Mathematical Programming 45 (1989) 193-210
North-Holland

ALGORITHMS FOR PROPORTIONAL MATRICES IN REALS AND INTEGERS

M.L. BALINSKI

C.N.R.S., Laboratoire d'Econométrie de l'Ecole Polytechnique, Paris, France, and Institute for Decision Sciences, S.U.N.Y., Stony Brook, USA

G. DEMANGE

Laboratoire d'Econométrie de l'Ecole Polytechnique, Paris, France, and C.E.R.M.S.E.M., Université Paris I, France

Received 15 March 1987
Revised manuscript received 29 January 1988

Let R be the set of nonnegative matrices whose row and column sums fall between specific limits and whose entries sum to some fixed $h > 0$. Closely related axiomatic approaches have been developed to ascribe meanings to the statements: the real matrix $f \in R$ and the integer matrix $a \in R$ are "proportional to" a given matrix $p \geq 0$.

These approaches are described, conditions under which proportional solutions exist are characterized, and algorithms are given for finding proportional solutions in each case.

Introduction

Regional councils in the Netherlands are composed of seats that simultaneously represent both townships and political parties (see [1]). The stated intent is that each township should receive a number of seats proportional to its population and each political party a number of seats proportional to its total vote. This gives rise to the following "matrix" problem. Suppose $h > 0$ is the size of the council, that $p_{ij} \geq 0$ is the number of votes for party j in township i, that r_i is the number of seats apportioned to township $i, \sum_i r_i = h$, and that c_j is the number of seats assigned to party $j, \sum_j c_j = h$. How many seats a_{ij} *should* be apportioned to the candidates of party j in township i? The reflex response is, numbers proportional to the p_{ij}, but what precisely does this mean? An axiomatic answer to this question was advanced for a special case of this problem, the "vector" apportionment problem [4]: $h > 0$ is the size of the council, $p_i \geq 0$ the population of township i (or the number of votes of party i) and the question is how many seats a_i should be apportioned to township i (or party i). Note that solutions to the vector apportionment problem with appropriate choices of the data p_i determine respectively the r_i and the c_j for the matrix problem.

Consider another problem (reviewed by Cox and Ernst [6]). In Canada, as well as most countries, census data are gathered with a guarantee of anonymity. Given,

however, arrays of many tables containing cross classifications concerning a community that is small in numbers, it may be possible to deduce the identities of certain subjects of the population, and so fail the guarantee of anonymity. To avoid this possibility the counts in Canada are recorded as multiples of 5. Given the row data p_{ij} of any table, row sums r_i, $\sum_i r_i = h$, and column sums c_j, $\sum_j c_j = h$, that are multiples of 5 (in both cases solutions to vector apportionment problems) what a_{ij}, multiples of 5, *should* replace the corresponding p_{ij}? Again the answer seems to be, proportional numbers; again, what precisely does this mean?

Indeed, what proportionality should mean when the integer requirement is relaxed and solutions in real numbers are sought is not clear either. In related work [3] we have developed an axiomatic approach to both of these problems. This paper contains a brief account of the axioms and results for defining proportionality when solutions are required in reals and when solutions are required in integers. Its primary objective, however, is to give algorithms for finding solutions in both cases and to characterize when solutions exist. Section 1 concerns proportionality in reals, and Section 2 proportionality in integers.

1. Proportional matrix allocations

1.1. Definitions

In the sequel $\boldsymbol{x} > \mathbf{0}$, for $\boldsymbol{x}$ a vector or matrix, means every component of $\boldsymbol{x}$ is positive, whereas $\boldsymbol{x} \geqslant \mathbf{0}$ means every component of $\boldsymbol{x}$ is nonnegative, and $M = \{1, \ldots, i, \ldots, m\}$ and $N = \{1, \ldots, j, \ldots, n\}$.

A *problem* is a pair $(\boldsymbol{p}, \boldsymbol{\sigma})$, where $\boldsymbol{p} = (p_{ij}) \geqslant \mathbf{0}$ is an m by n matrix containing no row or column of zeros, and $\boldsymbol{\sigma} = (\boldsymbol{r}^-, \boldsymbol{r}^+, \boldsymbol{c}^-, \boldsymbol{c}^+, h)$ is a vector with $\boldsymbol{r}^- = (r_i^-) \geqslant \mathbf{0}$ and $\boldsymbol{r}^+ = (r_i^+) > \mathbf{0}$ two m-vectors, $\boldsymbol{c}^- = (c_j^-) \geqslant \mathbf{0}$ and $\boldsymbol{c}^+ = (c_j^+) > \mathbf{0}$ two n-vectors, and h a positive scalar.

The set of allocations $R(\boldsymbol{\sigma})$ is

$$R(\boldsymbol{\sigma}) = \{\boldsymbol{f} = (f_{ij}) \geqslant \mathbf{0}\colon r_i^- \leqslant f_{iN} \leqslant r_i^+,\ i \in M;\ c_j^- \leqslant f_{Mj} \leqslant c_j^+,\ j \in N;\ f_{MN} = h\}$$

where $t_{IJ} = \sum_{I \times J} t_{ij}$.

From now on, we consider only nonempty regions of allocations.

The first question is: what does it mean to say that an allocation $\boldsymbol{f}$ in $R(\boldsymbol{\sigma})$ is proportional to $\boldsymbol{p}$?

Two special types of problems $(\boldsymbol{p}, \boldsymbol{\sigma})$ play key roles. If $\boldsymbol{p} > \mathbf{0}$ the problem is *positive*. If $\boldsymbol{r}^- = \boldsymbol{r}^+$ and $\boldsymbol{c}^- = \boldsymbol{c}^+$ the problem is *equality constrained*.

In discussing the existence of solutions several subsets of $R(\boldsymbol{\sigma})$ are singled out:

$$R^0(\boldsymbol{p}, \boldsymbol{\sigma}) = \{\boldsymbol{f} \in R(\boldsymbol{\sigma})\colon f_{ij} = 0 \text{ if } p_{ij} = 0\},$$

$$R^+(\boldsymbol{p}, \boldsymbol{\sigma}) = \{\boldsymbol{f} \in R(\boldsymbol{\sigma})\colon f_{ij} = 0 \text{ if and only if } p_{ij} = 0\}.$$

1.2. Axioms

A *method of allocation* F is a correspondence that assigns at least one allocation to every problem: $F(\boldsymbol{p}, \boldsymbol{\sigma})$ is a nonempty subset of $R(\boldsymbol{\sigma})$. The possibility of multiple allocations is not excluded. For the method F to be "proportional" it should satisfy a number of basic properties suggested by the usual (vector) idea of proportionality.

The following set of principles provides a seemingly reasonable approach.

Axiom 1 (exactness). If $\delta \boldsymbol{p} \in R(\boldsymbol{\sigma})$ for some scalar $\delta > 0$ then $F(\boldsymbol{p}, \boldsymbol{\sigma}) = \{\delta \boldsymbol{p}\}$.

This is simply asking for the usual idea to work if some scalar multiple of $\boldsymbol{p}$ happens to belong to the feasible set $R(\boldsymbol{\sigma})$.

Axiom 2 (relevance). If $F(\boldsymbol{p}, \boldsymbol{\sigma}) \cap R(\hat{\boldsymbol{\sigma}}) \neq \emptyset$ and $R(\hat{\boldsymbol{\sigma}}) \subset R(\boldsymbol{\sigma})$ then $F(\boldsymbol{p}, \hat{\boldsymbol{\sigma}}) \subset F(\boldsymbol{p}, \boldsymbol{\sigma}) \cap R(\hat{\boldsymbol{\sigma}})$.

This is a kind of "independence of irrelevant alternatives" property. It says that if some of the allocations of $F(\boldsymbol{p}, \boldsymbol{\sigma})$ (which are meant to be "proportional") belong to the more constrained region $R(\hat{\boldsymbol{\sigma}})$ then surely one can obtain no better set of allocations $F(\boldsymbol{p}, \hat{\boldsymbol{\sigma}})$ than those: the possibilities in $R(\boldsymbol{\sigma}) - R(\hat{\boldsymbol{\sigma}})$ are irrelevant.

In the sequel $\bar{X}$ denotes the complement of X. Let $\boldsymbol{t}_{I \times J}$ be the submatrix of $\boldsymbol{t}$ defined on rows $I \subset M$ and columns $J \subset N$. Given $\boldsymbol{f} \in F(\boldsymbol{p}, \boldsymbol{\sigma})$ the *subproblem* $(\boldsymbol{p}_{I \times J}, \boldsymbol{\sigma}_{I \times J})$ has the set $R(\boldsymbol{\sigma}_{I \times J})$ defined over rows I and columns J with lower bounds $r_i^- - f_{i\bar{J}}$ and $c_j^- - f_{\bar{I}j}$, upper bounds $r_i^+ - f_{i\bar{J}}$ and $c_j^+ - f_{\bar{I}j}$, and sum f_{IJ}.

Axiom 3 (uniformity). If $\boldsymbol{f} \in F(\boldsymbol{p}, \boldsymbol{\sigma})$ then $\boldsymbol{f}_{I \times J} \in F(\boldsymbol{p}_{I \times J}, \boldsymbol{\sigma}_{I \times J})$; and, conversely, if $\boldsymbol{g}_{I \times J} \in F(\boldsymbol{p}_{I \times J}, \boldsymbol{\sigma}_{I \times J})$ then for $\boldsymbol{g}$ defined to be equal to $\boldsymbol{g}_{I \times J}$ on $I \times J$ and $\boldsymbol{f}$ elsewhere, $\boldsymbol{g} \in F(\boldsymbol{p}, \boldsymbol{\sigma})$.

Uniformity is the familiar property that any part of a proportional solution must itself be proportional, and that if a part of the problem admits another solution then it may be substituted to obtain another solution to the whole problem. It was first introduced in the context of the vector apportionment problem.

Axiom 4 (monotonicity). If $\boldsymbol{f} \in F(\boldsymbol{p}, \boldsymbol{\sigma}), \boldsymbol{f}' \in F(\boldsymbol{p}', \boldsymbol{\sigma})$, and $\boldsymbol{p}'$ is equal to $\boldsymbol{p}$ except that $p_{kl} < p'_{kl}$ then $f_{kl} \leqslant f'_{kl}$.

Monotonicity asks that a change in population data be accompanied only by a reasonable change in possible allocations.

Axiom 5 (homogeneity). Suppose $(\boldsymbol{p}, \boldsymbol{\sigma})$ is equality constrained. If two rows of $\boldsymbol{p}$ are proportional and are constrained to the same sum, then the corresponding rows of any $\boldsymbol{f} \in F(\boldsymbol{p}, \boldsymbol{\sigma})$ are identical (and the same holds for columns).

Homogeneity insists that two proportional population rows (or columns) whose allocation sums must be identical should have identical row (or column) allocations, just as is the case in ordinary vector proportionality.

In [3], we characterize a unique method satisfying the above axioms over the set of positive problems. It will be convenient to use the following notation: if δ is a scalar, $\boldsymbol{\lambda} = (\lambda_i)$, $\boldsymbol{\mu} = (\mu_j)$ and $\boldsymbol{p} = (p_{ij})$ then $\delta\boldsymbol{\lambda}\boldsymbol{p}\boldsymbol{\mu}$ will represent the matrice $(\delta\lambda_i p_{ij}\mu_j)$.

A matrix $\boldsymbol{f}$ is said to be a *fair share matrix* for a problem $(\boldsymbol{p}, \boldsymbol{\sigma})$ if

$$\boldsymbol{f} = \delta\boldsymbol{\lambda}\boldsymbol{p}\boldsymbol{\mu}, \quad \boldsymbol{f} \in R(\boldsymbol{\sigma}), \tag{1}$$

for some $\delta > 0$, $\boldsymbol{\lambda} > \boldsymbol{0}$, $\boldsymbol{\mu} > \boldsymbol{0}$ satisfying:

$$\lambda_i > 1 \text{ implies } f_{iN} = r_i^- \quad \text{and} \quad \lambda_i < 1 \text{ implies } f_{iN} = r_i^+,$$

$$\mu_j > 1 \text{ implies } f_{Mj} = c_j^- \quad \text{and} \quad \mu_j < 1 \text{ implies } f_{Mj} = c_j^+.$$

Intuitively one can see that a fair share matrix departs from the usual proportional matrix only via multipliers of rows and columns, a multiplier being greater than one (or less than one) only if it must be to meet the lower bound (or to meet the upper bound) requirement.

We establish in [3] that a fair share matrix exists and is unique for any positive problem. This allows the definition of the *fair share method* F^* over the set of positive problems and then a proof of the following:

Characterization theorem. *The fair share method is the unique method of allocation satisfying Axioms* 1 *through* 5 *over the class of positive problems.* □

1.3. Existence

It is part of the folklore that in the equality constrained problem a fair share matrix of a positive problem is the solution of a convex program (see, for example, [2]). We extend these results in two ways: the matrix $\boldsymbol{p}$ may have zeros and inequality constraints replace equations.

Given $\boldsymbol{p} \geq \boldsymbol{0}$ denote by S the set of indices (i, j) for which $p_{ij} > 0$ and by $\bar{S}$ its complement. Consider the program (where ln denotes the natural logarithm)

$$\text{minimize} \quad \sum_S x_{ij}[\ln(x_{ij}/p_{ij}) - 1] \tag{2}$$

$$\text{subject to} \quad r_i^- \leq x_{iN} \leq r_i^+, \quad c_j^- \leq x_{Mj} \leq c_j^+, \quad x_{MN} = h, \text{ and}$$

$$x_{ij} \geq 0 \ (i, j) \in S, \quad x_{ij} = 0 \ (i, j) \in \bar{S}.$$

Here the feasible set is $R^0(\boldsymbol{p}, \boldsymbol{\sigma})$. Recall that $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \subset R^0(\boldsymbol{p}, \boldsymbol{\sigma}) \subset R(\boldsymbol{\sigma})$.

Theorem 1. *A fair share matrix exists if and only if* $R^+(\boldsymbol{p}, \boldsymbol{\sigma})$ *is nonempty, in which case it is the unique solution to* (2).

Proof. Suppose that a fair share matrix exists. No multiplier can be null: for if $\lambda_i = 0$ then $f_{iN} = r_i^+ = 0$, contradicting $r_i^+ > 0$, and the same for any μ_j and δ. Thus $f_{ij} = \delta\lambda_i p_{ij}\mu_j = 0$ if and only if $p_{ij} = 0$, and $\boldsymbol{f} \in R^+(\boldsymbol{p}, \boldsymbol{\sigma})$.

Now suppose $R^+(\boldsymbol{p}, \boldsymbol{\sigma})$ is nonempty. Since the objective function is convex, bounded on the feasible set and the constraints are affine, Kuhn-Tucker multipliers exist whenever there is a feasible point in the interior of the domain of the objective function (see [8, p. 279]). This interior is the set of $\boldsymbol{x}$ with $x_{ij} > 0$ for $(i, j) \in S$, and so the condition is satisfied precisely when $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$. Thus program (2) has a minimum $\boldsymbol{f}$ and there are nonnegative multipliers (α_i^-, α_i^+) for row i, (β_j^-, β_j^+) for column j, and a multiplier ν satisfying: $\boldsymbol{f}$ minimizes the Lagrangian L over $\boldsymbol{x} \geqslant \boldsymbol{0}$ and $x_{ij} = 0$ for $(i, j) \in \bar{S}$,

$$
\begin{aligned}
L(\boldsymbol{x}, \boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+, \boldsymbol{\beta}^-, \boldsymbol{\beta}^+, \nu) = {} & \sum_S x_{ij}[\ln(x_{ij}/p_{ij}) - 1] + \sum_M \alpha_i^-[r_i^- - x_{iN}] \\
& + \sum_M \alpha_i^+[x_{iN} - r_i^+] + \sum_N \beta_j^-[c_j^- - x_{Mj}] \\
& + \sum_N \beta_j^+[x_{Mj} - c_j^+] + \nu(h - x_{MN}). \qquad (3)
\end{aligned}
$$

Also $\boldsymbol{f} \in R(\boldsymbol{\sigma})$ and the orthogonality conditions are satisfied:

$$
\alpha_i^- > 0 \text{ implies } f_{iN} = r_i^-, \qquad \alpha_i^+ > 0 \text{ implies } f_{iN} = r_i^+, \qquad (4)
$$

and similarly for the columns. $f_{ij} > 0$ for $(i, j) \in S$ since otherwise $\partial L/\partial x_{ij} < 0$ at $x_{ij} = 0$. By the first order conditions,

$$
f_{ij} = p_{ij} \exp\{\alpha_i^- - \alpha_i^+ + \beta_j^- - \beta_j^+ + \nu\} \quad \text{for } (i, j) \in S,
$$

$$
f_{ij} = 0 \quad \text{for } (i, j) \in \bar{S}.
$$

Letting $\lambda_i = \exp\{\alpha_i^- - \alpha_i^+\}$, $\mu_j = \exp\{\beta_j^- - \beta_j^+\}$ and $\delta = \exp\{\nu\}$ one obtains

$$
f_{ij} = \delta\lambda_i p_{ij}\mu_j \quad \text{for all } (i, j).
$$

Moreover, if $\lambda_i > 1$ then necessarily $\alpha_i^- > 0$ so by (4), $f_{iN} = r_i^-$, and analogously for the other constraints. This shows that $\boldsymbol{f}$ is a fair share matrix.

It remains to prove uniqueness. Since the objective is strictly convex in the variables x_{ij}, $(i, j) \in S$, program (2) has a unique solution. Suppose now that $\boldsymbol{f}$ is a fair share matrix. Then define multipliers α_i^- and α_i^+ as follows: if $\lambda_i \geqslant 1$ set $\alpha_i^- = \ln \lambda_i$ and $\alpha_i^+ = 0$, otherwise set $\alpha_i^- = 0$ and $\alpha_i^+ = -\ln \lambda_i$. Define the other multipliers similarly. Then $\boldsymbol{f}$ minimizes the Lagrangian and the orthogonality conditions are satisfied, proving that $\boldsymbol{f}$ is the unique solution of (2). □

The natural questions that remain to be answered are: when is $R^+(\boldsymbol{p}, \boldsymbol{\sigma})$ nonempty and what can be done if in fact it is empty?

When the matrix $\boldsymbol{p}$ is strictly positive the answer is easy: $R^+(\boldsymbol{p}, \boldsymbol{\sigma})$ is nonempty if and only if

$$r_M^- \leq h \leq r_M^+, \quad c_N^- \leq h \leq c_N^+ \quad \text{and} \quad r_I^- < h, \quad c_J^- < h \quad \text{for all proper } I \subset M, J \subset N.$$

However, when the matrix $\boldsymbol{p}$ contains zeros, difficulties arise. Recall that $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \subset R^0(\boldsymbol{p}, \boldsymbol{\sigma}) \subset R(\boldsymbol{\sigma})$. Necessary and sufficient conditions for $R^0(\boldsymbol{p}, \boldsymbol{\sigma})$ to be nonempty are the following adaptation of the familiar "supply-demand" conditions of network flow theory. The necessary and sufficient conditions for $R^0(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ are that

$$c_J^+ \geq r_{\bar{I}}^-, \qquad r_{\bar{I}}^+ \geq c_{\bar{J}}^- \quad \text{and} \quad c_{\bar{J}}^- + r_{\bar{I}}^- \leq h \leq c_J^+ + r_{\bar{I}}^+, \tag{5}$$

for any $I \subset M, J \subset N$ with $p_{I\bar{J}} = 0$ ($r_{\emptyset}^- = 0$, etc.).

These conditions do not exclude the possibility $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) = \emptyset$ as the following equality constrained problem shows:

$$\boldsymbol{p} = \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix}, \qquad \boldsymbol{r} = (1, 1), \qquad \boldsymbol{c} = (1, 1).$$

$R^+(\boldsymbol{p}, \boldsymbol{\sigma})$ is empty because one of the inequalities (5) is satisfied as an equation: the first inequality with $I = \{1\}$ and $J = \{1\}$, forcing $x_{11} = 1$ and so $x_{21} = 0$ whereas $p_{21} > 0$.

A problem $(\boldsymbol{p}, \boldsymbol{\sigma})$ is said to be *irreducible* if the inequalities (5) hold and are satisfied strictly whenever the subsets are proper. It has been shown in [3] that $(\boldsymbol{p}, \boldsymbol{\sigma})$ irreducible implies $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$. The proof given there is embedded in the proof of the existence of fair shares. Now that we have Theorem 1 all that needs to be done is to show the existence of a matrix $\boldsymbol{x} \in R^0(\boldsymbol{p}, \boldsymbol{\sigma})$ satisfying $x_{ij} > 0$ if and only if $p_{ij} > 0$. This is easily done constructively by beginning with some $\boldsymbol{y} \in R(\boldsymbol{\sigma})$, and if $y_{ij} = 0$ for $p_{ij} > 0$, seeking a flow augmenting path from j to i in an associated network; and repeating. If no such path exists then some inequality of (5) is satisfied as an equation, contradicting the assumption of irreducibility.

If the conditions (5) hold for all $I \subset M$ and $(\boldsymbol{p}, \boldsymbol{\sigma})$ is reducible due to subsets $I \subset M$ and $J \subset N$, with $p_{I\bar{J}} = 0$ then $(\boldsymbol{p}, \boldsymbol{\sigma})$ is said to be *decomposable* into independent subproblems if also $p_{\bar{I}J} = 0$. If, for example, $c_J^+ = r_I^-$ then the subproblem on $(\bar{I}, \bar{J})$ must sum to $h - c_J^+$; and similarly if $r_{\bar{I}}^+ = c_{\bar{J}}^-$. If $c_{\bar{J}}^- + r_I^- = h$ then the subproblem on (I, J) must have an allocation summing to r_I^- while that on $(\bar{I}, \bar{J})$ must sum to $c_{\bar{J}}^-$; and similarly if $h = c_J^+ + r_{\bar{I}}^+$.

Theorem 2. *$R^+(\boldsymbol{p}, \boldsymbol{\sigma})$ is nonempty if and only if $(\boldsymbol{p}, \boldsymbol{\sigma})$ can be decomposed into a set of independent irreducible subproblems.*

Proof. If the condition is satisfied then $R^+(\boldsymbol{p}, \boldsymbol{\sigma})$ is nonempty. If $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ then $R^0(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ and so conditions (5) must hold. If they are all satisfied strictly then $(\boldsymbol{p}, \boldsymbol{\sigma})$ is irreducible. Otherwise, equality holds for some (I, J). But, then, if $\boldsymbol{x} \in R^+(\boldsymbol{p}, \boldsymbol{\sigma})$, $x_{\bar{I}J} = x_{I\bar{J}} = 0$, implying $p_{\bar{I}J} = p_{I\bar{J}} = 0$, so $(\boldsymbol{p}, \boldsymbol{\sigma})$ is decomposable into independent subproblems. □

An example of these results is the 5 by 7 problem $(\boldsymbol{p}, \boldsymbol{\sigma})$ defined by $h = 32$ and

$$\boldsymbol{p} = \begin{bmatrix} \times & & \times & & & & \\ \times & \times & \times & & & & \\ \varepsilon & & & \times & \times & & \\ & & & & & \times & \times \\ & & & \delta & & \times & \times \end{bmatrix},$$

$$\boldsymbol{r}^- = (5, 7, 1, 1, 1), \qquad \boldsymbol{r}^+ = (7, 9, 4, 9, 9),$$

$$\boldsymbol{c}^- = (0, 0, 0, 1, 1, 1, 1), \qquad \boldsymbol{c}^+ = (3, 4, 5, 6, 7, 8, 8),$$

where in $\boldsymbol{p}$ a $\times$ denotes that the corresponding p_{ij} is positive, $p_{31} = \varepsilon \geq 0$, $p_{54} = \delta \geq 0$ and blanks denote the corresponding $p_{ij} = 0$. Consider $I = \{1, 2\}$ and $J = \{1, 2, 3\}$, with $p_{I\bar{J}} = 0$. Then $r_I^- = c_J^+ = 12$, so $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) = \emptyset$ unless $\varepsilon = 0$. If $\varepsilon = 0$ then the problem decomposes into two subproblems: the 2 by 3 problem at the northwest corner with $h = 12$, and the 3 by 4 problem at the southeast corner problems with $h = 20$. Consider now only the southeast corner problem: If $I = \{4, 5\}$ and $J = \{6, 7\}$, then $\bar{I} = \{3\}$, $20 = h = r_{\bar{I}}^+ + c_J^+$, so $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) = \emptyset$ for this subproblem unless $\delta = 0$. Moreover, if $\delta = 0$ then there is a further decomposition into two subproblems. Summarizing, the example decomposes into three independent irreducible subproblems and $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ if and only if $\varepsilon = 0$ and $\delta = 0$.

Over the set of positive problems the fair share matrix $\boldsymbol{f}$ depends continuously on $\boldsymbol{p}$ since both the objective function and the constraints are continuous in $\boldsymbol{p}$. When, however, $\boldsymbol{p} \geq 0$ the result is no longer necessarily true. It has been shown in [3] that the fair share matrix $\boldsymbol{f}$ is continuous over those $\boldsymbol{p} \geq \boldsymbol{0}$ for which $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$.

If $\boldsymbol{p} \geq \boldsymbol{0}$, $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) = \emptyset$ and $R^0(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$, then the unique matrix $\boldsymbol{f}$ that solves (2) may also be obtained via continuity (see [3]). In this case some inequality in (5) is satisfied as an equation so that any feasible solution $\boldsymbol{f}$ of (2) must satisfy $f_{ij} = 0$ for $(i, j) \in (I, \bar{J}) \cup (\bar{I}, J)$. If one is willing to accept only $f_{ij} = 0$ if $p_{ij} = 0$ (admitting the possibility $f_{ij} = 0$ when $p_{ij} > 0$), then the problem effectively decomposes into the two independent subproblems defined on (I, J) and $(\bar{I}, \bar{J})$ and $f_{I\bar{J}} = 0 = f_{\bar{I}J}$. The constraint sets corresponding to $R^0(\boldsymbol{p}, \boldsymbol{\sigma})$ are nonempty on each subproblem, so repeating the analysis one sees that ultimately the nonzero terms of $\boldsymbol{f}$ are fair share matrices corresponding to some collection of independent subproblems.

When $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) = \emptyset = R^0(\boldsymbol{p}, \boldsymbol{\sigma})$ but $R(\boldsymbol{\sigma}) \neq \emptyset$ it is tempting to take the same approach and solve the program: minimize $\sum_S x_{ij}[\ln(x_{ij}/p_{ij}) - 1]$ over $R(\boldsymbol{\sigma})$. But the result is not satisfactory for such solutions are not continuous and indeed there is no way to extend the definition in a continuous and unique manner, as the following example shows. The matrix that solves the above program for the equality constrained problem

$$\boldsymbol{p} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \qquad \boldsymbol{r} = (2, 2, 2), \qquad \boldsymbol{c} = (1, 1, 4),$$

is

$$\begin{bmatrix} \alpha & \alpha & 2-2\alpha \\ \alpha & \alpha & 2-2\alpha \\ 1-2\alpha & 1-2\alpha & 4\alpha \end{bmatrix} \quad \text{for } \alpha = \frac{-1+\sqrt{3}}{4}.$$

However, the fair share of the matrices $\boldsymbol{p}(\varepsilon)$ when ε goes to zero,

$$\boldsymbol{p}(\varepsilon) = \begin{bmatrix} 1 & 1 & \varepsilon \\ 1 & 1 & \varepsilon \\ 1 & 1 & 1 \end{bmatrix} \quad \text{tend to} \quad \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 1 \\ \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 0 & 2 \end{bmatrix},$$

whereas the fair shares of the matrices $\boldsymbol{p}'(\varepsilon)$ when ε goes to zero,

$$\boldsymbol{p}'(\varepsilon) = \begin{bmatrix} 1 & 1 & \varepsilon \\ 1 & 1 & \varepsilon^2 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{tend to} \quad \begin{bmatrix} 0 & 0 & 2 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

This shows that to have satisfactory solutions it is necessary that $R^0(\boldsymbol{p}, \boldsymbol{\sigma})$ be nonempty.

1.4. Algorithm

The algorithm described below generalizes the well-known iterative process used for equality problems which alternatively scales the rows and columns of $\boldsymbol{p}$ to sum to their respective values (see, e.g., [2, 9]). The proof of convergence uses a theorem of Zangwill and can be interpreted as a cyclic coordinate ascent method. The same approach was used by Bigelow and Shapiro [5] for the positive equality constrained problem.

To begin we consider the *vector allocation problem* $(\boldsymbol{q}, \boldsymbol{c}^-, \boldsymbol{c}^+, h)$, for vectors $\boldsymbol{q} > \boldsymbol{0}, \boldsymbol{0} \leq \boldsymbol{c}^- \leq \boldsymbol{c}^+$ and $h > 0$: the set of feasible allocations is $\{\boldsymbol{x} = (x_j)\colon c_j^- \leq x_j \leq c_j^+, x_N = h\}$. Its *fair share vector* $\boldsymbol{f} = (f_j)$ is easily found: let $\delta > 0$ be such that

$$\sigma(\delta) = \sum_N \operatorname{mid}(c_j^-, \delta q_j, c_j^+) = h$$

and

$$c_j^- \leq \delta q_j \leq c_j^+ \quad \text{for at least one } j \in N,$$

where $\operatorname{mid}(x, y, z) = y$ if $x \leq y \leq z$. Such a value δ exists if and only if $c_N^- \leq h \leq c_N^+$ and $c_J^- < h$ for all proper $J \subset N$. To see this note that $\sigma(\delta)$ is continuous in δ, and that $\sigma(0) = c_N^-$ whereas $\sigma(\delta) = c_N^+$ for δ large. Therefore there exist values δ satisfying $\sigma(\delta) = h$. There is a $\delta > 0$, for otherwise, $c_J^- = h$ for some proper $J \subset N$. Appropriate multipliers and fair shares are then defined by

$$\mu_j = \operatorname{mid}(c_j^-, \delta q_j, c_j^+)/\delta q_j \quad \text{and} \quad f_j = \delta \mu_j q_j.$$

So

$$\mu_j > 1 \text{ implies } f_j = c_j^- \quad \text{and} \quad \mu_j < 1 \text{ implies } f_j = c_j^+.$$

The algorithm given below is based on the solutions of such vector allocation problems. The proof of its convergence requires, however, that $\min \mu_j \leq 1 \leq \max \mu_j$. This is easily met. Suppose that $\delta > 0$ and $\sigma(\delta) = h$ with (μ_j) the corresponding multipliers. If δq_j is between c_j^- and c_j^+ for some j then $\mu_j = 1$ and the required condition obtains. Otherwise, δ may be increased or decreased without changing the values of $\sigma(\delta)$ until either $\delta q_j = c_j^-$ or $\delta q_j = c_j^+$ for some j and so again $\mu_j = 1$. Thus, in the following, it is assumed that the multipliers of solutions to vector allocation problems are chosen to satisfy this property.

The *allocation algorithm* is simply a sequence of solutions to vector allocation problems:

Step 0. $\lambda_i^0 = 1 = \mu_j^0$ all i and j.

Step (2k+1). Compute the fair share vector of $((\sum_j p_{ij}\mu_j^{2k}), \boldsymbol{r}^-, \boldsymbol{r}^+, h)$. Let δ^{2k+1} and $\boldsymbol{\lambda}^{2k+1}$ be multipliers and set $\mu^{2k+1} = \boldsymbol{\mu}^{2k}$.

Step (2k+2). Compute the fair share vector of $((\sum_i \lambda_i^{2k+1} p_{ij}), \boldsymbol{c}^-, \boldsymbol{c}^+, h)$. Let δ^{2k+2} and $\boldsymbol{\mu}^{2k+2}$ be multipliers and set $\boldsymbol{\lambda}^{2k+2} = \boldsymbol{\lambda}^{2k+1}$.

Let $\boldsymbol{f}^k = \delta \boldsymbol{\lambda}^k \boldsymbol{p} \boldsymbol{\mu}^k$ be the matrix associated with Step k. The sum of its entries for $k \geq 1$ is always h. If k is odd, $\boldsymbol{f}^k$ satisfies the row constraints, whereas if k is even it satisfies the column constraints.

Theorem 3. *Let $(\boldsymbol{p}, \boldsymbol{\sigma})$ be a positive problem with $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$. Then the multipliers of the algorithm converge to some $(\delta, \boldsymbol{\lambda}, \boldsymbol{\mu})$ and $\boldsymbol{f} = \delta \boldsymbol{\lambda} \boldsymbol{p} \boldsymbol{\mu}$ is the fair share matrix of the problem.*

Proof. The idea of the proof is very natural (see, for example, [7, pp. 121–125]). Think of the algorithm as a point to set mapping φ that takes a trial solution $\boldsymbol{x}^k \in X$ into some (not necessarily unique) successor $\boldsymbol{x}^{k+1} \in \varphi(\boldsymbol{x}^k) \subset X$. φ is assumed to be "closed": If for $\boldsymbol{x}^k \in X$, $\boldsymbol{x}^k \to \boldsymbol{x}$ and $\boldsymbol{y}^k \in \varphi(\boldsymbol{x}^k)$, $\boldsymbol{y}^k \to \boldsymbol{y}$, then $\boldsymbol{y} \in \varphi(\boldsymbol{x})$. In addition, suppose H is a continuous function on X and $X^* \subset X$ is a "solution set" satisfying for any $\boldsymbol{y} \in \varphi(\boldsymbol{x})$: $H(\boldsymbol{x}) < H(\boldsymbol{y})$ if $\boldsymbol{x} \notin X^*$ and $H(\boldsymbol{x}) \leq H(\boldsymbol{y})$ if $\boldsymbol{x} \in X^*$. Then, if X is compact, Zangwill's theorem asserts that the limit of any convergent subsequence is a solution.

Take for H the (dual objective) function

$$H(\boldsymbol{\alpha}^-, \boldsymbol{\alpha}^+, \boldsymbol{\beta}^-, \boldsymbol{\beta}^+, \nu) = \sum_M (\alpha_i^- r_i^- - \alpha_i^+ r_i^+) + \sum_N (\beta_j^- c_j^- - \beta_j^+ c_j^+)$$
$$+ \nu h - \sum_{M \times N} p_{ij} \exp\{\alpha_i^- - \alpha_i^+ + \beta_i^- - \beta_i^+ + \nu\}$$

for $\boldsymbol{\alpha}^- \geq \boldsymbol{0}$, $\boldsymbol{\alpha}^+ \geq \boldsymbol{0}$, $\boldsymbol{\beta}^- \geq \boldsymbol{0}$, $\boldsymbol{\beta}^+ \geq \boldsymbol{0}$. If $\boldsymbol{\beta}^-$ and $\boldsymbol{\beta}^+$ are fixed, compute the fair share of the problem $((\sum_j p_{ij} \exp\{\beta_j^- - \beta_j^+\}), \boldsymbol{r}^-, \boldsymbol{r}^+)$. If δ and $\boldsymbol{\lambda}$ are the corresponding multipliers define ν, $\boldsymbol{\alpha}^-$ and $\boldsymbol{\alpha}^+$ as follows:

$$\alpha_i^- = \ln \lambda_i \text{ and } \alpha_i^+ = 0, \quad \text{if } \lambda_i > 1,$$

$$\alpha_i^- = 0 \text{ and } \alpha_i^+ = -\ln \lambda_i, \quad \text{if } \lambda_i \leq 1,$$

$$\nu = \ln \delta.$$

It is easy to check that for these values (where α_i is α_i^- or α_i^+),

$$\partial H/\partial \alpha_i = 0 \quad \text{or} \quad \partial H/\partial \alpha_i \leq 0 \text{ and } \alpha_i = 0, \qquad \partial H/\partial \nu = 0.$$

Thus, since H is concave, H is maximized with respect to $\boldsymbol{\alpha}^- \geq \mathbf{0}$, $\boldsymbol{\alpha}^+ \geq \mathbf{0}$ and ν, for $\boldsymbol{\beta}^-$ and $\boldsymbol{\beta}^+$ fixed. The same holds if $\boldsymbol{\alpha}^-$ and $\boldsymbol{\alpha}^+$ are fixed and H is maximized over $\boldsymbol{\beta}^- \geq \mathbf{0}$, $\boldsymbol{\beta}^+ \geq \mathbf{0}$ and ν, so this is a cyclic coordinate ascent method.

The algorithm is closed. Consider an odd step, and suppose that a sequence $\boldsymbol{\lambda}^s$ converges to $\boldsymbol{\lambda}$ and that the associated multipliers $(\delta^s, \boldsymbol{\mu}^s)$ converge to $(\delta, \boldsymbol{\mu})$. We must show that $(\delta, \boldsymbol{\mu})$ could be chosen by the algorithm when the multipliers are $\boldsymbol{\lambda}$. The argument is simple. Since all the formula for determining $(\delta^s, \boldsymbol{\mu}^s)$ are continuous in $\boldsymbol{\lambda}^s$ and there are only a finite number of inequalities $c_j^- \leq \delta^s \sum_i \lambda_i^s p_{ij} \leq c_j^+$, one of these must be satisfied infinitely often so that the multipliers $\boldsymbol{\lambda}$ must admit $(\delta, \boldsymbol{\mu})$ as a solution.

Finally, the sequence of points $(\delta^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$ is bounded. $\boldsymbol{p} > \mathbf{0}$ and $\sum_{M \times N} \delta^k \lambda_i^k p_{ij} \mu_j^k = h$ implies

$$\delta^k \max_M \lambda_i^k \max_N \mu_j^k \leq B \quad \text{for some } B \text{ and any } k. \tag{6}$$

Since $\max_M \lambda_i^k \geq 1$, $\max_N \mu_j^k \geq 1$ it follows that $\delta^k \leq B$, so δ^k is bounded.

Take k to be odd. Suppose $I = \{i: r_i^- > 0\}$ is empty; then $\lambda_i^k > 1$ implies $\sum_j \delta^k p_{ij} \mu_j^k = r_i^- = 0$, which is impossible. Thus, $\lambda_i^k \leq 1$ obtains for all i and $\boldsymbol{\lambda}^k$ is bounded. So suppose $I \neq \emptyset$. From $\sum_j \delta^k \lambda_i^k p_{ij} \mu_j^k \geq r_i^-$ for $i \in I$ it may be deduced that there is some $b > 0$ satisfying $\delta^k \lambda_i^k \max_N \mu_j^k \geq b$ for $i \in I$ or

$$\delta^k \min_I \lambda_i^k \max_N \mu_j^k \geq b \quad \text{for some } b \text{ and any odd } k. \tag{7}$$

Dividing (6) by (7) yields $\max_M \lambda_i^k \leq (B/b) \min_I \lambda_i^k$ for k odd. Therefore, if $\min_I \lambda_i^k \leq 1$ then $\lambda_i^k \leq B/b$ for all i so $\boldsymbol{\lambda}$ is bounded.

If, on the other hand, $\min_I \lambda_i^k > 1$ then $I \neq M$ and $\sum_{\bar{I}N} \delta^k \lambda_i^k p_{ij} \mu_j^k = r_{\bar{I}}^-$. But for $i \in \bar{I}$ we know $\lambda_i^k \leq 1$ since $r_i^- = 0$ so

$$h = \sum_{I \times N} \delta^k \lambda_i^k p_{ij} \mu_j^k + \sum_{\bar{I}N} \delta^k \lambda_i^k p_{ij} \mu_j^k \leq r_{\bar{I}}^- + \sum_{\bar{I} \times N} \delta^k p_{ij} \max_N \mu_j^k.$$

$R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ implies $r_I < h$ so

$$\delta^k \max_N \mu_j^k \geq (h - r_{\bar{I}}^-)/p_{\bar{I}N} = b' > 0.$$

Now use (6) to deduce $\max_M \lambda_i^k \leq B/b'$. Thus, in all cases λ^k is bounded for k odd and since (for k odd) $\boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k$, it is bounded for all k. A similar argument shows $\boldsymbol{\mu}^k$ is bounded for all k. □

We conjecture that the same algorithm works for $\boldsymbol{p} \geq \boldsymbol{0}$ when $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ and also that if $R^+(\boldsymbol{p}, \boldsymbol{\sigma}) = \emptyset$ and $R^0(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ the $\boldsymbol{f}^k$ of the algorithm converge to the unique solution $\boldsymbol{f} \in R^0(\boldsymbol{p}, \boldsymbol{\sigma})$ that minimizes the objective function of (2).

2. Proportional matrix apportionments

2.1. Definitions and axioms

A *problem* is a pair $(\boldsymbol{p}, \boldsymbol{\sigma})$ where $\boldsymbol{p}$ and $\boldsymbol{\sigma}$ are defined as before except that the data $\boldsymbol{\sigma}$ are assumed to be integer. The *set of apportionments* of a problem $(\boldsymbol{p}, \boldsymbol{\sigma})$ is composed of the integer valued allocations of $R(\boldsymbol{\sigma})$. Since $R(\boldsymbol{\sigma})$ is nonempty and $\boldsymbol{\sigma}$ is integer valued, the set of apportionments is nonempty.

A *method of apportionment* A is a correspondence that assigns at least one apportionment to every problem: $A(\boldsymbol{p}, \boldsymbol{\sigma})$ is a nonempty subset of $R(\boldsymbol{\sigma})$. For the method A to be "proportional" it should satisfy a number of basic properties. They are essentially the same as those postulated for allocations, but in one case a property is formulated which is deduced from the fair share method.

Axiom 1′ (exactness). If $\boldsymbol{f} = F^*(\boldsymbol{p}, \boldsymbol{\sigma})$ is integer in all components, then $A(\boldsymbol{p}, \boldsymbol{\sigma}) = \{\boldsymbol{f}\}$.

If the unique fair share allocation $\boldsymbol{f}$ happens to be integer valued then it must be the unique apportionment.

Axiom 2′ (relevance). If $A(\boldsymbol{p}, \boldsymbol{\sigma}) \cap R(\hat{\boldsymbol{\sigma}}) \neq \emptyset$ and $R(\hat{\boldsymbol{\sigma}}) \subset R(\boldsymbol{\sigma})$ then $A(\boldsymbol{p}, \hat{\boldsymbol{\sigma}}) = A(\boldsymbol{p}, \boldsymbol{\sigma}) \cap R(\hat{\boldsymbol{\sigma}})$.

This is almost the same "independence of irrelevant alternatives" property as that imposed on allocations F.

Axiom 3′ (uniformity). If $\boldsymbol{a} \in A(\boldsymbol{p}, \boldsymbol{\sigma})$ then the same statements hold as do in Axiom 3 with A replacing F and $\boldsymbol{a}$ replacing $\boldsymbol{f}$.

Again, "any part of a fair apportionment should be fair".

Axiom 4′ (monotonicity). If $\boldsymbol{a} \in A(\boldsymbol{p}, \boldsymbol{\sigma})$, $\boldsymbol{a}' \in A(\boldsymbol{p}', \boldsymbol{\sigma})$ and $\boldsymbol{p}'$ is equal to $\boldsymbol{p}$ except that $p_{kl} < p'_{kl}$ then $a_{kl} \leq a'_{kl}$.

If $\boldsymbol{f} = F^*(\boldsymbol{p}, \boldsymbol{\sigma})$ let $I^- = \{i \in M: f_{iN} = r_i^-\}$, $I^+ = \{i \in M: f_{iN} = r_i^+\}$ and $I^0 = \{i \in M: r_i^- < f_{iN} < r_i^+\}$, and define J^-, J^+, J^0 analogously. Then it is evident from the theorem characterizing fair share allocations that if $\delta > 0$, $\boldsymbol{\alpha} = (\alpha_i) > 0$, $\boldsymbol{\beta} = (\beta_j) > 0$ are reals satisfying $\alpha_i > 1$ for $i \in I^+$, $\alpha_i < 1$ for $i \in I^-$ and $\alpha_i = 1$ for $i \in I^0$ (and analogously for $\boldsymbol{\beta}$), then $\boldsymbol{f} \in F^*(\delta \boldsymbol{\alpha} \boldsymbol{p} \boldsymbol{\beta}, \boldsymbol{\sigma})$. One cannot hope for anything less when looking for integer apportionments. Accordingly, given $\boldsymbol{a} \in A(\boldsymbol{p}, \boldsymbol{\sigma})$ define the corresponding partitions $M = I^- \cup I^+ \cup I^0$ and $N = J^- \cup J^+ \cup J^0$.

Axiom 5' (homogeneity). If $\boldsymbol{a} \in A(\boldsymbol{p}, \boldsymbol{\sigma})$ and $\delta > 0$, $\boldsymbol{\alpha} > \boldsymbol{0}$ and $\boldsymbol{\beta} > \boldsymbol{0}$ are such that

$$\alpha_i > 1 \text{ implies } i \in I^+ \quad \text{and} \quad \alpha_i < 1 \text{ implies } i \in I^-,$$

$$\beta_i > 1 \text{ implies } j \in J^+ \quad \text{and} \quad \beta_j < 1 \text{ implies } j \in J^-,$$

then $\boldsymbol{a} \in A(\delta \boldsymbol{\alpha p \beta}, \boldsymbol{\sigma})$.

Intuitively one would expect small changes in $\boldsymbol{p}$ to leave the apportionment $\boldsymbol{a}$ unchanged. For example, if one population increases and another decreases then presumably no change of apportionment occurs until a point $\boldsymbol{p}^*$ is reached when any further change does cause a change in apportionment. Arbitrarily small changes around $\boldsymbol{p}^*$ can then produce different apportionments. At $\boldsymbol{p}^*$ all these apportionments should be admissible. Thus:

Axiom 6' (completeness). If $\boldsymbol{p}^s \to \boldsymbol{p}$ when s tends to infinity and $\boldsymbol{a} \in A(\boldsymbol{p}^s, \boldsymbol{\sigma})$ for every s, then $\boldsymbol{a} \in A(\boldsymbol{p}, \boldsymbol{\sigma})$.

The Axioms 1' through 6' are consistent and characterize a class of methods of apportionment called divisor methods. The characterization is proved in [3]; our objective here is to describe the class of divisor methods and for this several definitions are necessary.

A *divisor function* is a strictly monotone real function d defined on all nonnegative integers, satisfying $a \leq d(a) \leq a+1$ and $d(b)/(b+1) < d(a)/a$ for all $a \geq 1$ and $b \geq 0$. A *d-rounding* of the real number $x > 0$ is defined by

$$[x]_d = a \quad \text{if } d(a-1) \leq x \leq d(a),$$

and $[0]_d = 0$. So a d-rounding is unique unless $x = d(a)$, in which case $[x]_d = [d(a)]_d = a$ or $a+1$ and there is a *tie*. In effect, $d(a) \in [a, a+1]$ is a threshold below which x is rounded down, above which x is rounded up.

An apportionment matrix $\boldsymbol{a}$ belongs to the *divisor method* A^d *based on* d for a problem $(\boldsymbol{p}, \boldsymbol{\sigma})$ if

$$\boldsymbol{a} = (a_{ij}) = ([\delta \lambda_i p_{ij} \mu_j]_d), \quad \boldsymbol{a} \in R(\boldsymbol{\sigma}),$$

for some $\delta > 0$, $\boldsymbol{\lambda} > \boldsymbol{0}$, $\boldsymbol{\mu} > \boldsymbol{0}$ satisfying:

$$\lambda_i > 1 \text{ implies } a_{iN} = r_i^- \quad \text{and} \quad \lambda_i < 1 \text{ implies } a_{iN} = r_i^+,$$

$$\mu_j > 1 \text{ implies } a_{Mj} = c_j^- \quad \text{and} \quad \mu_j < 1 \text{ implies } a_{Mj} = c_j^+. \tag{8}$$

The set of all divisor method apportionments takes on the role of the fair share method: the integer requirement introduces a multiplicity in the choice of method. Moreover, a single divisor method may admit several apportionments because of ties.

Note that if $d(0) = 0$ then the d-rounding of any positive quantity is at least one, so if $\boldsymbol{a} \in A^d(\boldsymbol{p}, \boldsymbol{\sigma})$ then $p_{ij} > 0$ implies $a_{ij} \geq 1$. Thus, a necessary condition for the existence of apportionments is that

$$R^1(\boldsymbol{p}, \boldsymbol{\sigma}) = \{\boldsymbol{f} \in R^0(\boldsymbol{p}, \boldsymbol{\sigma}) \colon p_{ij} > 0 \text{ implies } a_{ij} \geq 1\}$$

be nonempty. Accordingly a problem is said to be *strongly positive* if $\boldsymbol{p} > \mathbf{0}$ and $R^1(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$. We prove in [3]:

Characterization theorem. *A method of apportionment satisfies Axioms* 1′ *through* 6′ *over the set of strongly positive problems if and only if it is a divisor method. It satisfies them over the set of positive problems if and only if it is a divisor method* A^d *with* $d(0) > 0$. □

2.2. Existence and algorithm

There are several well-known examples of divisor methods that have been used or proposed for apportioning seats in legislatures among regions or political parties [4]. The method of Adams, used in France in 1986 to apportion the Assemblée Nationale among the departments, has $d(a) = a$. The method of Webster, used for many years to apportion seats of the House of Representatives in the United States, and also popular in the Scandinavian countries, has $d(a) = a + \frac{1}{2}$. The method of Jefferson, also known as that of d'Hondt and frequently used for apportioning seats among political parties in P.R. systems, such as within departments in the French elections of 1986, has $d(a) = a + 1$.

Theorem 4. *Let d be a divisor function. If* $d(0) > 0$, $A^d(\boldsymbol{p}, \boldsymbol{\sigma})$ *is nonempty if and only if* $R^0(\boldsymbol{p}, \boldsymbol{\sigma})$ *is nonempty. If* $d(0) = 0$, $A^d(\boldsymbol{p}, \boldsymbol{\sigma})$ *is nonempty if and only if* $R^1(\boldsymbol{p}, \boldsymbol{\sigma})$ *is nonempty.*

Proof. The existence of divisor method apportionments is proved constructively by an algorithm that either provides an A^d apportionment or shows that $R^0(\boldsymbol{p}, \boldsymbol{\sigma})$ or $R^1(\boldsymbol{p}, \boldsymbol{\sigma})$, depending upon the d in hand, is empty.

Divisor method algorithm. At each step it is assumed that a *trial solution* $(\delta, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{a})$ is in hand that satisfies the following conditions:

$$\begin{aligned}
&\delta > 0, \boldsymbol{\lambda} > \mathbf{0}, \boldsymbol{\mu} > \mathbf{0};\\
&a_{ij} = [f_{ij}]_d \quad \text{where } \boldsymbol{f} = \delta \boldsymbol{\lambda} \boldsymbol{p} \boldsymbol{\mu} \text{ and } a_{MN} = h;\\
&\lambda_i > 1 \text{ implies } a_{iN} \leqslant r_i^- \quad \text{and} \quad \lambda_i < 1 \text{ implies } a_{iN} \geqslant c_i^+;\\
&\mu_j > 1 \text{ implies } a_{Mj} \leqslant c_j^- \quad \text{and} \quad \mu_j < 1 \text{ implies } a_{Mj} \geqslant c_j^+. \qquad (9)
\end{aligned}$$

The aim of the algorithm is to produce a matrix $\boldsymbol{a}$ that belongs to $R(\boldsymbol{\sigma})$ because, in the presence of conditions (9), this implies that $\boldsymbol{a}$ is an A^d-apportionment. Given a trial solution $\boldsymbol{a}$, define $I^- = \{i \in M: a_{iN} < r_i^-\}$, $I^+ = \{i \in M: a_{iN} > r_i^+\}$ and J^-, J^+ similarly. The *error* of the trial solution is defined to be the nonnegative integer

$$\sum_{I^-} (r_i^- - a_{iN}) + \sum_{I^+} (a_{iN} - r_i^+) + \sum_{J^-} (c_j^- - a_{Mj}) + \sum_{J^+} (a_{Mj} - c_j^+).$$

If the error is zero, an apportionment is in hand. If not, at most $m + n$ of the steps described below produce a new trial solution whose error is smaller by at least 1.

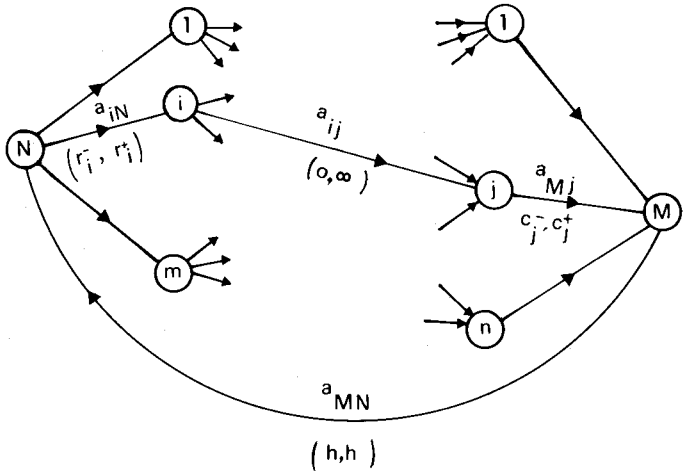


Diagram 1. Circulation network.

The underlying idea comes from the out-of-kilter algorithm: A labelling procedure identifies a change in trial solution that maintains the conditions (9) and strictly decreases the error within a finite number of steps.

An *initial* choice of trial solution that verifies conditions (9) is $\lambda_i = \mu_j = 1$ for all i, j and $\delta > 0$ chosen so that $a_{MN} = h$. This is clearly possible if $d(0) \neq 0$. In the case $d(0) = 0$ the number of nonzero p_{ij}'s cannot be greater than h, else $R^1(\boldsymbol{p}, \boldsymbol{\sigma})$ would be empty, so a $\delta > 0$ can be found. Consider the bipartite network with node set $M \cup N$, and arcs (i, j), with $i \in M$ and $j \in N$, if and only if $p_{ij} > 0$. Add a source called N and arcs (N, i), $i \in M$, with lower and upper capacities equal respectively to r_i^- and r_i^+. Similarly, add a sink called M and arcs (j, M) with lower and upper capacities respectively equal to c_j^- and c_j^+. Finally, add arc (M, N) with lower and upper capacity equal to h. Let $\mathscr{A}$ represent the set of all arcs. A trial solution is a (usually nonfeasible) circulation in this network.

There are several cases that may occur in the algorithm: We describe only the case where the step begins because $I^- \neq \emptyset$. The other cases are treated analogously.

Case $I^- \neq \emptyset$. Declare each arc (N, i), $i \in I^-$, to be a forward arc. Recursively define "labelled" sets of nodes I and J (and perhaps the node M) until either node N is labelled or no further labelling is possible, as follows:

(0) $I^- \subset I$.

(1) If $i \in I, j \notin J, (i, j) \in \mathscr{A}$ and $f_{ij} = d(a_{ij})$ then $j \in J$ and is labelled with $\{i\}$, and (i, j) is declared a forward arc. If $i \in I$, and either $a_{iN} > r_i^+$, or $\lambda_i = 1$ and $a_{iN} > r_i^-$ then label N with $\{i\}$ and declare (N, i) to be a backward arc.

(2) If $j \in J, i \notin I, (i, j) \in \mathscr{A}, f_{ij} = d(a_{ij} - 1)$ and $a_{ij} \geqslant 1$ then $i \in I$ and is labelled with $\{j\}$, and (i, j) is declared a backward arc. If $j \in J$ and either $a_{Mj} < c_j^-$, or $\mu_j = 1$ and $a_{Mj} < c_j^+$ then label M with $\{j\}$, and declare (j, M) to be a forward arc.

(3) If M is labelled, $j \notin J$ and either $a_{Mj} > c_j^+$, or $\mu_j = 1$ and $a_{Mj} > c_j^-$ then $j \in J$ and is labelled with $\{M\}$ and (j, M) is declared a backward arc.

Labelling terminates in one of three possibilities: (i) N is labelled, (ii) neither N nor M are labelled, or (iii) N is not labelled but M is.

(i) *N is labelled.* Then the path indicated by following the labels beginning with the label of N goes to some $i \in I^-$ and identifies with (N, i) a cycle along which the flow in every forward arc may be increased by 1 and along every backward arc decreased by 1 to obtain a new circulation $\boldsymbol{a}'$ satisfying (9) with error decreased by at least 1.

(ii) *Neither N nor M are labelled.* Then the following must hold:

$$\text{if } i \in I \text{ and } a_{iN} > r_i^-, \quad \text{then } \lambda_i < 1, \tag{10}$$

and

$$\text{if } j \in J \text{ and } a_{Mj} < c_j^+, \quad \text{then } \mu_j > 1. \tag{11}$$

Condition (10) holds because $a_{iN} > r_i^-$ implies that $\lambda_i \leq 1$ by (9), but since N is not labelled and i is labelled $\lambda_i \neq 1$, so $\lambda_i < 1$. Condition (11) holds for a similar reason.

Let

$$\varepsilon_1 = \min\{d(a_{ij})/f_{ij} \colon i \in I, j \notin J, (i,j) \in \mathscr{A}\},$$

$$\varepsilon_2 = \min\{f_{ij}/d(a_{ij}-1) \colon i \notin I, j \in J, (i,j) \in \mathscr{A}\},$$

$$\varepsilon_3 = \min\{1/\lambda_i \colon i \in I, a_{iN} > r_i^-\}, \qquad \varepsilon_4 = \min\{\mu_j \colon j \in J, a_{Mj} < c_j^+\},$$

unless ε_i is undefined in which case it is taken to be ∞. Thus, $\varepsilon_i > 1$ in each case and so $\varepsilon = \min_i \varepsilon_i > 1$. If ε is finite define the new trial solution by:

$$\delta' = \delta;$$

$$\lambda_i' = \varepsilon\lambda_i \quad \text{for } i \in I, \qquad \lambda_i' = \lambda_i, \quad \text{else;}$$

$$\mu_j' = \mu_j/\varepsilon \quad \text{for } j \in J, \qquad \mu_j' = \mu_j, \quad \text{else;}$$

$$\boldsymbol{a}' = \boldsymbol{a}.$$

The effect of this change is pictured in Diagram 2. The choice of ε guarantees that the properties (9) remains satisfied. To see this first note that $a_{ij} = [f_{ij}']_d$ for any (i,j) where $f_{ij}' = \delta'\lambda_i' p_{ij}\mu_j'$. If (i,j) is in $I \times J$ or $\bar{I} \times \bar{J}$ this holds because $f_{ij}' = f_{ij}$. If (i,j) is in $I \times \bar{J}$ we have $f_{ij}' = \varepsilon f_{ij}$ thus $d(a_{ij}-1) \leq f_{ij} \leq f_{ij}' \leq \varepsilon_1 f_{ij} \leq d(a_{ij})$ and $a_{ij} = [f_{ij}']_d$. A similar argument applies to (i,j) in $\bar{I} \times J$ by using $\varepsilon_2 \geq \varepsilon$. As for the constraints

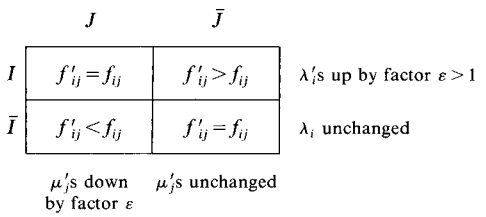
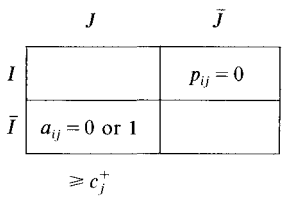


Diagram 2. Changes when neither M nor N labelled. $\boldsymbol{a}' = \boldsymbol{a}$.

on the multipliers note that the only change in row multipliers is for $i \in I$. One must only assure that if $a'_{iN} > r_i^-$ then $\lambda'_i \leq 1$, which is guaranteed by $\varepsilon_3 \geqslant \varepsilon$. A similar argument applies to the columns since $\varepsilon_4 \geqslant \varepsilon$. Moreover, at the next step all labels may be kept and further labelling must take place. $\varepsilon = \varepsilon_1$ means some new column $j \notin J$ will be labelled, $\varepsilon = \varepsilon_2$ that some new row $i \notin I$ will be labelled, $\varepsilon = \varepsilon_3$ that N will be labelled and $\varepsilon = \varepsilon_4$ that M will be labelled.

Suppose, however, that ε is not finite. Then the condition of Diagram 3 obtains (including the possibilities $\bar{I} = \emptyset$ or $\bar{J} = \emptyset$). ε_1 not finite means $p_{ij} = 0$ for $(i, j) \in I \times \bar{J}$ or $\bar{J} = \emptyset$. In either case J is not empty. ε_2 not finite means $d(a_{ij} - 1) = 0$ so either $a_{ij} = 0$ or $a_{ij} = 1$ and $d(0) = 0$ for $(i, j) \in \bar{I} \times J$, or $\bar{I} = \emptyset$. ε_3 not finite means $a_{iN} \leq r_i^-$ for $i \in I$. ε_4 not finite means $a_{Mj} \geqslant c_j^+$ for $j \in J$.



where for $(i, j) \in \bar{I} \times J$,
$d(0) \neq 0$ implies $a_{ij} = 0$,
$d(0) = 0$ implies $a_{ij} = 0$ if $p_{ij} = 0$,
$= 1$ if $p_{ij} > 0$.

Diagram 3. $\varepsilon = \infty$, M and N not labelled.

Consider the case $d(0) \neq 0$. Since $I^- \subset I$, $a_{IN} < r_I^-$, so $c_J^+ \leq a_{MJ} = a_{IJ} = a_{IN} < r_I^-$ where $J = \{j \in N: p_{ij} > 0 \text{ for some } i \in I\}$, showing the conditions (5) are violated and there is no feasible solution in $R^0(\boldsymbol{p}, \boldsymbol{\sigma})$.

If $d(0) = 0$ then $a_{ij} \geqslant 1$ for every $p_{ij} > 0$ is an added constraint. We know that $a_{\bar{I}J} = e_{\bar{I}J}$, where $e_{KL} = \{\text{the number of } p_{ij} > 0: (i, j) \in K \times L\}$. Thus, $c_J^+ \leq a_{MJ} = a_{IJ} + a_{\bar{I}J} = a_{IJ} + e_{\bar{I}J} = a_{IN} + e_{\bar{I}J} < r_I^- + e_{\bar{I}J}$. But there can be no $\boldsymbol{f} \in R^1(\boldsymbol{p}, \boldsymbol{\sigma})$: any feasible $\boldsymbol{f}$ must satisfy $f_{IN} \geqslant r_I$ and $f_{I\bar{J}} = 0$, so $f_{IJ} \geqslant r_I^-$, and $f_{\bar{I}J} \geqslant e_{\bar{I}J}$. Therefore $f_{MJ} = f_{IJ} + f_{\bar{I}J} \geqslant r_I^- + e_{\bar{I}J} > c_J^+$ for any feasible $\boldsymbol{f}$, violating the upper bound constraints on columns J.

(iii) *M is labelled and N is not.* Then (10) and the following obtain:

$$\text{if } j \in \bar{J} \text{ and } a_{Mj} > c_j^-, \quad \text{then } \mu_j < 1. \tag{12}$$

Let ε_1, ε_2 and ε_3 be defined as above, $\varepsilon_4 = \min\{1/\mu_j: j \in \bar{J}, a_{Mj} > c_j^-\}$ and $\varepsilon = \min_i \varepsilon_i > 1$. If ε is finite define the new trial solution by

$$\delta' = \delta/\varepsilon;$$

$$\lambda'_i = \varepsilon\lambda_i \quad \text{for } i \in I, \qquad \lambda'_i = \lambda_i, \quad \text{else;}$$

$$\mu'_j = \varepsilon\mu_j \quad \text{for } j \in \bar{J}, \qquad \mu'_j = \mu_j, \quad \text{else;}$$

$$\boldsymbol{a}' = \boldsymbol{a}.$$

The situation is depicted in Diagram 4.

As before the properties (9) remain satisfied. The next step all labels may be kept and further labelling must take place.

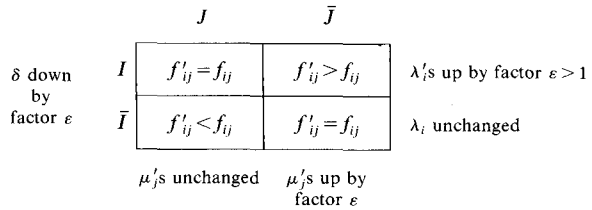


Diagram 4. Changes when M labelled, N not.

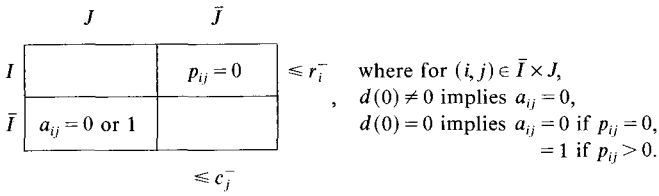


Diagram 5. $\varepsilon = \infty$, M labelled, N not labelled.

Suppose ε is not finite. Then the condition of Diagram 5 obtains (including the possibilities $\bar{I} = \emptyset$ or $\bar{J} = \emptyset$). If $d(0) \neq 0$ deduce $h < r_I^- + c_{\bar{J}}^-$, violating conditions (5) and showing $R^0(\boldsymbol{p}, \boldsymbol{\sigma}) = \emptyset$. If $d(0) = 0$ deduce that $r_I^- + c_{\bar{J}}^- > h - e_{\bar{I}J}$, showing $R^1(\boldsymbol{p}, \boldsymbol{\sigma}) = \emptyset$.

Observe, in summary, that after at most $m + n$ steps either a case (i) must occur and so a decrease in the integer error measure or the problem is found to have no feasible solution. Thus, the algorithm converges in a finite number of steps. Of course, we have only described what to do when $I^- \neq \emptyset$, where the steps are motivated by the desire to increase the values of λ_i for $i \in I^-$. If $I^+ \neq \emptyset$, a similar procedure works where the motivation is to decrease the values of λ_i for $i \in I^+$. The "transpose" procedure works for J^- and J^+. This completes the justification of the algorithm, and so proves the theorem. □

As a byproduct the algorithm proves the following.

Corollary. *$R^0(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ if and only if the conditions (5) hold. $R^1(\boldsymbol{p}, \boldsymbol{\sigma}) \neq \emptyset$ if and only if*

$$c_J^+ - e_{\bar{I}J} \geqslant r_{\bar{I}}^-, \qquad r_{\bar{I}}^+ - e_{\bar{I}J} \geqslant c_{\bar{J}}^-$$

and

$$c_{\bar{J}}^- + r_I^- + e_{\bar{I}J} \leqslant h \leqslant c_J^+ + r_{\bar{I}}^+ - e_{\bar{I}J},$$

for any $I \subset M$, $J \subset N$ with $p_{I\bar{J}} = 0$. □

This algorithm, persuasive as an existence theorem, leaves something to be desired as a method for finding solutions in practice. A "good" initial solution needs to be specified, presumably one that begins with the fair share allocation $\boldsymbol{f}$.

Whereas the set of allocations or proportional solutions in reals $F^*(\boldsymbol{p}, \boldsymbol{\sigma})$ is unique, the set of apportionments or proportional solutions in integers admits any divisor method solution $A^d(\boldsymbol{p}, \boldsymbol{\sigma})$. What d should be chosen? The answer depends on the particular application. If it is desirable that over many problems $(\boldsymbol{p}, \boldsymbol{\sigma})$ with $\boldsymbol{\sigma}$ fixed the average of the a_{ij} be equal to the average of the f_{ij} for each i, j (that there be "no bias" in apportionments), then there is reason to believe that the method of Webster, $d(a) = a + \frac{1}{2}$, should be used. This is suggested by results concerning vector apportionment [4] and has been confirmed by some limited computational trials. It would be a reasonable approach to apportioning seats.

The data of census rounding problems, satisfies $\sum_j p_{ij} = r_i$, $\sum_i p_{ij} = c_j$ and $\sum r_i = \sum c_j = h$. A sensible rounding procedure is to first solve the associated vector rounding problems $(\boldsymbol{r}, h)$ and $(\boldsymbol{c}, h)$, then solve the matrix rounding problem. The current approach to solving the equivalent problem of rounding to integers is to compute the associated *quotas* $\bar{p}_{ij} = p_{ij}h/p_{MN}$, $\bar{r}_i = \sum_j \bar{p}_{ij}$, $\bar{c}_j = \sum_i p_{ij}$, then find an (integer) apportionment $\boldsymbol{a}$ satisfying $\lfloor \bar{p}_{ij} \rfloor \leq a_{ij} \leq \lceil \bar{p}_{ij} \rceil$, $\lfloor \bar{r}_i \rfloor \leq a_{iN} \leq \lceil \bar{r}_i \rceil$, $\lfloor \bar{c}_j \rfloor \leq a_{Mj} \leq \lceil \bar{c}_j \rceil$ and $a_{MN} = h$, that minimizes the distance from the quotas for some measure of distance. ($\lceil x \rceil$ = least integer greater than or equal to x. $\lfloor x \rfloor$ = greatest integer less than or equal to x.) This is not a "proportional" idea: it is akin to using the method of Hamilton in vector apportionment, which admits unfortunate behaviour. The reason for this is simply that rounding large numbers $\bar{p}_{ij}$ (or $\bar{r}_i$, $\bar{c}_j$) up or down to the nearest integers is, from the "proportional" perspective, more restrictive than rounding small numbers up or down to the nearest integers. But understanding the impact of this observation on actual problems awaits computation.

References

[1] J.M. Anthonisse, "Proportional representation in a regional council," CWI Newsletter 5, Centrum voor Wiskunde en Informatica (Amsterdam, December 1984) 22-29.

[2] M. Bacharach, *Biproportional Matrices and Input-Output Change* (Cambridge University Press, Cambridge, 1970).

[3] M.L. Balinski and G. Demange, "An axiomatic approach to proportionality between matrices," *Mathematics of Operations Research* (1989), to appear.

[4] M.L. Balinski and H.P. Young, *Fair Representation: Meeting the Ideal of One Man, One Vote* (Yale University Press, New Haven, CT, 1982).

[5] J. Bigelow and N. Shapiro, "A scaling theorem and algorithms," *SIAM Journal of Applied Mathematics* 33 (1977) 348-352.

[6] L.H. Cox and L.R. Ernst, "Controlled rounding," *INFOR* 20 (1982) 423-432.

[7] D.G. Luenberger, *Introduction to Linear and Nonlinear Programming* (Addison-Wesley, Reading, MA, 1973).

[8] R.T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[9] R. Sinkhorn, "Diagonal equivalence to matrices with prescribed row and column sums," *American Mathematical Monthly* 74 (1967) 402-405.