DM872

Mathematical Optimization at Work

Formulating Equity and Fairness in Optimization Models

Marco Chiarandini

Department of Mathematics & Computer Science
University of Southern Denmark

Outline

Optimization with Multiple Criteria

Let the mappings $f_i : X \to \mathbb{R}$ be criteria or objective functions:

$$\begin{aligned}&\text{“min”}(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}))\\ &\qquad\text{subject to } \boldsymbol{x} \in X\end{aligned}$$

$Y := f(X) := \{y \in \mathbb{R}^2 : y = f(x) \text{ for some } x \in X\}$ image of X under $f = (f_1, f_2)$ aka the image of the feasible set, or the feasible set in criterion space (the space from which the criterion values are taken).

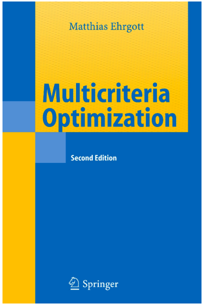
$\boldsymbol{y}_1 \preceq \boldsymbol{y}_2$ if $y_{1i} \leq y_{2i}\ \forall i = 1, 2$
$\boldsymbol{y}_1 \prec \boldsymbol{y}_2$ if $y_{1i} \leq y_{2i}\ \forall i = 1, 2$ and $\exists i = 1, 2 : y_{1i} < y_{2i}$

$X^* \subseteq X$ set of Pareto optimal solutions if $\forall \boldsymbol{x}^* \in X^* \quad \nexists x \in X : f(x) \prec f(x^*)$

Pareto Optimal Solutions

How do we find the set of Pareto Optimal solutions?

- ϵ-constraint method
- scalarization method



See: Matthias Ehrgott, Multicriteria Optimization, Second edition, 2005, Springer Berlin

Alternative ways to deal with Multiple Criteria

- lexicographic optimization
- weighted sums
- distance from ideal points

Outline

Modeling Equity

Growing interest in incorporating ethics-related criteria, including those that integrate efficiency and equity concerns, into optimization models.

Practical applications in:

- health care
- disaster management
- telecommunications
- facility location

Fair resource allocation.

Modeling Equity

Example: disaster recovery

- power restoration can focus on urban areas first (efficiency)
- this can leave rural areas without power for weeks/months
- happened in Puerto Rico after hurricane Maria (2017)

A more equitable solution

- ...would give some priority to rural areas without overly scarifying efficiency.

Modeling Equity

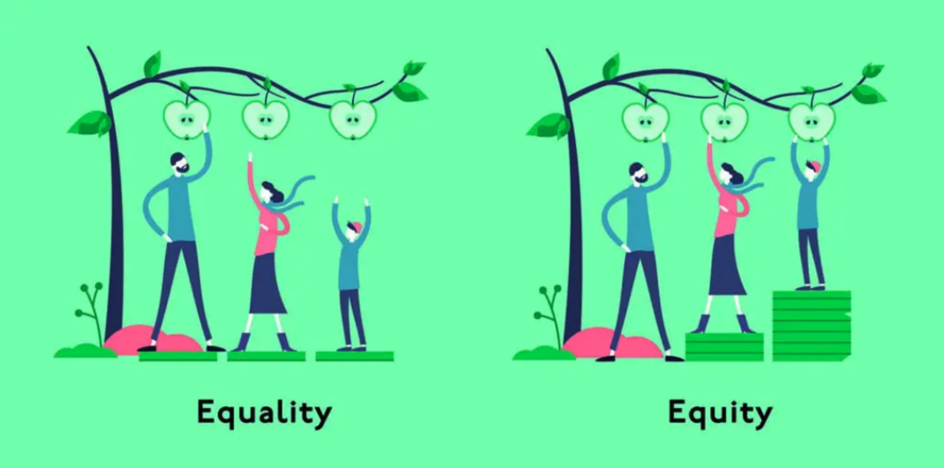
Mathematical formulation

- normally straightforward to reflect efficiency or cost in an objective function
- fairness can be understood in multiple ways, with no generally accepted method for representing any of them.

[Chen and Hooker, 2021] survey a wide range of formulations:

- described their mathematical properties
- indicate strength and weaknesses
- state what appears to be the most practical models
- so that one can select the formulation that best suites the practical application
- make the link with (computational) social choice theory

Equality vs Equity and Fairness



Equality vs Equity and Fairness

Equality vs fairness

Two views on ethical importance of equality:

Parfit 1997

- **Irreducible:** Inequality is inherently unfair. Scanlon 2003
- **Reducible:** Inequality is unfair only insofar as it reduces utility.

Frankfurt 2015

Possible problems with inequality measures:

- No preference for an identical distribution with **higher utility**.
- Even when average utility is fixed, no preference for reducing inequality at the **bottom** rather than the **top** of the distribution.

Modeling Equity

- Inequality measures
- Fairness for the disadvantaged (grounding in social choice theory)
- Combining efficiency and fairness — convex combinations
- Combining efficiency and fairness — classical methods
- Combining efficiency and fairness — threshold models
- Statistical bias metrics from machine learning

Methods

- inequality metrics,
- Rawlsian maximin and leximax criteria,
- convex combinations of these,
- alpha fairness and proportional fairness (the latter also known as the Nash bargaining solution),
- Kalai-Smorodinsky bargaining solution,
- utility-threshold and equity-threshold criteria for combining utilitarianism with maximin and leximax criteria.
- n-person model for the equity-threshold criterion.
- statistical fairness metrics
- demographic parity,
- equalized odds,
- accuracy parity,
- and predictive rate parity.

Generic Model

Given a model to maximize efficiency $f(\boldsymbol{x})$:

$$\max_{\boldsymbol{x}}\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in S_{\boldsymbol{x}}\}$$

we incorporate equity by formulating a fairness criterion as a social welfare function (SWF) of the individual utilities

$$W(\boldsymbol{u}) = W(u_1, \ldots, u_n)$$

- measures desirability of the magnitude and distribution of utilities across individuals
- utility can be wealth, health, negative cost, etc.
- the SWF becomes the objective function of the optimization model:

$$\max_{\boldsymbol{u},\boldsymbol{x}}\{W(\boldsymbol{u}) \mid \boldsymbol{u} = U(\boldsymbol{x}), \boldsymbol{x} \in S_{\boldsymbol{x}}\}$$

Generic Model

The social welfare optimization problem

Vector of utilities enjoyed by individuals

Vector of resource allocations to individuals

$$\max_{\boldsymbol{u},\boldsymbol{x}} \{W(\boldsymbol{u}) \mid \boldsymbol{u} = U(\boldsymbol{x}),\ \boldsymbol{x} \in S_{\boldsymbol{x}}\}$$

Social welfare function

Vector of utility functions $U_1, \ldots, U_n$

Problem constraints

Notation simplification:

$$\max_{\boldsymbol{u},\boldsymbol{x}} \{W(\boldsymbol{u}) \mid (\boldsymbol{u},\boldsymbol{x}) \in S\}$$

Also:

$$\max_{\boldsymbol{u},\boldsymbol{x}} \{f(\boldsymbol{x}) \mid W(\boldsymbol{u}) \geq LB, (\boldsymbol{u},\boldsymbol{x}) \in S\}$$

Example

Medical triage

- n patients requiring treatment
- c_i cost of treatment for patient i
- B limited budget
- u_i utility in quality-adjusted life years (QALY),
  $u_i = a_i$ without treatment, $u_i = a_i + b_i$ with treatment
- Task: allocate treatments in equitable and efficient way.
- binary variables x_i

$$
\begin{aligned}
\max \quad & W(\boldsymbol{u}) \\
& \sum_i c_i x_i \leq B \\
& u_i = a_i + b_i x_i \qquad \forall i \\
& x_i \in \{0, 1\} \qquad \forall i
\end{aligned}
$$

$W(\boldsymbol{u})$ should reflect how equity and effectiveness are balanced

Pigou-Dalton Condition

- The Pigou-Dalton condition checks whether a SWF reflects **equality**.
  - A utility transfer from a **better-off** individual to a **worse-off** individual **never decreases** social welfare.
  - **Problem:** such a transfer can **increase inequality** with respect to some other individuals.

Pigou-Dalton Condition

P-D transfer if $W(\boldsymbol{u} + \epsilon \boldsymbol{e}_i - \epsilon \boldsymbol{e}_j) \geq W(\boldsymbol{u})$ for any i, j and any $\epsilon > 0$ for which $u_i + \epsilon \leq u_j - \epsilon$, where $\boldsymbol{e}_i$, $\boldsymbol{e}_j$ are the ith and jth unit vectors, respectively.

A stricter form of the condition requires $W(\boldsymbol{u} + \epsilon \boldsymbol{e}_i - \epsilon \boldsymbol{e}_j) > W(\boldsymbol{u})$.

While a pairwise Pigou–Dalton transfer reduces inequality between two individuals, it may increase inequality between those individuals and others.

Chateuneuf-Moyes Condition

- Addresses weakness of Pigou-Dalton condition.
  - A utility transfer from **top of distribution** to **bottom of distribution** never decreases social welfare.
  - Loss/gain due to transfer is distributed equally in each class.

Chateauneuf & Moyes 2006

Chateuneuf-Moyes Condition

A C-M transfer is a transfer of utility from $\boldsymbol{u}$ to $\boldsymbol{u}'$ such that $u_1 \leq \ldots \leq u_n$ as well as $u'_1 \leq \ldots \leq u'_n$, and for some pair of integers ℓ, h with $1 \leq \ell < h \leq n$, we have $u_\ell < u_h$ and

$$\boldsymbol{u}' = \boldsymbol{u} + \frac{\epsilon}{\ell} \sum_{i=1}^{\ell} \boldsymbol{e}_i - \frac{\epsilon}{n-h+1} \sum_{i=h}^{n} \boldsymbol{e}_i$$

for some $\epsilon > 0$.

A SWF $W(\boldsymbol{u})$ satisfies the C-M condition if C-M transfers never decreases social welfare. That is, $W(\boldsymbol{u}') \geq W(\boldsymbol{u})$ for any C-M transfer from $\boldsymbol{u}$ to $\boldsymbol{u}'$.

A C-M transfer does not incur th P-D problem, because the donor and recipient classes respectively lie completely above and below the rest of the population.

Inequality measures

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Relative range	yes	yes	yes	no
Relative mean deviation	yes	yes	yes	no
Coefficient of variation	yes	yes	no	no
Gini coefficient	yes	yes	yes	no
Hoover index	yes	yes	yes	no

Fairness for the disadvantaged

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Maximin (Rawlsian)	yes	yes	yes	no
Leximax (lexicographic)	yes	yes	yes	no
McLoone index	no	yes	yes	yes

P-D = Pigou-Dalton
C-M = Chateauneuf-Moyes
Linear = all constraints linear
Discrete = some variables discrete

Combining efficiency & fairness
Convex combinations

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Utility + Gini coefficient	yes	yes	no	no
Utility * Gini coefficient	yes	yes	yes	no
Utility + maximin	yes	yes	yes	no

Combining efficiency & fairness
Classical methods

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Alpha fairness	yes	yes	yes	no
Proportional fairness (Nash bargaining)	yes	yes	yes	no
Kalai-Smorodinsky bargaining	yes	yes	no	no

P-D = Pigou-Dalton
C-M = Chateauneuf-Moyes
Linear = all constraints linear
Discrete = some variables discrete

Combining efficiency & fairness
Threshold methods

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Utility + maximin – Utility threshold	no	yes	yes	yes
Utility + maximin – Equity threshold	yes	yes	yes	no
Utility + leximax – Predefined priorities	no	no	yes	yes
Utility + leximax – No predefined priorities	no	yes	yes	yes

Statistical fairness metrics

Criterion	*P-D?*	*C-M?*	*Linear?*	*Discrete?*
Demographic parity			yes	no
Equalized odds			yes	no
Accuracy parity			yes	no
Predictive rate parity			no	yes

P-D = Pigou-Dalton
C-M = Chateauneuf-Moyes
Linear = all constraints linear
Discrete = some variables discrete

Outline

Inequality measures

- Relative range
- (Relative mean deviation)
- Coefficient of variation
- Gini coefficient
- (Hoover index)

All dispersion measures are normalized by the mean utility so as to be invariant under rescaling of utilities.

Relative range

Relative range

$$W(\boldsymbol{u}) = -\frac{u_{max} - u_{min}}{\bar{u}}$$

Rationale:

- Perceived inequality is relative to the best off
- So, move everyone closer to the best off

Problem:
Ignores distribution between extremes

Relative range

Linearization via linear-fractional programming (Charnes and Cooper 1962), see next slide:
Let $\boldsymbol{u} = \mathbf{u}'/t$ and $\boldsymbol{x} = \mathbf{x}'/t$:

$$\min_{\substack{\boldsymbol{x}',\boldsymbol{u}',t\\ u'_{\min},u'_{\max}}} \left\{ u'_{\max} - u'_{\min} \;\middle|\; \begin{array}{c} u'_{\min} \le u'_i \le u'_{\max},\ \text{all } i \\ \bar{u}' = 1,\ t \ge 0,\ (\boldsymbol{u}', \boldsymbol{x}') \in S' \end{array} \right\}$$

where t, u'_{min}, u'_{max} are new variables.

Digression: Linear-Fractional Programming

Formally, a linear-fractional program is defined as the problem of maximizing (or minimizing) a ratio of affine functions over a polyhedron,

$$\begin{array}{rl} \text{maximize} & \dfrac{\boldsymbol{c}^T\boldsymbol{x}+\alpha}{\boldsymbol{d}^T\boldsymbol{x}+\beta} \\ \text{subject to} & A\boldsymbol{x} \leq \boldsymbol{b} \end{array}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ represents the vector of variables to be determined, $\boldsymbol{c}, \boldsymbol{d} \in \mathbb{R}^n$, $\boldsymbol{b} \in \mathbb{R}^m$ are vectors of (known) coefficients, $A \in \mathbb{R}^{m\times n}$ is a (known) matrix of coefficients and $\alpha, \beta \in \mathbb{R}$ are constants.

The constraints have to restrict the feasible region to $\{\boldsymbol{x} | \boldsymbol{d}^T\boldsymbol{x}+\beta > 0\}$, i.e. the region on which the denominator is positive. Alternatively, the denominator of the objective function has to be strictly negative in the entire feasible region.

Digression: Linear-Fractional Programming: Transformation to a linear program

Under the assumption that the feasible region is non-empty and bounded, the Charnes-Cooper transformation

$$\boldsymbol{y} = \frac{1}{\boldsymbol{d}^T \boldsymbol{x} + \beta} \cdot \boldsymbol{x} \, ; \quad t = \frac{1}{\boldsymbol{d}^T \boldsymbol{x} + \beta}$$

translates the linear-fractional program to the equivalent linear program:

$$\begin{aligned} \text{maximize} \quad & \boldsymbol{c}^T \boldsymbol{y} + \alpha t \\ \text{subject to} \quad & A\boldsymbol{y} \leq \boldsymbol{b} t \\ & \boldsymbol{d}^T \boldsymbol{y} + \beta t = 1 \\ & t \geq 0. \end{aligned}$$

Then the solution for $\boldsymbol{y}$ and t yields the solution of the original problem as

$$\boldsymbol{x} = \frac{1}{t} \boldsymbol{y}.$$

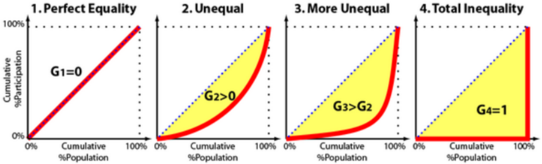
Coefficient of variation

Non linear.

$$W(\boldsymbol{u}) = -\frac{1}{\bar{u}}\left[\frac{1}{n}\sum_i (u_i - \bar{u})^2\right]^{\frac{1}{2}}$$

Gini Index

Used to measure income/wealth inequality.
Lorenz curve shows for the bottom x% of individuals, what percentage (y%) of the total utility they have



The SWF is $W(\boldsymbol{u}) = -G(\boldsymbol{u})$, where

$$G(\boldsymbol{u}) = \frac{1}{2\bar{u}n^2} \sum_{i,j} |u_i - u_j|$$

It can be linearized:

$$\min_{\boldsymbol{x}',\boldsymbol{u}',V,t} \left\{ \frac{1}{2n^2} \sum_{i,j} v_{ij} \;\middle|\; \begin{array}{l} -v_{ij} \le u'_i - u'_j \le v_{ij}, \ \text{all } i,j \\ \bar{u}' = 1, \ t \ge 0, \ (\boldsymbol{u}', \boldsymbol{x}') \in S' \end{array} \right\}$$

Outline

1. Multiobjective Optimization
2. Modeling Fairness
3. Inequality measures
4. Fairness of the Disadvantaged
5. Combinations

Fairness for the Disadvantaged

Maximin

$$W(\boldsymbol{u}) = \min_i \{u_i\}$$

Rationale:

- Based on **difference principle** of John Rawls.
- Inequality is justified only to the extent that it increases the utility of the worst-off.
- Originally intended only for the design of **social institutions** and distribution of **primary goods** (goods that any rational person would want).
- Can be adopted as a general principle of equity: maximize the minimum utility.

Rawls 1971, 1999

Fairness for the Disadvantaged

Maximin

$$W(\boldsymbol{u}) = \min_i \{u_i\}$$

Social contract argument:

- We decide on social policy in an "original position," behind a "veil of ignorance" as to our position on society.
- All parties must be willing to **endorse** the policy, no matter what position they end up assuming.
- No rational person can endorse a policy that puts him/her on the **bottom** of society – unless that person would be even **worse off** under another social arrangement.
- Therefore, an agreed-upon social policy must maximize the welfare of the worst-off.

Leximax

The maximin criterion can be plausibly extended to lexicographic maximization (leximax)
Leximax is achieved by first maximizing the smallest utility subject to resource constraints, then the second smallest, and so forth.
A leximax solution can computed by solving a sequence of optimization problems

$$\max_{\boldsymbol{x},\boldsymbol{u},w} \left\{ w \;\middle|\; \begin{array}{c} w \le u_i,\; u_i \ge \hat{u}_{i_{k-1}},\; i \in I_k \\ (\boldsymbol{u},\boldsymbol{x}) \in S \end{array} \right\} \tag{5}$$

for $k = 1, \ldots, n$, where $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{u}})$ is an optimal solution of problem k, $\hat{u}_{i_0} = -\infty$, and i_k is defined so that

$$\hat{u}_{i_k} = \min_{i \in I_k}\{\hat{u}_i\}, \text{ with } I_k = \{1, \ldots, n\} \setminus \{i_1, \ldots, i_{k-1}\}$$

McLoone Index

The McLoone index compares the total utility of individuals at or below the median utility to the utility they would enjoy if all were brought up to the median utility.

The index is 1 if nobody's utility is strictly below the median, and it approaches 0 if the utility distribution has a very long lower tail (on the assumption that all utilities are positive.)

rewarding equality in the lower half of the distribution, but it is unconcerned by the existence of very rich individuals in the upper half.

$$W(\boldsymbol{u}) = \frac{1}{|I(\boldsymbol{u})|\tilde{u}} \sum_{i \in I(\boldsymbol{u})} u_i$$

where $\tilde{u}$ is the median of utilities in $\boldsymbol{u}$ and $I(\boldsymbol{u})$ is the set of indices of utilities at or below the median, so that $I(\boldsymbol{u}) = \{i \mid u_i \leq \tilde{u}\}$.

We can formulate the McLoone index optimization problem as a mixed integer programming (MIP) problem with a fractional objective function, by using standard "big-M" modeling techniques from integer programming. The model uses 0–1 variables δ_i, where $\delta_i = 1$ when $i \in I(\boldsymbol{u})$. The constant M is a large number chosen so that $u_i < M$ for all i. The model is

$$\max_{\substack{\boldsymbol{x},\boldsymbol{u},m \\ \boldsymbol{y},\boldsymbol{z},\boldsymbol{\delta}}} \left\{ \frac{\sum_i y_i}{\sum_i z_i} \;\middle|\; \begin{array}{c} m - M\delta_i \leq u_i \leq m + M(1-\delta_i), \text{ all } i \\ y_i \leq u_i, \; y_i \leq M\delta_i, \; \delta_i \in \{0,1\}, \text{ all } i \\ z_i \geq 0, \; z_i \geq m - M(1-\delta_i), \text{ all } i \\ \sum_i \delta_i \leq n/2, \; (\boldsymbol{u},\boldsymbol{x}) \in S \end{array} \right\}$$

where the new variable m represents the median, variable y_i is u_i if $\delta_i = 1$ and 0 otherwise, and variable z_i is m if $\delta_i = 1$ and 0 otherwise in the optimal solution. The objective function can be linearized by using the same change of variable as in linear-fractional programming:

$$\max_{\substack{\boldsymbol{x}',\boldsymbol{u}',m' \\ \boldsymbol{y}',\boldsymbol{z}',t,\boldsymbol{\delta}}} \left\{ \sum_i y'_i \;\middle|\; \begin{array}{c} u'_i \geq m' - M\delta_i, \text{all } i \\ u'_i \leq m' + M(1-\delta_i), \text{ all } i \\ y'_i \leq u'_i, \; y'_i \leq M\delta_i, \; \delta_i \in \{0,1\}, \text{ all } i \\ z'_i \geq 0, \; z'_i \geq m' - M(1-\delta_i), \text{ all } i \\ \sum_i z'_i = 1, \; t \geq 0 \\ \sum_i \delta_i \leq n/2, \; (\boldsymbol{u}',\boldsymbol{x}') \in S' \end{array} \right\}$$

The model is an MILP problem when the constraints defining S are linear.

Outline

Convex Combination

$$F(\boldsymbol{u}) = (1-\lambda)\sum_i u_i + \lambda\Phi(\boldsymbol{u})$$

Utility & Fairness – Convex Combinations

Utility + Gini coefficient

$$W(\boldsymbol{u}) = (1-\lambda)\sum_i u_i + \lambda\big(1 - G(\boldsymbol{u})\big)$$

Rationale.

- Takes into account both efficiency and equity.
- Allows one to adjust their relative importance.

Problem.

- Combines utility with a dimensionless quantity.
- How to interpret λ, or choose a λ for a given application?
- Choice of λ is an issue with convex combinations in general.

Utility and maximin

utility with the Rawlsian maximin criterion by using the convex combination

$$W(\boldsymbol{u}) = (1-\lambda)\sum_i u_i + \lambda \min_i\{u_i\}$$

Rationale
- combines quantities that are measured in the same units.

Problem
- again unclear how to select a suitable value of λ.

Note that if we index utilities so that $u_1 \leq \ldots \leq u_n$, is simply a weighted sum $u_1 + (1-\lambda)\sum_{i>1} u_i$ that gives somewhat more weight to the lowest utility.

Following this path:

$$W(u) = \sum_i w_i u_i$$

with gradually decreasing weights $w_1 > w_2 > \ldots > w_n$ to the utilities. Yet modeling challenge of ensuring that weight w_i is assigned to the ith smallest utility.