, where $I_k = [1,n] \setminus \{i_1, \ldots, i_{k-1}\}$ and $u_{i_1}, \ldots, u_{i_{k-1}}$ are the variables determined by solving $P_1, \ldots, P_{k-1}$ respectively. Solving each P_k determines u_{i_k} by setting it equal to $\bar{u}_{i_k} = \min_{i \in I_k}\{\bar{u}_i^k\}$, where $\{\bar{u}_i^k \mid i \in I_k\}$ is the optimal solution obtained for P_k. It is convenient to let $\boldsymbol{u}^k$ be the result of removing variables u_i for $i \in I_k$ from the tuple $\boldsymbol{u} = (u_1, \ldots, u_n)$. Chen and Hooker state mixed integer programming models of $P_1, \ldots, P_n$ that can be readily solved in practice.

The solutions of $P_2, \ldots, P_n$ may be neither purely utilitarian nor purely maximin, even when there are no upper bounds on utilities. This is true despite the fact that the functions $F_2, \ldots, F_n$ are rather similar to F_1, which yields a pure solution. A small example illustrates this and other points. Let $n = 3$ and $\Delta = 3$ with budget constraint $3u_1 + 4u_2 + 8u_3 \leq 24$. Problem P_1 has the pure utilitarian solution $\boldsymbol{u} = (8, 0, 0)$ since Δ satisfies Proposition 8, so that $u_{\min} = 0$. We therefore fix, say, $\bar{u}_{\langle 1 \rangle} = u_3 = 0$. Solving P_2, we obtain $(u_1, u_2) = (3, 4)$, which coincides with neither the utilitarian solution $(8, 0)$ nor the maximin solution $(\frac{24}{7}, \frac{24}{7})$. To complete solution of the social welfare problem, we set $\bar{u}_{\langle 2 \rangle} = u_1 = 2$. The solution of P_3 is trivially $\bar{u}_{\langle 3 \rangle} = u_2 = 3$, resulting in a welfare maximizing solution $\boldsymbol{u} = (4, 3, 0)$.

The functions $F_2, \ldots, F_n$ do share a property with F_1, however, in that the solution of each P_k contains at most one utility that is strictly between $u_{\min}$ and its upper bound.

Proposition 13. *Problem P_k for $k \geq 2$ has an optimal solution in which at most one u_i is strictly between $u_{\min}$ and d_i.*

Proposition 14. *If a given solution of (18) for $k \geq 2$ contains exactly one utility strictly between $u_{\min}$ and its upper bound, then some utility u_j that is equal to $u_{\min}$ is at its lower or upper bound.*

Even though the solution of each P_k has the above properties, this need not be the case for the solution of the social welfare problem. In the above 3-person example, the welfare maximizing solution remains $(u_1, u_2, u_3) = (4, 3, 0)$ if we impose bounds $u_1 \in [0, 8]$, $u_2 \in [0, 6]$, and $u_3 \in [0, 3]$. Two utilities (namely, u_1 and u_2) are strictly between $u_{\min} = 0$ and their upper bounds. Thus while one can count on structured solutions while solving each P_k, the resulting solution of the overall social welfare problem can be much more complex, perhaps reflecting the subtleties of the problem.

10. Conclusions. One might construct a narrative from the foregoing observations as follows. The inadequacy of popular optimization objectives becomes evident when they are applied to a generic constraint set consisting of a budget limitation and perhaps bounds on the utilities of each party concerned. A utilitarian objective is by far the most widely used but leads to results that almost any observer would find unacceptable. It allocates all utility to a single party, an outcome that is only marginally ameliorated by placing bounds on individual utilities. While this extreme result is not evident in most practical optimization models, due to the complexity of the constraint set, this complexity only serves to conceal the basic unreasonableness of a purely utilitarian criterion.

Objectives based solely on equity can yield equally extreme and unacceptable solutions. Perhaps the most famous fairness criterion, the maximin objective that derives from the Rawlsian difference principle, forces all parties to accept the same level of utility, except, again, where this is blocked by other constraints—constraints that may reflect only the situation and no coherent understanding of what is just. For example, if one person is difficult to accommodate, due to unfortunate circumstances such as incurable disease, it is necessary to lavish resources on that person to the point that all others are reduced to the same level of suffering. Even if we prevent this outcome by placing a low upper bound on the disadvantaged party's utility, the maximin objective allows us to allot others that same low level of utility, when they could receive much more. A leximax objective largely removes this second anomaly, but the extreme solutions remain.

A natural strategy for avoiding extreme solutions is to combine equity and utilitarian considerations in some fashion. A simple convex combination is both unprincipled and difficult to calibrate, particularly since equity and efficiency tend to be measured in different units. Alpha fairness, perhaps the best known composite criterion, allows a parameterized balancing of fairness and efficiency that avoids the extreme solutions just descibed. Yet it creates an extreme result of its own, because it can regard an egalitarian distribution as no better than one in which there is extreme inequality. This does not become evident in optimal solutions subject to simple budgetary and bounding constraints, but it can emerge in nonconvex constraint sets. It is also unclear how to select and interpret the balancing parameter. The Kalai-Smorodinsky bargaining solution avoids the extreme outcomes of alpha fairness, but at the cost of another extreme that is opposite to that of the maximin criterion. It awards wealthy and privileged individuals the same fraction of their potential utility as individuals who have far less potential, perhaps due to some physical or mental impairment. There is also no parameter for regulating the equity/efficiency balance.

Threshold functions provide an alternate means for combining equity and efficiency, where the balance is governed a parameter Δ that is more easily interpreted in practice. A utility threshold function employs a maximin criterion but switches to a utilitarian criterion when utility cost fairness crosses a specified threshold, while an equity threshold function does the reverse. Yet threshold criteria can lead to extreme solutions, at least in the presence of the simplest constraints. The utility threshold function, for example, yields a purely utilitarian or purely maximin solution in the presence of a single budget constraint without utility bounds, although one can state in closed form which values of Δ produce one or the other. The addition of utility bounds results in much more reasonable solutions, unlike the situation with a simple utilitarian objective. In addition, the resulting solutions have an interesting structure that can ease implementation and provide managerial insights.

Nonetheless, threshold functions that combine utilitarian with maximin objectives inherit a shortcoming of the latter, if only in attenuated form. This is the tendency to give insufficient attention to disadvantaged parties other than the very worst off. The problem can again be addressed by replacing a maximin with a leximax criterion, in this case by optimizing a certain sequence of threshold functions rather than a single one. The resulting solutions avoid all the extremes described here, even in the presence of a simple budget constraint without utility bounds. While the individual threshold functions share the structural properties mentioned earlier, the socially optimal solutions that result are more complex in nature. A threshold-based combination of utilitarian and leximax critertia doubtless has shortcomings of its own, but it illustrates the thesis that we must move beyond the naïveté of simpler social welfare functions if we are to avoid unacceptable results.

This narrative is enriched by observing the behavior of the various criteria on hierarchical networks in which resources can be passed from a national to a regional level. Each region combines its own resources with those received from above and distributes them to subregions (which can be interpreted as hospitals or other institutions). All of the extremes discussed above persist in this context. However, the simpler models are more likely to be regionally decomposable, due to a technical property (monotone separability) of the corresponding social welfare functions. Regional decomposability means that if each region distributes its allotted resources using the same social welfare criterion as used at the national level, it will obtain an allocation to subregions that is consistent with that prescribed at the national level. The allocation computed by the national authority assumes, in fact, that the allocation within regions will follow this pattern. When there is no regional decomposability, the national solution is valid only if the regions follow the national prescription for intra-regional distribution rather than computing the distribution themselves.

To be specific, the pure utilitarian and maximin criteria are regionally decomposable, as is alpha fairness. The Kalai-Smorodinsky model is regionally decomposable when it is collapsible, meaning that the multilevel problem can be solved as a single-level problem. A sufficient condition for collapsibility can be checked by applying a simple test. The utility threshold model can be regionally nondecomposable even when it is collapsible. Thus the more sophisticated models are progressively less prone to be decomposable. This might be regarded as an inconvenience, but it can also signal greater adequacy and subtlety as an equity measure. Perhaps local decisions should reflect a larger perspective if they are to be truly fair.

References

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2(3):244–263.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation*. reprinted by Oxford University Press (1907).

Bertsimas, D., Farias, V. F., and Trichakis, N. (2012). On the efficiency-fairness trade-off. *Management Science*, 58(12):2234–2250.

Chen, V. and Hooker, J. N. (2020a). Balancing fairness and efficiency in an optimization model. *ArXiv preprint 2006.05963*.

Chen, V. and Hooker, J. N. (2020b). A just approach balancing Rawlsian leximax fairness and utilitarianism. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 221–227.

Dalton, H. (1920). The measurement of the inequality of incomes. *The Economic Journal*, 30(119):348.

Freeman, S., editor (2003). *The Cambridge Companion to Rawls*. Cambridge University Press.

Hooker, J. N. and Williams, H. P. (2012). Combining equity and utilitarianism in a mathematical programming model. *Management Science*, 58(9):1682–1693.

Kalai, E. and Smorodinsky, M. (1975). Other solutions to Nash's bargaining problem. *Econometrica*, 43:513–518.

Kelly, F. P., Maulloo, A. K., and Tan, D. K. H. (1998). Rate control for communication networks: Shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49(3):237–252.

Lan, T., Kao, D., Chiang, M., and Sabharwal, A. (2010). An axiomatic theory of fairness in network resource allocation. In *Conference on Information Communications (INFOCOM 2010)*, pages 1343–1351. IEEE.

Mazumdar, R., Mason, L., and Douligeris, C. (1991). Fairness in network optimal flow control: Optimality of product forms. *IEEE Transactions on Communications*, 39(5):775–782.

Mo, J. and Walrand, J. (2000). Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8:556–567.

Nash, J. (1950). The bargaining problem. *Econometrica*, 18:155–162.

Rawls, J. (1999). *A Theory of Justice* (revised). Harvard University Press (original edition 1971).

Richardson, H. S. and Weithman, P. J., editors (1999). *The Philosophy of Rawls* (5 volumes). Garland.

Simchi-Levi, D., Trichakis, N., and Zhang, P. Y. (2019). Designing response supply chain against bioattacks. *Operations Research*, 67(5):1246–1268.

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

Appendix (Proofs)

Proof of Proposition 1. Let $\bar{\boldsymbol{u}}$ be an optimal solution of (12) that satisfies $\boldsymbol{a}^k\bar{\boldsymbol{u}}^k \geq B_k$ for $k = 1, \ldots, r$. If we let $\bar{y}_k = \boldsymbol{a}^k\bar{\boldsymbol{u}}^k - B_k \geq 0$, then $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{u}})$ is feasible in (11) because

$$\sum_{k=1}^{r} \bar{y}_k = \sum_{k=1}^{r} \boldsymbol{a}^k\bar{\boldsymbol{u}}^k - B_k \leq B \tag{19}$$

where the inequality is due to the constraint in (12). Also any $\boldsymbol{u}$ feasible in (11) is feasible in (12), which implies $F(\boldsymbol{u}) \leq F(\bar{\boldsymbol{u}})$ since $\bar{\boldsymbol{u}}$ is optimal in (12). Thus $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{u}})$ is optimal in (11). □

Proof of Proposition 2. Let $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{u}})$ be optimal in (11) and $\hat{\boldsymbol{u}}^k$ optimal in (13) for $k = 1, \ldots, r$. Then if we let $\hat{y}_k = \boldsymbol{a}^k\hat{\boldsymbol{u}}^k - B_k$, we have from (19) and the constraint in (13) that

$$\sum_{k=1}^{r} \hat{y}_k = \sum_{k=1}^{r} \boldsymbol{a}^k\hat{\boldsymbol{u}}^k - B_k \leq \sum_{k=1}^{r} \boldsymbol{a}^k\bar{\boldsymbol{u}}^k - B_k \leq B$$

and so $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{u}})$ is feasible in (11). Also if a given $\boldsymbol{u}$ is feasible in (11), then $\boldsymbol{u}^k$ is feasible in (13) for each k. But $F(\boldsymbol{u}^k) \leq F(\hat{\boldsymbol{u}}^k)$ for each k because $\hat{\boldsymbol{u}}^k$ is optimal in (13). Since $F(\boldsymbol{u})$ is monotonically separable, this implies $F(\boldsymbol{u}) \leq F(\hat{\boldsymbol{u}})$, and $(\hat{\boldsymbol{y}}, \hat{\boldsymbol{u}})$ must be optimal in (11). □

Propositions 3–5 are straightforward.

Proof of Proposition 6. The K–S problem is

$$\max_{\beta,\boldsymbol{u}} \left\{ \beta \;\middle|\; \boldsymbol{u} = \beta\boldsymbol{u}^{\max},\ \boldsymbol{a}^{\mathsf{T}}\boldsymbol{u} \leq B,\ 0 \leq \boldsymbol{u} \leq \boldsymbol{d},\ 0 \leq \beta \leq 1 \right\}$$

Substituting $\beta\boldsymbol{u}^{\max} = \beta\boldsymbol{d}$ for $\boldsymbol{u}$, this becomes

$$\max_{\beta} \left\{ \beta \mid \beta\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d} \leq B,\ 0 \leq \beta\boldsymbol{d} \leq \boldsymbol{d},\ 0 \leq \beta \leq 1 \right\}$$

The constraint $0 \leq \beta\boldsymbol{d} \leq \boldsymbol{d}$ is redundant, and $B/\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d} \leq 1$ since we are given that $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d} \geq B$. The optimal solution is therefore $\beta = B/\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d} \leq 1$, and $\boldsymbol{u} = \beta\boldsymbol{d} = B\boldsymbol{d}/\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d}$. If there are no upper bounds, we can set $d_i = B/a_i$ for each i, so that $\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d} = nB$. Thus we have the optimal solution $u_j = (1/n)B/a_j$ for all j. □

Proof of Proposition 7. The collapsed problem (12) is

$$\max_{\beta,\boldsymbol{u}} \left\{ \beta \;\middle|\; \boldsymbol{u} = \beta \boldsymbol{d},\ \boldsymbol{a}^{\mathsf{T}}\boldsymbol{u} \le B + \sum_i B_i,\ 0 \le \beta \le 1 \right\}$$

By Proposition 6, the solution of this problem is

$$\bar{\boldsymbol{u}} = \Big(B + \sum_i B_i\Big)\boldsymbol{d}/\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d}, \ \text{ or } \ \bar{\beta} = \big(1/\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d}\big)\Big(B + \sum_i B_i\Big) \tag{20}$$

where the latter expression is due to $\bar{\boldsymbol{u}} = \bar{\beta}\boldsymbol{d}$. By Proposition 1, this solves the original problem (11) if $\boldsymbol{a}^k\bar{\boldsymbol{u}}^k \ge B_k$ for all k. Substituting the value of $\bar{\boldsymbol{u}}$, we obtain (15).

To show that the hierarchical problem is regionally decomposable, we note that region k's problem (13) is

$$\max_{\beta_k} \left\{ \beta_k \;\middle|\; \boldsymbol{a}^k\boldsymbol{d}^k\beta_k \le \boldsymbol{a}^k\bar{\boldsymbol{u}}^k,\ 0 \le \beta_k \le 1 \right\}$$

where $\boldsymbol{u}^k = \beta_k\boldsymbol{d}^k$. Substituting the value of $\bar{\boldsymbol{u}}^k$ in (20), this becomes

$$\max_{\beta_k} \left\{ \beta_k \;\middle|\; \boldsymbol{a}^k\boldsymbol{d}^k\beta_k \le \frac{\boldsymbol{a}^k\boldsymbol{d}^k}{\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d}}\Big(B + \sum_i B_i\Big),\ 0 \le \beta_k \le 1 \right\}$$

By Proposition 6, the solution of this problem is

$$\hat{\boldsymbol{u}}^k = \hat{\beta}_k\boldsymbol{d} = \Big(B + \sum_i B_i\Big)\boldsymbol{d}/\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d}, \ \text{ or } \ \hat{\beta}_k = \big(1/\boldsymbol{a}^{\mathsf{T}}\boldsymbol{d}\big)\Big(B + \sum_i B_i\Big)$$

Thus we have $\hat{\beta}_k = \bar{\beta}$ for all k, and $\hat{\boldsymbol{u}} = \hat{\beta}\boldsymbol{d} = \bar{\beta}\boldsymbol{d} = \bar{u}$ solves the original problem (11). □

To prove Proposition 8, it is useful to reformulate (4) as follows:

$$\max_{v_0,\boldsymbol{v}} \left\{ \bar{F}_\Delta(v_0,\boldsymbol{v}) \;\middle|\; a_N v_0 + \boldsymbol{a}^{\mathsf{T}}\boldsymbol{v} \le B,\ \boldsymbol{v} \ge \boldsymbol{0} \right\} \tag{21}$$

where $\boldsymbol{v} = (v_1,\ldots,v_n)$ and

$$\bar{F}_\Delta(v_0,\boldsymbol{v}) = (n-1)\Delta + nv_0 + \sum_{i=1}^n \big(v_i - \Delta\big)^+$$

Lemma 15. *Formulation (21) has the same optimal value as (4).*

Proof. It suffices to show that for any feasible solution of (4), there is a feasible solution of (21) with value at least as large as that of (4), and vice-versa. First consider any feasible solution $\boldsymbol{u}$ of (4). If we let $v_0 = u_{\min}$ and $v_j = u_j - u_{\min}$ for all j, the solution $(v_0, \boldsymbol{v})$ is feasible in (21), given the constraints of (4). Also the objective function $\bar{F}_\Delta$ is identical to F_Δ, and so $\bar{F}_\Delta(v_0, \boldsymbol{v}) = F_\Delta(\boldsymbol{u})$.

Now suppose that $(v_0, \boldsymbol{v})$ is feasible in (21). Set $u_j = v_0 + v_j$ for all j, which implies $u_{\min} = v_0 + v_{\min}$, where $v_{\min} = \min_j\{v_j\}$. The constraint $\boldsymbol{a}^\mathsf{T}\boldsymbol{u} \leq B$ of (4) becomes $a_N v_0 + \boldsymbol{a}^\mathsf{T}\boldsymbol{v} \leq B$ in (21), so that $\boldsymbol{u}$ is feasible in (4). The objective function of (4) becomes

$$(n-1)\Delta + nv_{\min} + nv_0 + \sum_j \big(v_j - v_{\min} - \Delta\big)^+$$

which can be written

$$(n-1)\Delta + nv_0 + \sum_j \max\big\{v_j - \Delta,\ v_{\min}\big\}$$

This is no smaller than the objective function of (21) because $v_{\min} \geq 0$. □

Proof of Proposition 8. We will show that $(v_0, \boldsymbol{v}) = (0, (B/a_1)\boldsymbol{e}_1)$ is an optimal solution of (21) if (17) holds, and $(v_0, \boldsymbol{v}) = (B/a_N, \boldsymbol{0})$ is an optimal solution otherwise. This proves the theorem because (4) and (21) have the same optimal value by Lemma 15, and because

$$\begin{aligned} F\big((B/a_1)\boldsymbol{e}_1\big) &= (n-2)\Delta + B/a_1 = \bar{F}\big(0, (B/a_1)\boldsymbol{e}_1\big) \\ F\big((B/a_N)\boldsymbol{e}\big) &= (n-1)\Delta + nB/a_N = \bar{F}\big(B/a_N, \boldsymbol{0}\big) \end{aligned}$$

We first observe that the objective function $\bar{F}(v_0, \boldsymbol{v})$ is convex because $(v_j - \Delta)^+$ is a convex function of v_j, and a sum of convex functions is convex. It follows that some extreme point of the feasible set of (21) is optimal. Yet every extreme point is the solution of some linearly independent subset T of $n+1$ of the following equations:

$$\begin{aligned} & a_N v_0 + \boldsymbol{a}\boldsymbol{v} = B && (a) \\ & v_0 = 0 && (b) \\ & v_j = 0,\ j \in [1, n] && (c) \end{aligned}$$

We can suppose T contains (a), since otherwise the corresponding extreme point $(v_0, \boldsymbol{v}) = (0, \boldsymbol{0})$ is clearly dominated by $(0, (B/a_1)\boldsymbol{e}_1)$ and $(B/a_N, \boldsymbol{0})$. Then T either contains (b) and all but one equation $v_j = 0$ in (c), or else all equations in (c). The former yields extreme point $(v_0, \boldsymbol{v}) = (0, (B/a_j)\boldsymbol{e}_j)$ and the latter $(v_0, \boldsymbol{v}) = (B/a_N, \boldsymbol{0})$. Now if (17) holds, both $\bar{F}(0, (B/a_j)\boldsymbol{e}_j) = (n-2)\Delta + B/a_t$ and $\bar{F}(B/a_N, \boldsymbol{0}) = (n-1)\Delta + nB/a_N$ are less than or equal to $\bar{F}(0, (B/a_t)\boldsymbol{e}_t)$ because $a_1 \le a_j$. If (17) does not hold, both of these expressions are less than or equal to $\bar{F}(B/a_N, \boldsymbol{0})$. The proposition follows. □

Proof of Proposition 9. Let S be the feasible set of (6), and define

$$S_i = S \cap \{\boldsymbol{u} \mid u_i \le u_j, \text{ all } j\}$$

Since S is the union of all S_i, the maximum of $F(\boldsymbol{u})$ over some S_i is optimal in (6). Suppose without loss of generality that the maximum of $F(\boldsymbol{u})$ over S_1 is optimal in (6). For $\boldsymbol{u} \in S_1$, the function $F(\boldsymbol{u})$ can be written

$$F_1(\boldsymbol{u}) = (n-1)\Delta + nu_1 + \sum_{i=1}^{n} \big(u_i - u_1 - \Delta\big)^+$$

Since $F_1(\boldsymbol{u})$ is convex, some extreme point $\bar{\boldsymbol{u}}$ of S_1 maximizes $F_1(\boldsymbol{u})$ and therefore $F(\boldsymbol{u})$ over S_1. Since $\bar{\boldsymbol{u}}$ is an extreme point of S_1, it is the solution of some linearly independent[2] subset E of n of the equations

$$\begin{array}{ll} u_1 - u_j = 0, \ j \in [2, n] & (a) \\ \boldsymbol{A}\boldsymbol{u} = \boldsymbol{B} & (b) \\ u_i = c_i, \ i \in [1, n] & (c) \\ u_i = d_i, \ i \in [1, n] & (d) \end{array}$$

Let T be the subset of equations in E that appear in (c) or (d). This means that $n - |T|$ variables are not fixed to one of their bounds by (c) and (d). Suppose that $m+1$ of these variables are not set equal to u_1 by (a). Then all of the nonzeros in the corresponding columns must occur in the m rows of (b). The $m+1$ columns must therefore be linearly dependent, which is impossible because is E is nonsingular. We conclude that at most m variables u_i are fixed neither to a bound by (c) and (d) nor to $u_1 = u_{\min}$ by (a). Since these u_is cannot be strictly between their lower bound c_i and $u_{\min}$, the proposition follows. □

Proof of Proposition 10. Let S_i, E and T be as in the previous proof, and let Q be the set of inequalities (a) in E. Then Q contains $|Q| + 1$ variables,

[2]We consider equations to be linearly independent when their coefficient rows are linearly independent.

and T contains $|T|$ variables. Since m variables are strictly between $u_{\min}$ and their upper bounds, these variables are in neither Q nor T, and E must therefore contain all m rows of (b). If Q and T have no variables in common, the number of variables is at least $|Q|+|T|+m+1$. But $|Q|+|T|+m+1 > n$ because $|Q|+|T|+m = n$ due to the linear independence of the equations in E. Thus one variable u_j belongs to both Q and T, which implies that $u_j = u_1 = u_{\min}$ and u_j is at its lower or upper bound. □

Corollary 11 is immediate.

Proof of Proposition 12. Let S be the feasible set of (11), and define

$$S_i = S \cap \{(\boldsymbol{y}, \boldsymbol{u}) \mid u_i \leq u_j, \text{ all } j\}$$

Since S is the union of all S_i, the maximum of $F(\boldsymbol{u})$ over some S_i is optimal in (11). Suppose without loss of generality that the maximum of $F(\boldsymbol{u})$ over S_1 is optimal in (11). For $(\boldsymbol{y}, \boldsymbol{u}) \in S_1$, the function $F(\boldsymbol{u})$ can be written

$$F_1(\boldsymbol{u}) = (n-1)\Delta + nu_1 + \sum_{i=1}^{n} \big(u_i - u_1 - \Delta\big)^+$$

Since $F_1(\boldsymbol{u})$ is convex, some extreme point $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{u}})$ of S_1 maximizes $F_1(\boldsymbol{u})$ and therefore $F(\boldsymbol{u})$ over S_1. Since $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{u}})$ is an extreme point of S_1, it is the solution of some linearly independent subset E of $n+r$ of the equations

$$\begin{array}{ll}
u_1 - u_j = 0, \ j = 2, \ldots, n & (a) \\
\boldsymbol{e}^{\mathsf{T}}\boldsymbol{y} = B & (b) \\
\boldsymbol{a}^k\boldsymbol{u} - y_k = B_k, \ k = 1, \ldots, r & (c) \\
u_i = c_i, \ i = 1, \ldots, n & (d) \\
u_i = d_i, \ i = 1, \ldots, n & (e) \\
y_k = 0, \ k = 1, \ldots, r & (f)
\end{array}$$

We first demonstrate (i). Let T be the subset of equations in E that appear in (d) or (e). This means that $n - |T|$ variables u_j are not fixed to one of their bounds by (d) and (e). Suppose that $r+1$ of these variables are not set equal to u_1 by (a). Then all of the nonzeros in the corresponding columns must occur in the r rows of (c). The $r+1$ columns must therefore be linearly dependent, which is impossible because is E is nonsingular. We conclude that at most r variables u_j are fixed neither to a bound by (d) and (e) nor to $u_1 = u_{\min}$ by (a). Since these u_js cannot be strictly between their lower bound c_j and $u_{\min}$, (i) follows.

We now demonstrate (ii). Let T be as before. This means that $n - |T|$ variables u_j are not fixed to one of their bounds by (d) and (e). Suppose for any given k that 2 of these variables that are in $\boldsymbol{u}^k$ are not set equal to u_1 by (a). Then all of the nonzeros in the corresponding columns must occur in row k of (c). The 2 columns must therefore be linearly dependent, which is impossible because is E is nonsingular. We conclude that at most one variable u_j in $\boldsymbol{u}^k$ is strictly between its lower bound c_j and $u_{\min}$, and (ii) follows. □

Proof of Proposition 13. Let S be the feasible set of (18), and define

$$S_i = S \cap \{\boldsymbol{u}^k \mid u_i \leq u_j, \text{ all } j \in I_k\}, \; i \in I_k$$

Since S is the union of all S_i, the maximum of $F_k(\boldsymbol{u}^k)$ over some S_i is optimal in (18). Suppose without loss of generality that the maximum of $F_k(\boldsymbol{u}^k)$ over S_1 is optimal in (18). For $\boldsymbol{u}^k \in S_1$, the function $F_k(\boldsymbol{u}^k)$ can be written

$$F_k^1(\boldsymbol{u}^k) = (n-k+1)u_1 + \sum_{i \in I_k} \big(u_i - u_1 - \Delta\big)^+$$

Since $F_k^1(\boldsymbol{u}^k)$ is convex, some extreme point $\bar{\boldsymbol{u}}^k$ of S_1 maximizes $F_k^1(\boldsymbol{u}^k)$ and therefore $F_k(\boldsymbol{u}^k)$ over S_1. Since $\bar{\boldsymbol{u}}^k$ is an extreme point of S_1, it is the solution of some linearly independent subset E of $n-k+1$ of the equations

$$\begin{aligned}
& u_1 - u_j = 0, \; j \in I_k \setminus \{1\} && (a) \\
& \sum_{j \in I_k} a_j u_j = B_k && (b) \\
& u_j = \max\{c_j, \; \bar{u}_{i_{k-1}}\}, \; j \in I_k && (c) \\
& u_j = d_j, \; j \in I_k && (d)
\end{aligned}$$

Let T be the subset of equations in E that appear in (c) or (d). This means that $n-k+1-|T|$ variables are not fixed to one of their bounds by (c) and (d). If two or more of these variables, say u_i and u_j, are not set equal to u_1 by (a), then the corresponding columns of E (which must be nonzero columns because E is linearly independent) have the form $(\mathbf{0}, a_i, \mathbf{0})$ and $(\mathbf{0}, a_j, \mathbf{0})$, with a_i and a_j at the same position in the two vectors. But these columns are linearly dependent, which is impossible because is E is linearly independent. We conclude that at most one variable u_j not fixed to one of its bounds by (c) and (d) is not set equal to $u_1 = u_{\min}$ by (a). Since u_j cannot be strictly between $\max\{c_j, \bar{u}_{i_{k-1}}\}$ and $u_{\min}$, the proposition follows. □

Proof of Proposition 14. Let S_1, u_1, E and T be as in the previous proof, and let Q be the set of inequalities (a) in E. Then Q contains $|Q|+1$ variables, and T contains $|T|$ variables. Since one variable is strictly between $u_{\min}$ and its upper bound, this variable is in neither Q nor T. If Q and T have no variables in common, then the number of variables is at least $|Q|+|T|+2$. But $|Q|+|T|+2 > n-k+1$ because $|Q|+|T|+1 = n-k+1$ due to the linear independence of the equations in E. Thus one variable u_j belongs to both Q and T, which implies that $u_j = u_1 = u_{\min}$ and u_j is at its lower or upper bound. $\square$