Optimization Letters (2020) 14:1533–1541
https://doi.org/10.1007/s11590-020-01568-1

ORIGINAL PAPER

Fairness criteria for allocating scarce resources

Bismark Singh[1]

Received: 15 June 2019 / Accepted: 4 March 2020 / Published online: 13 March 2020

Abstract
We develop an optimization model to provide a fair allocation of multiple resources to multiple users. All resources might not be suitable to all users. We develop a notion of fairness, and then provide a general class of functions achieving it. Next, we develop more restricted notions of fairness—special cases of which exist in literature. Finally, we distinguish between scarce and abundant resources, and show that if a resource is abundant, all users seeking it achieve the maximum possible coverage.

Keywords Proportional fairness · Equity · Optimization · Welfare · Resource allocation · KKT conditions

Notation

Sets and indices

$i \in I$	Types of users
$k \in K$	Types of resources
$i \in I_k$	Subset of users eligible to receive resource k
$k \in K_i$	Subset of resources eligible for user i

Data and parameters

n_i	Population of user i, $n_i > 0$
b_k	Amount of resource k, $b_k > 0$
f_i	Original coverage of user i, $0 \leq f_i < 1$
w_i	Weight of user i, $w_i > 0$

Decision variables

x_{ik}	Amount of resource k allocated to user i
y_i	Final coverage of user i

✉ Bismark Singh
bismark.singh@fau.de

1 Discrete Mathematics, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany

1 Background

We study the problem of allocating different kinds of substitutable resources to different populations seeking them. This is related to a fundamental problem in economics—how should different resources be produced and made available to agents; see, e.g., [10]. Our aim is to allocate resources in a *fair* manner, a term we make precise below, and we develop a class of optimization models that achieve specific notions of fairness. This work finds applications in resource allocation problems that have been studied in specific settings; see, e.g., distribution of coal among power companies [2], several military and defense examples [11], multiperiod manufacturing of high-tech products [12], wireless networks [9], healthcare [6], education [3], and conservation of threatened species [8]. Special cases of this general problem include the so-called waterfilling algorithm; see, e.g., [13]. We present general results, based on the KKT optimality conditions, that provide essential conditions for fair allocation of resources, as well as special cases in which resources are abundant. For an introduction to resource allocation problems, see, e.g., [7].

An excellent introduction to different concepts in fairness is available in Bertsimas et al. [1]. Similar to our study, the authors consider a central decision planner seeking to allocate resources in a “fair” manner. The authors describe several concepts of fairness based on utilities of each user, and primarily distinguish between max-min fairness and proportional fairness. Our work differs from this existing body of literature in the following sense. We do not seek to maximize utilities, but rather to ensure equitable coverages (or, shortages) when allocating scarce resources with varying priorities (or, claims) for users. Bertsimas et al. [1] call this as an Aristotle’s equity principle, which they do not analyze; see also [4,14]. To this end, we develop mathematical models and analyze their properties under specific notions of fairness that are significantly different from the existing literature. As with Aristotle’s equity principle, our models serve to answer the question: how should allocation decisions be conducted when resources are due to users with pre-existing claims?

2 Mathematical models for fairness

We consider a set of user types (henceforth simply users), $i \in I$, seeking a set of resources, $k \in K$. Define a subset of resources, $k \in K_i$, as the set of resources that user i is eligible to receive. Define a subset of users, $i \in I_k$, as the set of users that are eligible to receive resource k. Define the population of user i and an associated weight as $n_i > 0$ and $w_i > 0$, $\forall i \in I$, respectively, and the availability of resource k as $b_k > 0$, $\forall k \in K$. The prior coverage of user i is denoted by $0 \leq f_i < 1$, where we assume f_i is defined relative to the same population of size n_i. Let $x_{ik} \geq 0$ be the amount of discretionary resource k allocated to user i. We denote the final coverage, after allocation of all resources, by y_i, where $f_i \leq y_i \leq 1$. Our resources are partially substitutable in the sense that user i is equally able to make use of any resource in K_i, but can make no use of a resource in $K \setminus K_i$.

Consider the following optimization model:

$$\min_{x,y} \quad \sum_{i\in I} w_i n_i F(y_i) \tag{1a}$$

$$\text{s.t.} \quad \sum_{i\in I_k} x_{ik} \le b_k, \forall k \in K \quad : \nu_k \tag{1b}$$

$$f_i + \frac{1}{n_i}\sum_{k\in K_i} x_{ik} = y_i, \forall i \in I : \alpha_i \tag{1c}$$

$$y_i \le 1, \forall i \in I \quad : \lambda_i \tag{1d}$$

$$x_{ik} \ge 0, \forall k \in K_i, i \in I, \quad : -\mu_{ik}, \tag{1e}$$

where $F : [0, 1] \to \mathbb{R}$ is a loss function. The dual multipliers indicated in constraints (1b)–(1e) satisfy $\nu_k \ge 0$, $\lambda_i \ge 0$ and $\mu_{ik} \ge 0$, respectively. Model (1) provides an optimal allocation, denoted x^*_{ik}, and, in turn, the optimal final coverage of each user population, denoted $y^*_i = f_i + n_i^{-1}\sum_{k\in K_i} x^*_{ik}$. The objective function in model (1) seeks to minimize the weighted sum of the users' loss values. Constraint (1b) ensures that the allocated resources of type k do not exceed what is available. Constraints (1c) and (1d) indicate that the maximum final coverage is at most 1 for every user type. Constraint (1e) ensures non-negative allocation of resources. In the rest of this article, we drop the superscript, $(\cdot)^*$, on the optimal solution for simplicity. We start with the following notion of priority which is fundamental to our notions of fairness.

Definition 1 (*Priority*) In an optimal solution to model (1) if $w_{i'} \ge w_i$, then i' is said to have a *priority* over i. If i' has a priority over i and $y_{i'} \le y_i$ (with at least one inequality strict), then i' has an *inappropriate priority*, denoted by $i' \gg i$; while, if i' has a priority over i and $y_{i'} \ge y_i$ (with at least one inequality strict) then i has an *appropriate priority*, denoted by $i' \ggg i$.

In the above definition, "with at least one inequality strict" means that both the weights and the coverages cannot be equal. Definition 1 distinguishes two kinds of priorities. "Inappropriate" can be understood as follows: a user has a larger weight, but receives a lower coverage; "Appropriate" refers to a larger weight and a larger coverage. Excluding the trivial case of $w_i = w_{i'}$, $y_i = y_{i'}$, we have the following four exclusive cases (i) $i' \gg i$, (ii) $i \gg i'$, (iii) $i' \ggg i$, or (iv) $i \ggg i'$. In Sect. 2.1, we study the first two cases, while in Sect. 2.2 we study the latter two.

2.1 Fairness: inappropriate priorities

Definition 2 (*Fairness*) In an optimal solution to model (1) if $i' \gg i$ implies $x_{ik} = 0$ for all $k \in K_i \cap K_{i'}$ and all $i, i' \in I$ with $y_{i'} < 1$, then the solution is said to be *fair*.

We can interpret Definition 2 by understanding whether the inappropriately prioritized user i' can "complain" about the allocation given to user i; i.e., the distribution being not fair. According to Definition 2, user i' can have a justifiable complain only if: (a) user i' has coverage less than 100%, (b) there is a user, i, over which i' has an inappropriate priority, and (c) user i received positive allocation from a resource shared with user i'. The following theorem provides a class of functions, $F(\cdot)$, that provide fairness for model (1).

Theorem 1 *If $F : [0, 1] \to \mathbb{R}$ is continuously differentiable, decreasing, and strictly convex, then optimization model (1) achieves fairness.*

Proof The feasible region of model (1) is nonempty and compact, and hence the model has a finite optimal solution. The constraints are linear, and F is differentiable and convex, and so an optimal solution is characterized by the KKT conditions of primal feasibility, i.e., constraints (1b)–(1e) along with multipliers that satisfy the following conditions:

$$\lambda_i + \alpha_i + w_i n_i F'(y_i) = 0, \quad \forall i \in I \tag{2a}$$

$$\nu_k - \frac{\alpha_i}{n_i} - \mu_{ik} = 0, \quad \forall k \in K_i, i \in I \tag{2b}$$

$$\nu_k \geq 0,\ \forall k \in K,\ \lambda_i \geq 0, \quad \forall i \in I,\ \mu_{ik} \geq 0,\ \forall k \in K_i, i \in I \tag{2c}$$

and

$$\nu_k \left(\sum_{i \in I_k} x_{ik} - b_k \right) = 0, \quad \forall k \in K \tag{3a}$$

$$\alpha_i \left(f_i + \frac{1}{n_i} \sum_{k \in K_i} x_{ik} - y_i \right) = 0, \quad \forall i \in I \tag{3b}$$

$$\lambda_i (y_i - 1) = 0, \quad \forall i \in I \tag{3c}$$

$$\mu_{ik} x_{ik} = 0, \quad \forall k \in K_i, i \in I. \tag{3d}$$

Because F is strictly convex and decreasing, we have

$$w_i F'(y_i) \geq w_{i'} F'(y_{i'}), \text{ with strict inequality if either } y_i > y_{i'} \text{ or } w_{i'} > w_i. \tag{4}$$

Assume that the optimal solution is not fair. Hence, there exists at least one user pair, with $i' \gg i$, $y_{i'} < 1$, and with $k \in K_i \cap K_{i'} \neq \emptyset$ such that $x_{ik} > 0$. From the complementary slackness conditions (3c) and (3d), we have $\lambda_{i'} = 0$ and $\mu_{ik} = 0$.

From Eqs. (2a) and (2b), we obtain $w_i F'(y_i) = -\nu_k - \frac{\lambda_i}{n_i}$ and $w_{i'} F'(y_{i'}) = \mu_{i'k} - \nu_k$. Thus, $w_i F'(y_i) + \frac{\lambda_i}{n_i} + \mu_{i'k} = w_{i'} F'(y_{i'})$, and because $\lambda_i, \mu_{i'k} \geq 0$, we have $w_i F'(y_i) \leq w_{i'} F'(y_{i'})$. Given that $y_i > y_{i'}$ or $w_{i'} > w_i$, inequality (4) implies the contradiction of $w_i F'(y_i) > w_{i'} F'(y_{i'})$. □

Examples of functions that satisfy the hypothesis of Theorem 1 include $F(z) = (1 - z)^m$, where $m > 1$; $F(z) = -\ln(z + \varepsilon)$, where $\varepsilon > 0$; and, $F(z) = e^{-z}$.

2.2 Proportional fairness: appropriate priorities

The previous notion of fairness covers the case of inappropriate priorities. A stronger notion of fairness can be developed by a slight change in the definition of fairness. This encompasses the case when the prioritized user i' receives a larger coverage than i; i.e., the priority is appropriate. We term this notion as *proportional fairness*. Before we explain the motivation for this term, we define another term as follows.

Definition 3 (*Balance*) An optimal solution to model (1) satisfying $w_i F'(y_i) = w_{i'} F'(y_{i'})$ is said to be *balanced* between users i and i'.

In other words, a balance between two users requires their marginal losses to be in inverse proportion to their weights. The following lemma proves that excluding the trivial case of $w_i = w_{i'}$, $y_i = y_{i'}$, only an appropriate priority could result in a balanced coverage.

Lemma 1 *Let $F : [0, 1] \to \mathbb{R}$ is continuously differentiable, decreasing, and strictly convex. Then, an optimal solution to model (1) with an inappropriate priority cannot be balanced.*

Proof Since $w_i, w_i' > 0$ and F is decreasing, $w_i F'(y_i) = w_{i'} F'(y_{i'})$ cannot hold if either i or i' has an inappropriate priority. □

Definition 4 (*Proportional fairness*) Consider a balanced optimal solution to model (1). If $y_{i'} < 1$ and $x_{ik} > 0$ for $k \in K_i \cap K_{i'}$ implies and is implied by $y_i < 1$ and $x_{i'k} > 0$, then the solution is said to be *proportionally fair* between user i and i'.

This stronger notion encompasses the situation when a user has an appropriate priority, did not receive 100% coverage, and yet the other user received a positive allocation of the shared resource. According to Definition 4, a user i' can have a justifiable complain only if: (a) i' has coverage less than 100%, (b) another user, i, received a positive allocation from a resource shared with i', (c) the coverages of the two users are balanced, and (d) i' did not receive any allocation of this shared resource or user i received 100% coverage. We note from Lemma 1 that statement (c) can hold only for the case of appropriate priorities.

The following theorem shows that the same class of functions as Theorem 1, but with an additional condition, provide proportional fairness for model (1).

Theorem 2 *Let $F : [0, 1] \to \mathbb{R}$ is continuously differentiable, decreasing, and strictly convex. If an optimal solution to model (1) is balanced and satisfies strict complementary slackness, then it is proportionally fair.*

Proof The proof is similar to that of Theorem 2. Equations (2a) and (2b) yield

$$w_i F'(y_i) + \frac{\lambda_i}{n_i} - \mu_{ik} = w_{i'} F'(y_{i'}) + \frac{\lambda_{i'}}{n_{i'}} - \mu_{i'k}.$$

From Definition 4, consider an (i, i') pair satisfying $y_{i'} < 1$ and $x_{ik} > 0$ for $k \in K_i \cap K_{i'}$. We prove that $y_i < 1$ and $x_{i'k} > 0$. Because the optimal solution satisfies strict complementary slackness $y_{i'} < 1$ implies $\lambda_{i'} = 0$ and $x_{ik} > 0$ implies $\mu_{ik} = 0$. Since the solution is balanced, we have $\mu_{i'k} = -\frac{\lambda_i}{n_i}$. Because the dual multipliers are non-negative, strict complementary slackness further implies $y_i < 1$ and $x_{i'k} > 0$. □

Most interior point algorithms yield solutions that satisfy strict complementary slackness; see, e.g., [16]. Also note, the assumption of differentiability ensures we can compute the gradients of F, and the assumption of convexity on F ensures the KKT conditions are necessary and sufficient. We thus have the following central results for model (1) for the chosen class of F:

1. An optimal solution with an inappropriate priority is fair.
2. An optimal balanced solution with an appropriate priority that satisfies strict complementary slackness is proportionally fair.

2.3 Scarce and abundant resources

We define a resource k to be abundant if $\sum_{i \in I_k} x_{ik}$ is strictly less than b_k, and scarce if $\sum_{i \in I_k} x_{ik}$ is equal to b_k . Since F is decreasing, we expect that if b_k is sufficiently large, all users sharing resource k achieve 100% coverage; i.e., $y_i = 1, \forall i \in I_k$. Since this is the maximum possible coverage, no user should complain; the hypotheses of both Definitions 2 and 4 are not satisfied and there is no concept of fairness. However, if there exists even a single user with less than 100% coverage, then we expect a complete allocation of all resources the user is eligible for. And then, the user can justifiably complain and seek fairness. The following theorem makes this precise.

Theorem 3 *Abundant resources: In an optimal solution to model (1), if resource k is not completely allocated, i.e.,* $\sum_{i \in I_k} x_{ik} < b_k$*, then* $y_i = 1, \forall i \in I_k$*. Scarce resources: In an optimal solution to model (1), if* $y_i < 1$ *for some* $i \in I$*, then* $\sum_{i \in I_k} x_{ik} = b_k, \forall k \in K_i$.

Proof *Abundant resources*: Consider a resource $k \in K$ such that $\sum_{i \in I_k} x_{ik} < b_k$. From Eq. (3a), we have $\nu_k = 0$. Assume that there exists some $i \in I_k$ with $y_i < 1$. Then, by Eq. (3c), we have $\lambda_i = 0$. From Eqs. (2a) and (2b) we have for this $(i, k) : \mu_{ik} = w_i F'(y_i)$. Since $F'(y_i) < 0$ and $w_i > 0$, we have $\mu_{ik} < 0$ which is a contradiction of Eq. (2c).

Scarce resources: Assume that there exists some $i \in I_k$ with $y_i < 1$. Then, by Eq. (3c), we have $\lambda_i = 0$. From Eqs. (2a) and (2b), we have for all k in K_i: $\mu_{ik} - \nu_k = w_i F'(y_i)$. Since $w_i > 0$, $F'(y_i) < 0$ and $\mu_{ik} \geq 0$, we have $\nu_k > 0$. From Eq. (3a) we have $\sum_{i \in I_k} x_{ik} = b_k, \forall k \in K_i$. □

The proof of Theorem 3 uses the KKT conditions for model (1). However, we can also prove the same without requiring the assumptions of convexity or differentiability. Consider an optimal solution, (x^*, y^*), and resource k' is abundant. Assume that $y_{i'} < 1$ for some $i' \in I_{k'}$. We show that this is not possible. Consider another solution, (x', y'), as follows: $x'_{i'k'} = x^*_{i'k'} + \delta$ and $x'_{ik} = x^*_{ik}$ for all other (i, k) pairs; $y'_{i'} = y'_{i'} + \frac{\delta}{n_{i'}}$ and $y'_i = y^*_i$ for all other i. Here $0 < \delta < \min\{b_{k'} - \sum_{i \in I_{k'}} x^*_{ik'}, n_{i'}(1 - y^*_{i'})\}$. Then, this solution is feasible to model (1). Further, since F is decreasing and $y'_{i'} > y^*_{i'}$, we

Table 1 Mapping between model (1) and model (3.1) of [15]

Model (1)	Model (3.1) of [15]
$i \in I$	$(i, j) \in (I \times J)$
$k \in K$	$k \in K$
$i \in I_k$	–
$k \in K_i$	$k \in K_j$
n_i	n_{ij}
b_k	b_k
f_i	f_{ij}, or $\frac{\sum_{k \in K_j} m_{ijk}}{n_{ij}}$
w_i	w_{ij}
x_{ik}	Q_{ijk}
y_i	$f_{ij} + \frac{\sum_{k \in K_j} Q_{ijk}}{n_{ij}}$, or $1 - s_{ij}$

have $F(y'_{i'}) < F(y^*_{i'})$. This contradicts the hypothesis that (x^*, y^*) is optimal. The analogous proof for scarce resources mirrors this.

3 A motivating example

This research was motivated by a collaboration with the Texas Department of State Health Services, aimed at allocating a limited amount of vaccines in an equitable manner during an influenza pandemic [5,6,15]. Here, we use model (3.1) from [15] to allocate four types of vaccines to five different population groups seeking them. In this example, vaccines are scarce at the start of the pandemic. The allocation ensures an equitable access to vaccines across all five population priority groups in 254 different counties of Texas. Vaccine types constitute our "resources" and the five priority groups across the 254 counties of Texas constitute our "users". Table 1 presents a map between our notation and that of [15].

Singh's optimization model in [15] presents a special case of model (1), with $F(y_i) = (1 - y_i)^2$. A similar loss function is used in [5,6]. This loss function, $F(z) = (1 - z)^2$, satisfies the hypothesis of Theorem 2 and hence achieves fairness for inappropriately prioritized users per Definition 2. Next, Theorem 3.2.1 of [15] provides sufficient conditions for a balanced solution $w_i F'(y_i) = w_{i'} F'(y_{i'})$: (i) $y_i < 1$ and $y_{i'} < 1$, and (ii) $x_{ik} > 0$ and $x_{i'k} > 0$ for some $k \in K_i \cap K_{i'}$. Then, the solution is proportionally fair per Definition 4 for the given loss function. Singh [15] thus proves the special cases of inappropriate priorities and appropriate priorities with balance, using the quadratic loss function.

4 Conclusion

We develop a general optimization model to study the allocation of partially substitutable resources to different classes of users seeking them. Not all resources are suitable to all classes of users, and different classes of users have different priorities.

When a resource is abundant, all users seeking it achieve the maximum possible coverage. When resources are scarce, we provide different notions of fairness depending on the priorities between users. And, we provide a class of objective functions as well as restrictions that achieve these notions of fairness.

Inappropriate priorities always result in fairness. Appropriate priorities, with two additional restrictions of balance and complementary slackness, always result in proportional fairness. Future research could establish what happens when these two restrictions are not met, as well as sufficient and/or necessary conditions to guarantee a balanced solution.

Finally, we demonstrate how our model is reducible to an existing model for equitable allocation of vaccines during an influenza pandemic.

Acknowledgements Open Access funding provided by Projekt DEAL. I sincerely thank David Morton for helpful and long discussions as well as for structuring this article. I also thank an anonymous reviewer for correcting an error in an earlier version of a proof. Anything overlooked in the manuscript is, of course, my responsibility.

References

1. Bertsimas, D., Farias, V.F., Trichakis, N.: The price of fairness. Oper. Res. **59**(1), 17–31 (2011)
2. Brown, J.R.: The sharing problem. Oper. Res. **27**(2), 324–340 (1979)
3. Byrne, E.M.: Planning and educational inequality: a study of the rationale of resource-allocation. Tech. rep., ERIC (1974)
4. Hoffman, R.L.: Defining, measuring, and implementing equity. Tech. rep., Metron Aviation (2003)
5. Huang, H.C.: Stockpiling and resource allocation for influenza preparedness and manufacturing assembly. Ph.D. thesis, The University of Texas at Austin (2014)
6. Huang, H.C., Singh, B., Morton, D.P., Johnson, G.P., Clements, B., Meyers, L.A.: Equalizing access to pandemic influenza vaccines through optimal allocation to public health distribution points. PloS One **12**(8), e0182720 (2017)
7. Ibaraki, T., Katoh, N.: Resource allocation problems: algorithmic approaches. MIT Press, Cambridge (1988)
8. Joseph, L.N., Maloney, R.F., Possingham, H.P.: Optimal allocation of resources among threatened species: a project prioritization protocol. Conserv. Biol. **23**(2), 328–338 (2009)
9. Julian, D., Chiang, M., O'Neill, D., Boyd, S.: QoS and fairness constrained convex optimization of resource allocation for wireless cellular and ad hoc networks. In: Proceedings Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 477–486 (2002). https://doi.org/10.1109/INFCOM.2002.1019292
10. Kantorovich, L.V.: Mathematical methods of organizing and planning production. Manag. Sci. **6**(4), 366–422 (1960)
11. Kaplan, S.: Application of programs with maximin objective functions to problems of optimal resource allocation. Oper. Res. **22**(4), 802–807 (1974)
12. Klein, R.S., Luss, H., Rothblum, U.G.: Multiperiod allocation of substitutable resources. Eur. J. Oper. Res. **85**(3), 488–503 (1995)

13. Palomar, D.P., Fonollosa, J.R.: Practical algorithms for a family of waterfilling solutions. IEEE Trans. Signal Process. **53**(2), 686–695 (2005)
14. Shiner, R.A.: Aristotle's theory of equity. Loy. LAL Rev. **27**, 1245 (1993)
15. Singh, B.: Optimal spatiotemporal resource allocation in public health and renewable energy. Ph.D. thesis, The University of Texas at Austin (2016)
16. Zhang, S.: On the strictly complementary slackness relation in linear programming. In: Du, D.Z., Sun, J. (eds.) Advances in Optimization and Approximation, pp. 347–361. Springer, Boston (1994)