An Axiomatic Approach to Proportionality between Matrices

Michel L. Balinski, Gabrielle Demange

HAL Id: hal-00686748
https://hal.science/hal-00686748

Submitted on 11 Apr 2012

AN AXIOMATIC APPROACH TO PROPORTIONALITY BETWEEN MATRICES*

M. L. BALINSKI AND G. DEMANGE

Ecole Polytechnique

Given a matrix $\mathbf{p} \geqslant \mathbf{0}$ what does it mean to say that a matrix $\mathbf{f}$ (of the same dimension), whose row and column sums must fall between specific limits, is "proportional to" $\mathbf{p}$? This paper gives an axiomatic solution to this question in two distinct contexts. First, for any real "allocation" matrix $\mathbf{f}$. Second, for any integer constrained "apportionment" matrix $\mathbf{f}$.

In the case of $\mathbf{f}$ real the solution turns out to coincide with what has been variously called biproportional scaling and diagonal equivalence and has been much used in econometrics and statistics. In the case of $\mathbf{f}$ integer the problem arises in the simultaneous apportionment of seats to regions and to parties and also in the rounding of tables of census data.

Introduction. Regional councils in the Netherlands are composed of seats that simultaneously represent both townships and political parties [2]. The stated intent is that each township should receive a number of seats proportional to its population and each party a number of seats proportional to its total vote. The ideal or proportionality is as old as the hills and by now commonplace: given a vector of real numbers $\mathbf{p} = (p_1, \ldots, p_n)$, $p_j \geqslant 0$, $\mathbf{f} = (f_1, \ldots, f_n)$ summing to $h > 0$ is proportional to $\mathbf{p}$ if $\mathbf{f} = \lambda\mathbf{p}$ for $\lambda > 0$, and that is the end of the story.

But when $\mathbf{f}$ is subject to other constraints—for example, $c_j^- \leqslant f_j \leqslant c_j^+$ for all j—then what *should* it mean to say that $\mathbf{f}$ is "proportional to" $\mathbf{p}$? And when, in addition, the f_j are to be numbers of seats and so integer, what then? This is the (vector) *apportionment problem*: how to allocate representation in a house of h seats among regions or townships (of populations $\mathbf{p}$) or among political parties (having vote totals $\mathbf{p}$). A theory of apportionment has been developed that gives a meaning to proportionality in this setting [6], and so an approach to determining how many seats each township should receive and how many each party should receive in a Dutch regional council. But if p_{ij} is the number of votes in township i for party j how many candidates on the list of party j in township i should be declared elected? The reflex response is, proportional numbers: but what precisely does this mean? There are two types of constraints. One type demands that the total number allocated to a township add up to the number it deserves, and to a party the number it deserves; the other type asks that these allocations be integer values. Our aim is to arrive at axiomatic definitions of proportionality between matrices when one must satisfy (i) the first constraints only (the real case) and (ii) both types of constraints (the integer case).

The simpler real case is addressed first: given a matrix $\mathbf{p} \geqslant \mathbf{0}$ of real numbers what should it mean to say that a matrix of real numbers subject to bounds on row and column sums is "proportional to" $\mathbf{p}$? Axioms are proposed that yield a unique solution (when such exist) which turns out to have been the subject of prior work much used in

a wide variety of contexts. Referred to as diagonal equivalence [10], [18] and bipropor-tional scaling [4] it has been used in statistics and econometrics for adjusting contin-gency tables [14], balancing social accounting matrices [17] and estimating traffic flows (see [16]) or international trade accounts [15], and in allocating marketing budgets. A comprehensive overview of approaches and uses was collected in one volume in 1976 [1].

The unique solution minimizes an entropy function, a fact frequently noted, but no other justification has previously been forthcoming. It is also easily computed by alternately scaling rows and columns (the "RAS" method), as many people have discovered independently. However, the prior work has treated the real case with equation constraints (as indicated above) whereas we treat the more general case of inequality constraints, where the result obtained gives additional credence to the claim that the sense of proportionality has been properly extended.

The integer case is addressed next: the identical question is posed except that now the matrix $\mathbf{f}$ is further constrained to be integer valued. The set of axioms used are in spirit identical to those used in the real case, but in the integer case uniqueness cannot be realized. Indeed, without further restrictions (that should be chosen as a function of the application at hand), our axioms determine an infinite class of "divisor methods", any one of which can yield more than one solution. However, the set of solutions of any one method are unique "up to ties", and solutions of different methods bear close relationships to each other.

Part I describes the axiomatic approach to proportionality between matrices in reals, and characterizes when solutions exist. Part II treats the integer case.

I. Proportionality in reals.

I.1. *Definitions and axioms.* A *problem* is a pair $(\mathbf{p}, \sigma)$, where $\mathbf{p} = (p_{ij}) \geq \mathbf{0}$ is a matrix containing no row or column of zeros, $i \in M = \{1, 2, \ldots, m\}$ and $j \in N = \{1, 2, \ldots, n\}$, and $\sigma = (\mathbf{r}^-, \mathbf{r}^+, \mathbf{c}^-, \mathbf{c}^+, h)$ is a nonnegative vector with $\mathbf{r}^- = (r_i^-)$ and $\mathbf{r}^+ = (r_i^+)$, $i \in M$, $\mathbf{c}^- = (c_j^-)$ and $\mathbf{c}^+ = (c_j^+)$, $j \in N$, and h a positive scalar. The *region of allocations* $R(\sigma)$ is

$$R(\sigma) = \left\{ \mathbf{f} = (f_{ij}) \geq \mathbf{0}: r_i^- \leq f_{iN} \leq r_i^+, i \in M, c_j^- \leq f_{Mj} \leq c_j^+, j \in N, f_{MN} = h \right\}$$

where $t_{IJ} = \{\Sigma t_{ij}: i \in I \subset M \text{ and } j \in J \subset N\}$.

From now on, we consider only nonempty regions of allocations: $r_i^- \leq r_i^+$ for all i in M, $r_M^- \leq h \leq r_M^+$ and similarly for columns.

The central question of the first part of the paper is: what does it mean to say that an allocation $\mathbf{f} \in R(\sigma)$ is proportional to $\mathbf{p}$ and when do such allocations exist?

It will be necessary to distinguish several special cases of problems $(\mathbf{p}, \sigma)$. If $\mathbf{p} > \mathbf{0}$ the problem is *positive*. If $\mathbf{r}^- = \mathbf{r}^+$ and $\mathbf{c}^- = \mathbf{c}^+$ the problem is *equality constrained*, and we write $\mathbf{r} = \mathbf{r}^- = \mathbf{r}^+$ and $\mathbf{c} = \mathbf{c}^- = \mathbf{c}^+$. If $\mathbf{r}^- = \mathbf{0}$, $\mathbf{c}^- = \mathbf{0}$, and $r_i^+ \geq h$, $i \in M$, and $c_j^+ \geq h$, $j \in N$, the problem is *free* so we may drop the superfluous parameters and write the problem $(\mathbf{p}, h)$. If either $m = 1$ or $n = 1$, it is a *vector problem*. Throughout $\bar{I}$ denotes the complement of I in M, and $\bar{J}$ the complement of J in N. A problem is said to be *weakly irreducible* if $r_i^+ > 0$ for all i and $c_j^+ > 0$ for all j, and $r_I^- < h$ and $c_J^- < h$ for every proper subset $I \subset M$ and $J \subset N$. Weak irreducibility simply excludes obvious instances where $\mathbf{f} \in R(\sigma)$ implies $\mathbf{f}$ has some row or column of zeros.

A *method of allocation* F is a rule or mapping that assigns at least one allocation to every problem: $F(\mathbf{p}, \sigma)$ is a nonempty subset of $R(\sigma)$. The possibility of multiple

allocations is not excluded. The ideal of proportionality suggests that F should satisfy a number of elementary properties. These are specified in the following list of axioms. They are justified below.

AXIOM 1 (Exactness).   *If* $(\mathbf{p}, \boldsymbol{\sigma})$ *is free then* $\mathbf{f} = (h/p_{MN})\mathbf{p} = F(\mathbf{p}, \boldsymbol{\sigma})$.

AXIOM 2 (Relevance).   *If* $F(\mathbf{p}, \boldsymbol{\sigma}) \cap R(\hat{\boldsymbol{\sigma}}) \neq \varnothing$ *and* $R(\hat{\boldsymbol{\sigma}}) \subset R(\boldsymbol{\sigma})$ *then* $F(\mathbf{p}, \hat{\boldsymbol{\sigma}}) \subset$ $F(\mathbf{p}, \boldsymbol{\sigma}) \cap R(\hat{\boldsymbol{\sigma}})$.

Let $\mathbf{t}_{I \times J}$ be the submatrix of $\mathbf{t}$ defined on rows $I \subset M$ and columns $J \subset N$. Given $\mathbf{f} \in F(\mathbf{p}, \boldsymbol{\sigma})$ the *subproblem* $(\mathbf{p}_{I \times J}, \boldsymbol{\sigma}_{I \times J})$ has the region $R(\boldsymbol{\sigma}_{I \times J})$ 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* $\hat{\mathbf{f}} \in F(\mathbf{p}, \boldsymbol{\sigma})$ *then* $\mathbf{f}_{I \times J} \in F(\mathbf{p}_{I \times J}, \boldsymbol{\sigma}_{I \times J})$; *and, conversely, if* $\mathbf{g}_{I \times J}$ *is another F-allocation of the subproblem then* $\mathbf{g}$ *defined to be equal to* $\mathbf{g}_{I \times J}$ *on* $I \times J$ *and* $\mathbf{f}$ *elsewhere is also an F-allocation:* $\mathbf{g} \in F(\mathbf{p}, \boldsymbol{\sigma})$.

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

AXIOM 5 (Homogeneity).   *Suppose* $(\mathbf{p}, \boldsymbol{\sigma})$ *is equality constrained. If two rows of* $\mathbf{p}$ *are proportional,* $\mathbf{p}_{i.} = \lambda \mathbf{p}_{k.}$, *and are constrained to the same sum,* $r_i = r_k$, *then the corresponding rows of any* $\mathbf{f} \in F(\mathbf{p}, \boldsymbol{\sigma})$ *are identical:* $\mathbf{f}_{i.} = \mathbf{f}_{k.}$ *(and the same holds for columns).*

Exactness asks that if the usual notion of proportionality works it is the unique solution. Axiom 2 is a kind of "independence of irrelevant alternatives" property and says that if the extra room of $R(\boldsymbol{\sigma})$ lands one in the smaller $R(\hat{\boldsymbol{\sigma}})$ then surely one can do no better in $R(\hat{\boldsymbol{\sigma}})$ itself. Uniformity is the familiar property that any part of a proportional solution must itself be proportional. It is a key idea in many fair division problems, and often called "consistency" ([7], [3], [21]). Homogeneity simply requires answers consistent with what one would expect for vector proportionality.

I.2.   *The fair share matrix.*   It will be convenient to use the notations $\mathbf{py} = (p_{ij} y_j)$ and $\mathbf{xpy} = (x_i p_{ij} y_j)$ where $\mathbf{x} = (x_i)$ and $\mathbf{y} = (y_j)$.

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

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

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

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

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

An alternative description is sometimes useful. A matrix $\mathbf{f}$ satisfies conditions (2) for a problem $(\mathbf{p}, \boldsymbol{\sigma})$ if

$$(2) \qquad\qquad \mathbf{f} = \mathbf{xpy}, \quad \mathbf{f} \in R(\boldsymbol{\sigma})$$

for some $\mathbf{x} > \mathbf{0}$, $\mathbf{y} > \mathbf{0}$ satisfying

$x_i > x_k$ implies $f_{iN} = r_i^-$ or $f_{kN} = r_k^+$,

$y_j > y_l$ implies $f_{Mj} = c_j^-$ or $f_{Ml} = c_l^+$.

LEMMA 1. *If* **f** *is a fair share matrix then it satisfies conditions* (2). *Conversely, if conditions* (2) *hold and* (**p**, σ) *is weakly irreducible, then* **f** *is a fair share matrix.*

PROOF. If **f** is a fair share matrix it clearly satisfies (2). Conversely, if (2) holds there is an $\alpha > 0$ such that $x_i \leqslant \alpha \leqslant x_k$ for every i and k satisfying $f_{iN} > r_i^-$ and $f_{kN} < r_k^+$. Indeed, if $f_{kN} = r_k^+$ for every $k \in M$ take $\alpha = \max x_i > 0$. If for some k, $f_{kN} < r_k^+$ then take $\alpha = \min\{x_k : f_{kN} < r_k^+\}$. It is strictly positive since otherwise, $x_k = 0$ and $0 = f_{kN} < r_k^+$ for some k, so that by (2), for all $x_i > 0$, $f_{iN} = r_i^-$ must hold. This implies that $h = r_{M-k}^-$, contradicting weak irreducibility. Choose similarly a $\beta > 0$ for the columns. Now take $\lambda_i = x_i/\alpha$ and $\mu_j = y_j/\beta$. ∎

LEMMA 2. *Let* (**p**, σ) *and* (**p**′, σ) *be two problems, and* **f** *and* **f**′ *associated fair share matrices. If* $p_{kl} \leqslant p'_{kl}$ *for some* (k, l) *and* $p_{ij} = p'_{ij}$ *for* $(i, j) \neq (k, l)$ *then* $f_{kl} \leqslant f'_{kl}$.

PROOF. The argument is by contradiction. Suppose $f_{kl} > f'_{kl}$ and let **x**, **y** and **x**′, **y**′ be the respective multipliers.

If $f_{ij} > f'_{ij}$ then $x_i y_j > x'_i y'_j$. This follows from $p_{ij} \leqslant p'_{ij}$ and $x_i p_{ij} y_j > x'_i p'_{ij} y'_j$.

If $f_{ij} < f'_{ij}$ then $x_i y_j < x'_i y'_j$. Indeed, $(i, j) \neq (k, l)$ so $p_{ij} = p'_{ij}$, implying $p_{ij} \neq 0$.

Suppose first that there is a "loop" of entries of $\mathbf{f} - \mathbf{f}'$, $(i_1, j_1), (i_2, j_1)$, $(i_2, j_2), \ldots, (i_t, j_t), (i_1, j_t)$, where entries corresponding to (i_s, j_s) are positive and entries corresponding to (i_{s+1}, j_s) are negative (including $i_{t+1} = i_1$). So

$$x_{i_s} y_{j_s} > x'_{i_s} y'_{j_s} \quad \text{and} \quad x'_{i_{s+1}} y'_{j_s} > x_{i_{s+1}} y_{j_s}, \qquad s = 1, \ldots, t,$$

and multiplying all these $2t$ inequalities yields $1 > 1$: a contradiction.

Given $\mathbf{f} - \mathbf{f}' \neq \mathbf{0}$ and the fact that loops of alternating signs are forbidden there is some row or column where the respective sums over **f** and **f**′ are different, say $f_{1N} > f'_{1N}$. We will show below that then there exists some $j \in N$ such that $f_{Mj} > f'_{Mj}$ and $x_1 y_j > x'_1 y'_j$. But $f_{1N} > f'_{1N}$ implies there is some other row, say row 2, with $f_{2N} < f'_{2N}$. An identical argument shows that there exists some $g \in N$ with $f_{Mg} < f'_{Mg}$ and $x_2 y_g < x'_2 y'_g$, and thus $x'_1 x_2 y'_j y_g < x_1 x'_2 y_j y'_g$. This gives the sought for contradiction: $f_{1N} > f'_{1N}$ and $f_{2N} < f'_{2N}$ implies $x_1 \leqslant x_2$ and $x'_1 \geqslant x'_2$, whereas $f_{Mj} > f'_{Mj}$ and $f_{Mg} < f'_{Mg}$ implies $y_j \leqslant y_g$ and $y'_j \geqslant y'_g$ so $x'_1 x_2 y'_j y_g \geqslant x_1 x'_2 y_j y'_g$.

It remains to show that given $f_{1N} > f'_{1N}$, there is a j in N such that $f_{Mj} > f'_{Mj}$ and $x_1 y_j > x'_1 y'_j$. Label row 1 and continue to label as follows:

—if row i is labelled, and a column j with $f_{ij} - f'_{ij} > 0$ is not, then label j with $\{i\}$,

—if column j is labelled, and a row i with $f_{ij} - f'_{ij} < 0$ is not, then label i with $\{j\}$,

until either (a) some row i is labelled with $f_{iN} < f'_{iN}$ or (b) no further labelling is possible.

If (a) occurs $i = i_1$ labelled with $\{j_1\}$, j_1 labelled with $\{i_2\}, \ldots$, on to row 1 singles out a path of alternating signs with $x_{i_s} y_{j_s} < x'_{i_s} y'_{j_s}$ for the negative signs, and $x_{i_{s+1}} y_{j_s} > x'_{i_{s+1}} y'_{j_s}$ for the positive ones. Multiplying as before yields the result $x_i x'_1 < x'_i x_1$. But $r_1^+ \geqslant f_{1N} \geqslant f'_{1N} \geqslant r_1^-$ and $r_i^+ \geqslant f'_{iN} > f_{iN} \geqslant r_i^-$ so $x_i \geqslant x_1$ and $x'_1 \geqslant x'_i$ implying $x_i x'_1 \geqslant x'_i x_1$, a contradiction.

This leaves (b) as the only possibility. Let I be the set of labelled rows ($\bar{I}$ the complement), J the set of labelled columns ($\bar{J}$ the complement). (Refer to Diagram 1 for the sequel.) By the labelling rules $f_{ij} \geqslant f'_{ij}$ for $(i, j) \in \bar{I} \times J$ and $f_{ij} \leqslant f'_{ij}$ for $(i, j) \in I \times \bar{J}$. Moreover, since (a) did not occur, $f_{iN} \geqslant f'_{iN}$ for $i \in I$, and since $f_{1N} > f'_{1N}$, $f_{IN} > f'_{IN}$ and therefore $f_{IJ} - f'_{IJ} = (f_{IN} - f'_{IN}) - (f_{I\bar{J}} - f'_{I\bar{J}}) > 0$. This in turn implies $f_{MJ} > f'_{MJ}$ and, therefore, for some $j \in J$, $f_{Mj} > f'_{Mj}$.

	J	$\bar{J}$
I	$f_{IJ} > f'_{IJ}$	$f_{ij} \leqslant f'_{ij}$
$\bar{I}$	$f_{ij} \geqslant f'_{ij}$	

$f_{IN} \geqslant f'_{IN}$ so $f_{IN} > f'_{IN}$

so $f_{MJ} > f'_{MJ}$

DIAGRAM 1.   Case (b) after labelling.

Since $j \in J$, there is a path of alternating signs from column j to row 1 beginning and ending with positive signs. Multiplying as before yields $x_1 y_j > x'_1 y'_j$, as was to be shown. The argument for the case $f_{2N} < f'_{2N}$ is the same except that the signs are reversed. ∎

THEOREM 1.   *If* $(\mathbf{p}, \boldsymbol{\sigma})$ *is positive and weakly irreducible, a fair share matrix exists and is unique.*

PROOF. If a fair share matrix exists, uniqueness is immediate by Lemma 2. Existence is established via a fixed point argument.

To construct the mapping observe, first, that the result holds for any vector problem $(m = 1)$, $(\mathbf{p}, \mathbf{c}^-, \mathbf{c}^+, h)$. To see this, by weak irreducibility there is a $\delta > 0$ such that $\sum_j \text{mid}(c_j^-, \delta p_j, c_j^+) = h$, where $\text{mid}(x, y, z)$ is the middle in value of the three numbers x, y, z. The vector $\mathbf{f} = (f_j)$ where $f_j = \text{mid}(c_j^-, \delta p_j, c_j^+)$ is the fair share vector: if $c_j^- \leqslant \delta p_j \leqslant c_j^+$ take $\mu_j = 1$; if $\delta p_j < c_j^-$ take $\mu_j = c_j^-/\delta p_j > 1$; and if $c_j^+ < \delta p_j$ take $\mu_j = c_j^+/\delta p_j < 1$.

Define $\varphi(\mathbf{p}, \mathbf{c}^-, \mathbf{c}^+, h)$ to be the function with value equal to the unique fair share vector $\mathbf{f}$. The mapping φ is continuous in $\mathbf{p}$. For consider a sequence $\mathbf{p}^s$ converging to $\mathbf{p}$ and the corresponding sequence of fair shares $\mathbf{f}^s$. Since the $\mathbf{f}^s$ all belong to the compact set $\{\mathbf{f} \geqslant 0 \colon f_N = h\}$, a subsequence converges to some $\mathbf{f}^0$; moreover the δ^s such that $f_j^s = \text{mid}(c_j^-, \delta^s p_j^s, c_j^+)$ may be chosen from the bounded set $\min_j c_j^-/p_j^s \leqslant \delta^s \leqslant \max_j c_j^+/p_j^s$. Letting δ be an accumulation point of the δ^s, $f_j^0 = \text{mid}(c_j^-, \delta p_j, c_j^+)$, so $\mathbf{f}^0$ is the fair share of $\mathbf{p}$ and the continuity of φ is proven.

Now, let $(\mathbf{p}, \boldsymbol{\sigma})$ be a matrix problem and $S = \{\mathbf{x} = (x_i) \geqslant 0 \colon x_M = 1\}$. Construct the mapping $\Phi = \Psi \circ \nu$ with

$$\nu(\mathbf{x}) = \mathbf{y} \quad \text{where } y_j = \varphi_j(\boldsymbol{\eta}, \mathbf{c}^-, \mathbf{c}^+, h)/\eta_j \quad \text{and } \eta_j = \sum_i x_i p_{ij}, \quad \text{and}$$

$$\Psi(\mathbf{y}) = \mathbf{z}/\sum_i z_i \quad \text{where } z_i = \varphi_i(\boldsymbol{\xi}, \mathbf{r}^-, \mathbf{r}^+, h)/\xi_i \quad \text{and } \xi_i = \sum_j p_{ij} y_j.$$

One can think of φ as a process of repeated adjustment: given a set of row multipliers $\mathbf{x}$, obtain a row vector problem and define $\mathbf{y}$ to be the column multipliers for the corresponding fair share row vector; then given $\mathbf{y}$ obtain a column vector problem and define a new normalized set of row multipliers $\mathbf{x}$ (namely, $\mathbf{z}/\sum_i z_i$). From this perspective Φ generalizes the standard algorithmic approach—the RAS method—to finding the fair share matrix in the equality constrained case.

The positivity of $\mathbf{p}$ implies $\boldsymbol{\eta} > \mathbf{0}$, and $\sum_j \eta_j y_j = h$ together with $\mathbf{y} \geqslant \mathbf{0}$ implies $\boldsymbol{\xi} > \mathbf{0}$. Thus Φ is well defined and, by the continuity of φ, is continuous on the simplex S, and so by Brouwer's theorem must have a fixed point $\mathbf{x}^*$. Define $\mathbf{y}^* = \nu(\mathbf{x}^*)$: then $\mathbf{f}^*$ with $f_{ij}^* = x_i^* p_{ij} y_j^*$ is a fair share matrix of $(\mathbf{p}, \boldsymbol{\sigma})$. To see this, using $\boldsymbol{\eta}^*$, $\boldsymbol{\xi}^*$ and $\mathbf{z}^*$ to designate the associated values, note, first, that $f_{MN} = \sum_j(\sum_i x_i^* p_{ij}) y_j^* = \sum_j \eta_j^* y_j^* = h$.

Second, observe that $\mathbf{z}^* = \mathbf{x}^*$: by definition of Φ and of $\mathbf{x}^*$, $\mathbf{x}^* = \Phi(\mathbf{x}^*) = \alpha \mathbf{z}^*$ for some $\alpha > 0$ so $f_{MN} = \Sigma_i x_i^* (\Sigma_j p_{ij} y_j^*) = \alpha(\Sigma_i z_i^* \xi_i^*) = \alpha h$, thus $\alpha = 1$. Finally, since $\mathbf{y}^*$ is a set of multipliers for a vector problem it is already known that $y_j^* > y_i^*$ implies $y_j^* \eta_j^* = c_j^-$ or $y_{i'}^* \eta_i^* = c_{i'}^+$, or $f_{Mj}^* = \Sigma_i x_i^* p_{ij} y_j^* = c_j^-$ or $f_{Ml} = c_l^+$. The analogous conditions for $\mathbf{x}$ are verified similarly, and this completes the proof. ∎

We may now refer unequivocally to "the fair share method" F^*. It should be remarked that a fair share matrix minimizes the entropy function over the feasible set $R(\sigma)$. The existence and uniqueness of solutions may be obtained by standard convex optimization techniques. This is well known for the equality constrained problem (see, e.g., [4]). The RAS method may be generalized to the inequality problem and interpreted as a cyclic coordinate ascent method for maximizing the dual program [5] (see also [11] for interval convex programs).

I.3. Characterization.

THEOREM 2. *The fair share method F^* is the unique method of allocation satisfying Axioms 1 through 5 over the class of positive, weakly irreducible problems.*

PROOF. *Sufficiency.* We begin by checking that F^* satisfies the axioms. Exactness is immediate. F^* is relevant because suppose (the unique) $\mathbf{f} = F^*(\mathbf{p}, \sigma) \in R(\hat{\sigma})$ and $R(\hat{\sigma}) \subset R(\sigma)$. By Lemma 1 $\mathbf{f} = (f_{ij})$ may be defined by conditions (2). If $x_i > x_k$ then either $f_{kN} = r_k^+$ or $f_{iN} = r_i^-$. We then claim that either $f_{kN} = \hat{r}_k^+$ or $f_{iN} = \hat{r}_i^-$. For suppose not: then $f_{kN} < \hat{r}_k^+$ and $f_{iN} > \hat{r}_i^-$. Choose a column l such that $f_{il} > 0$ and define $\hat{\mathbf{f}}$ to be equal to $\mathbf{f}$ except for indices (k, l) and (i, l) where $\hat{f}_{kl} = f_{kl} + \epsilon$ and $\hat{f}_{il} = f_{il} - \epsilon$. For $\epsilon > 0$ and small, $\hat{\mathbf{f}} \in R(\hat{\sigma})$ but $\hat{\mathbf{f}} \notin R(\sigma)$, an contradiction. The remaining conditions (2) showing that $\mathbf{f}$ is the fair share matrix for the problem $(\mathbf{p}, \hat{\sigma})$ are verified similarly. Uniformity and homogeneity are immediate, and monotonicity is a consequence of Lemma 2.

Necessity is established in three stages beginning with vector problems, then going on to matrix equality constrained problems and finally to the general problem. F is any method of allocation satisfying the Axioms 1 through 5.

Vector problems. Let $(\mathbf{p}, \mathbf{c}^-, \mathbf{c}^+, h)$ be a vector problem and $\mathbf{f} = (\delta p_j \mu_j)$ be its fair share. By exactness $\mathbf{f} = F((p_j \mu_j), \mathbf{c}^-, \mathbf{c}^+, h)$. Suppose $\mu_1 \neq 1$, say $\mu_1 < 1$ and let $\mathbf{f}' \in F((p_1, p_2 \mu_2, \ldots, p_n \mu_n), \mathbf{c}^-, \mathbf{c}^+, h)$. By monotonicity $f_1' \geqslant f_1$, but $\mu_1 < 1$ implies $f_1 = c_1^+$ so that $f_1' = f_1 = c_1^+$. Therefore, by uniformity, $(f_2, \ldots, f_n)$ and $(f_2', \ldots, f_n')$ are both allocations of $F((p_2 \mu_2, \ldots, p_n \mu_n), (c_2^-, \ldots, c_n^-), (c_2^+, \ldots, c_n^+), h - c_1^+)$. But again by exactness $(f_2, \ldots, f_n) = (\delta p_2 \mu_2, \ldots, \delta p_n \mu_n)$ is the unique F-allocation so $f_j = f_j'$ for $j = 2, \ldots, m$ and $\mathbf{f} \in F(p_1, p_2 \mu_2, \ldots, p_n \mu_n, \mathbf{c}^-, \mathbf{c}^+, h)$. Repeating the same argument at most n times shows $\mathbf{f} = F(\mathbf{p}, \mathbf{c}^-, \mathbf{c}^+, h)$.

Matrix equality constrained problems. Given an $m \times n$ matrix problem $(\mathbf{p}, \sigma)$, where row sums are constrained to sum to $\mathbf{r}$ and column sums to $\mathbf{c}$, consider the $m \times 2n$ problem $((\mathbf{p}, \mathbf{p}\mathbf{y}), 2\mathbf{r}, (\mathbf{c}, \mathbf{c}), 2h)$ for some $\mathbf{y} > \mathbf{0}$. By the homogeneity of F an F-allocation must have the form $(\mathbf{f}, \mathbf{f})$; by uniformity $\mathbf{f} \in F(\mathbf{p}, \mathbf{r}, \mathbf{c}, h)$ and, moreover, if $\mathbf{g}$ were another F-allocation of the original problem, then $(\mathbf{f}, \mathbf{g})$ would have to be an F-allocation of the replicated problem. $\mathbf{g} \neq \mathbf{f}$ contradicts homogeneity so F assigns a unique allocation to any equality constrained problem. Furthermore, $\mathbf{f}$ belongs also to $F(\mathbf{p}\mathbf{y}, \mathbf{r}, \mathbf{c}, h)$ and so is its unique allocation. By the same argument on columns one concludes that $F(\mathbf{x}\mathbf{p}\mathbf{y}, \sigma) = F(\mathbf{p}, \sigma)$ for any equality constrained problem and any $\mathbf{x} > \mathbf{0}$, $\mathbf{y} > \mathbf{0}$. Choose then the multipliers $\mathbf{x}^*$ and $\mathbf{y}^*$ that give the fair shares $\mathbf{f}^*$ of $(\mathbf{p}, \sigma)$: by exactness $\mathbf{f}^* = F(\mathbf{x}^* \mathbf{p} \mathbf{y}^*, \sigma)$, so the fair share matrix is the unique allocation of $F(\mathbf{p}, \sigma)$.

The general case. Suppose $\mathbf{f}^* \in F(\mathbf{p}, \sigma)$, and let $R(\sigma^*) = \{\mathbf{f} \geqslant \mathbf{0}: f_{iN} = f_{iN'}^*, f_{Mj} = f_{Mj}^*$ and $f_{MN} = h\}$. Since F is relevant, $\mathbf{f}^* \in F(\mathbf{p}, \sigma^*)$, and by the results of the previous paragraph, $\mathbf{f}^* = \mathbf{xpy}$ for some $\mathbf{x} > \mathbf{0}$ and $\mathbf{y} > \mathbf{0}$. To see that the conditions (2) are verified suppose that $f_{Mj}^* > c_j^-$ and $f_{M_l}^* < c_l^+$ and let i be any row. Then, by uniformity, (f_{ij}^*, f_{il}) is an F-solution to the 2-vector problem with vector (p_{ij}, p_{il}) and region $R' = \{(f_{ij}, f_{il}): c_k^- - f_{Mk}^* + f_{ik}^* \leqslant f_{ik} \leqslant c_k^+ - f_{Mk}^* + f_{ik}$ for $k = j, l$, and $f_{ij} + f_{il} = f_{ij}^* + f_{il}^*\}$, and, since a vector, we already know it must be the fair share for this vector problem. But $f_{il}^* < c_l^+ - f_{M_l}^* + f_{il}^*$ and $f_{ij}^* > c_j^- - f_{Mj}^* + f_{ij}^*$, so $f_{ij}^*/p_{ij} \leqslant f_{il}^*/p_{il}$ implying $y_j \leqslant y_l$. The same argument applies for $\mathbf{x}$. Therefore, $\mathbf{x}$ and $\mathbf{y}$ satisfy the conditions (2) and $\mathbf{f}^*$ is the fair share matrix of $(\mathbf{p}, \sigma)$, completing the proof. ∎

I.4. *Extending fair shares to matrices with zeros.* When $\mathbf{p} \geqslant \mathbf{0}$ is not strictly positive, a fair share matrix does not necessarily exist even when $\mathbf{p}$ contains no row or column of zeros and is weakly irreducible, a fact known and analyzed in the equality constrained problem [4], [19]. The object of this section is to give conditions under which the definition of a fair share can be extended in a unique way by continuity. As a by-product the continuity of the fair shares over the set of positive weakly irreducible problems is proven. Consider first two examples. Both are equality constrained problems.

EXAMPLE 1.

$$\mathbf{p} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{r} = (1,1), \quad \mathbf{c} = (1,1).$$

All the allocations in $R(\sigma)$ are as follows;

$$\mathbf{f}(\alpha) = \begin{bmatrix} 1 - \alpha & \alpha \\ \alpha & 1 - \alpha \end{bmatrix} \quad \text{with } 0 \leqslant \alpha \leqslant 1.$$

Since a fair share has exactly the same zeros as $\mathbf{p}$, none can exist here. However one can define an extended fair share without ambiguity. Indeed take a sequence $\mathbf{p}^s$ converging to $\mathbf{p}$. The fair share of $(\mathbf{p}^s, \sigma^s)$ is the matrix $\mathbf{f}(\alpha^s)$ where α^s is the unique positive solution to

$$\frac{(1 - \alpha)^2}{p_{11}^s p_{22}^s} = \frac{\alpha^2}{p_{12}^s p_{21}^s}.$$

Since $p_{11}^s p_{22}^s$ tends to 1 and $p_{12}^s p_{21}^s$ tends to zero, α^s converges to zero and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ may be qualified as "a fair share".

This is not always possible as the second example shows.

EXAMPLE 2.

$$\mathbf{p} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{r} = (2,2,2), \quad \mathbf{c} = (1,1,4) \quad \text{and} \quad h = 6.$$

If we consider the sequence:

$$\mathbf{p}^s = \begin{bmatrix} 1 & 1 & 1/s \\ 1 & 1 & 1/s \\ 1 & 1 & 1 \end{bmatrix},$$

the corresponding fair shares tend to

$$\begin{bmatrix} 1/2 & 1/2 & 1 \\ 1/2 & 1/2 & 1 \\ 0 & 0 & 2 \end{bmatrix},$$

whereas for the sequence

$$\begin{bmatrix} 1 & 1 & 1/s \\ 1 & 1 & 1/s^2 \\ 1 & 1 & 1 \end{bmatrix}$$

they tend to

$$\begin{bmatrix} 0 & 0 & 2 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

A matrix $\mathbf{f}$ is an *extended fair share* for the problem $(\mathbf{p}, \sigma)$ if there exists some sequence of problems $(\mathbf{p}^s, \sigma^s)$, $\mathbf{p}^s > \mathbf{0}$ converging to $(\mathbf{p}, \sigma)$ such that the sequence $F^*(\mathbf{p}^s, \sigma^s)$ converges to $\mathbf{f}$.

Any problem with a nonempty feasible region admits an extended fair share: there always exists a sequence of positive weakly irreducible converging to it (just relax the constraints defined by σ) and the corresponding fair shares are in a compact set. So the interesting question is the uniqueness of this extension. It turns out to be closely related to the nonemptiness of the following subset of $R(\sigma)$:

$$R^0(\mathbf{p}, \sigma) = \left\{ \mathbf{f} \in R(\sigma) : f_{ij} = 0 \text{ if } p_{ij} = 0 \right\}.$$

In the first example this set is nonempty whereas it is empty in the second one.

THEOREM 3. *If the set $R^0(\mathbf{p}, \sigma)$ is nonempty, then the extended fair share of the problem $(\mathbf{p}, \sigma)$ is unique. Moreover it belongs to $R^0(\mathbf{p}, \sigma)$.*

COROLLARY. *The function F^* is continuous over the set of positive weakly irreducible problems.*

Before proving the theorem we need to investigate the structure of the set $R^0(\mathbf{p}, \sigma)$.

The structure of $R^0(\mathbf{p}, \sigma)$. First we characterize the problems with nonempty $R^0(\mathbf{p}, \sigma)$.

LEMMA 3. *The set $R^0(\mathbf{p}, \sigma)$ is nonempty if and only if*

(3)

$$\begin{cases} c_J^+ \geqslant r_I^-, \quad r_I^+ \geqslant c_J^- \quad \text{and} \quad c_J^- + r_I^- \leqslant h \leqslant c_J^+ + r_I^+ \quad \text{for any } I \subset M, J \subset N \text{ with} \\ p_{I\bar{J}} = 0 \ (r_\varnothing^- = 0, \text{ etc.}). \end{cases}$$

PROOF. Consider the bipartite network with nodes M, N, where arc (i, j) exists if and only if $p_{ij} > 0$. Add a source s and arcs (s, i) with lower and upper capacities equal respectively to r_i^- and r_i^+. Similarly, add a sink t and arcs (j, t) with lower and upper capacities respectively equal to c_j^- and c_j^+. Finally, add arc (t, s) with lower and upper capacity equal to h. A feasible circulation exists in this network if and only if $R^0(\mathbf{p}, \sigma)$ is nonempty. But conditions (3) of Lemma 3 are exactly the necessary and

sufficient conditions for such a circulation to exist (see Hoffman's theorem in [13, p. 51]). ∎

Suppose now that the conditions (3) are met but one of the inequalities involving proper subsets I and J is satisfied as an equation. Any matrix $\mathbf{f}$ in $R^0(\mathbf{p}, \sigma)$ can be obtained by solving two independent subproblems. Indeed if for example $r_I^- = c_J^+$ with $p_{I\bar{J}} = 0$ any such $\mathbf{f}$ satisfies $f_{I\bar{J}} = 0$ by definition of $R^0(\mathbf{p}, \sigma)$ so that $r_I^- \leqslant f_{IN} = f_{IJ} \leqslant c_J^+ = r_I^-$ and thus surely $f_{\bar{I}J} = 0$. The matrix $\mathbf{f}$ is of the following form:

	J	$\bar{J}$
I	$\mathbf{f}'$	$\mathbf{0}$
$\bar{I}$	$\mathbf{0}$	$\mathbf{f}''$

Moreover $\mathbf{f}'$ is in $R^0(\mathbf{p}_{I \times J}, \sigma')$ where σ' agrees with σ for the constraints on the sums of the rows in I and columns in J and where the total sum h' is equal to c_J^+. Analogously $\mathbf{f}''$ is in $R^0(\mathbf{p}_{\bar{I} \times \bar{J}}, \sigma'')$ where $h'' = h - c_J^+$. Conversely take any $\mathbf{f}'$ in $R^0(\mathbf{p}_{I \times J}, \sigma'')$ (such $\mathbf{f}'$ necessarily exist), any f'' in $R^0(\mathbf{p}_{\bar{I} \times \bar{J}}, \sigma'')$, and form the matrix $\mathbf{f}$ equal to $\mathbf{f}'$ on $I \times J$, to $\mathbf{f}''$ on $\bar{I} \times \bar{J}$ and to zero elsewhere; then $\mathbf{f}$ is in $R^0(\mathbf{p}, \sigma)$. All other cases of a binding inequality in (3) may be treated in the same way. This leads to the following definitions:

A problem $(\mathbf{p}, \sigma)$ with $R^0(\mathbf{p}, \sigma) \neq \emptyset$ is said to be *reducible* at (I, J) into independent subproblems on (I, J) and $(\bar{I}, \bar{J})$, respectively, if I and J are proper subsets of M and N for which an inequality in (3) is binding.

A problem is *irreducible* if it is not reducible.

The key lemma is the following:

LEMMA 4. *Let $\mathbf{f}$ be an extended fair share of a problem $(\mathbf{p}, \sigma)$ with $R^0(\mathbf{p}, \sigma) \neq \emptyset$. Then either $\mathbf{f}$ is the fair share of the problem or (not exclusive) the problem is reducible at some (I, J), and $\mathbf{f}_{I \times J}$ and $\mathbf{f}_{\bar{I} \times \bar{J}}$ are extended fair shares of the corresponding two subproblems.*

PROOF. If $\mathbf{f}$ is an extended fair share there is a sequence $(\mathbf{p}^s, \sigma^s)$ converging to $(\mathbf{p}, \sigma)$ with $\mathbf{p}^s > \mathbf{0}$ such that the fair shares $\mathbf{f}^s$ converge to $\mathbf{f}$. Choose the multipliers $\mathbf{x}^s$ so that $\sum_i x_i^s = 1$. Since $\mathbf{x}^s$ is in a simplex, one may assume the $\mathbf{x}^s$ converge to some $\mathbf{x}$.

Assume that the sequence $\mathbf{y}^s$ is bounded. Then if $\mathbf{y}$ is one of its accumulation points surely $\mathbf{f} = \mathbf{x}\mathbf{p}\mathbf{y}$. Moreover, $x_i > x_k$ implies $x_i^t > x_k^t$ for infinitely many t, implying $f_{iN}^t = r_i^{-t}$ or $f_{kN}^t = r_k^{+t}$ so $f_{iN} = r_i^-$ or $f_{kN} = r_k^+$, and analogously for the columns. Therefore $\mathbf{f}$ is the fair share of $(\mathbf{p}, \sigma)$: this is the first case of Lemma 4.

Suppose now that the sequence $\mathbf{y}^s$ is not bounded. Let $J = \{ j : y_j^t \text{ bounded} \}$, and $\bar{J}$ the complement; let $I = \{ i : x_i > 0 \}$ and $\bar{I} = \{ i : x_i = 0 \}$ the complement. By assumption $\bar{J} \neq \emptyset$; also $I \neq \emptyset$ since $\sum_i x_i = 1$; an immediate implication is that $p_{ij} = 0$ for $(i, j) \in I \times \bar{J}$. For if $(i, j) \in I \times \bar{J}$, x_i^t converges to $x_i > 0$ and y_j^t is unbounded, so $f_{ij}^t = x_i^t p_{ij}^t y_j^t$ converges to a finite value only if $p_{ij} = 0$. Hence, I and $\bar{J}$ (which are not empty) must be proper subsets of M and N else $\mathbf{p}$ would contain a column or a row of zeros. Thus $J \neq \emptyset$; suppose $f_{Ml} > c_l^-$ for some $l \in \bar{J}$; then $j \in J$ implies $y_j^t < y_l^t$ for infinitely many t, so $f_{Mj} = c_j^+$. This shows that

either (a) $f_{Mj} = c_j^+$ for all $j \in J$,

or (b) $f_{Ml} = c_l^-$ for all $l \in \bar{J}$.

Since I and $\bar{I}$ are not empty, an argument parallel to the one above shows that

either (c) $f_{iN} = r_i^-$ for all $i \in I$,

or (d) $f_{kN} = r_k^+$ for all $k \in \bar{I}$.

Consider the four cases in turn.

	J	$\bar{J} \neq \varnothing$	
$\varnothing \neq I$		$p_{ij} = 0$	$x_i > 0$
$\bar{I}$	$f_{ij} = 0$		$x_i = 0$
	y_j bounded	y_j unbounded	

DIAGRAM 2.   The Situation in the Limit.

If (a) and (c) hold, then $c_J^+ = f_{MJ} = f_{IJ} \leqslant f_{IN} \leqslant r_I^-$. But since $R^0(\mathbf{p}, \boldsymbol{\sigma})$ is nonempty and $p_{I\bar{J}} = 0$, $c_J^+ \geqslant r_I^-$ by Lemma 3. Therefore all the above inequalities are equations implying $r_I^- = c_J^+$ and $f_{I\bar{J}} = 0$. By a similar argument one shows that if (a) and (d) hold then $h = c_J^+ + r_I^+$, and if (b) and (c) hold then $h = c_{\bar{J}}^- + r_I^-$, and if (b) and (d) hold then $c_{\bar{J}}^- = r_I^+$. Moreover in all cases $f_{I\bar{J}} = 0$. Thus the problem is reducible and since $f_{I\bar{J}} = f_{\bar{I}J} = 0$ the submatrices $\mathbf{f}_{I \times J}$ and $\mathbf{f}_{\bar{I} \times \bar{J}}$ are extended fair shares of the corresponding subproblems. Indeed, by uniformity $\mathbf{f}_{I \times J}^s$ is the fair share matrix of the problem $(\mathbf{p}_{I \times J}^s, \boldsymbol{\sigma}_{I \times J}^s)$ where for example the lower bound on row i is equal to $r_i^{-s} - f_{iJ}^s$ and on column j to $c_j^{-s} - f_{\bar{I}j}^s$. But these two expressions tend respectively to r_i^- and c_j^- since f_{iJ}^s and $f_{\bar{I}j}^s$ tend to zero for i in I and j in J. The same argument applies for $\mathbf{f}_{\bar{I} \times \bar{J}}$: this ends the proof. ∎

PROOF OF THEOREM 3.   The proof is by induction on $n + m$. If $n + m = 1$, it is obvious. Take now an $m \times n$ problem. Lemma 4 implies that an extended fair share is in $R^0(\mathbf{p}, \boldsymbol{\sigma})$. Indeed if it is a fair share it is obvious; if not the problem is reducible at (I, J) and $f_{I\bar{J}} = f_{\bar{I}J} = 0$, $f_{I \times J}$ and $\mathbf{f}_{\bar{I} \times \bar{J}}$ are extended fair shares of the corresponding subproblems so they are by the induction hypothesis in the corresponding sets R^0 and thus $\mathbf{f}$ is in $R^0(\mathbf{p}, \boldsymbol{\sigma})$. Suppose now that $\mathbf{f}$ and $\mathbf{g}$ are distinct extended fair shares. At least one of them, say $\mathbf{f}$, is not the fair share so that the problem is reducible at some (I, J). But since $\mathbf{g}$ is in $R^0(\mathbf{p}, \boldsymbol{\sigma})$ surely $g_{I\bar{J}} = g_{\bar{I}J} = 0$, thus $\mathbf{g}_{I \times J}$ and $\mathbf{g}_{\bar{I} \times \bar{J}}$ are extended fair shares of the corresponding subproblems. By the induction hypothesis (recall that the sets R^0 of these problems are not empty) $\mathbf{g}_{I \times J} = \mathbf{f}_{I \times J}$ and $\mathbf{g}_{\bar{I} \times \bar{J}} = \mathbf{f}_{\bar{I} \times \bar{J}}$ so that $\mathbf{f} = \mathbf{g}$. ∎

I.5.   *Existence of a fair share..*   In this section we characterize the problems which admit a fair share. This generalizes earlier results for the equality problem (e.g. [4], [9]). Obviously a fair share belongs to the set $R^0(\mathbf{p}, \boldsymbol{\sigma})$ thus a necessary condition is the nonemptiness of this set. It is not sufficient (see the first example in I.4). However an immediate implication of Lemma 4 is that a fair share exists for any irreducible problem. So let us investigate the reducible problems. Consider the family of equality constrained problems:

$$\mathbf{p}(t) = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ t & t & 1 \end{bmatrix}, \quad \mathbf{r} = (9, 6, 5), \quad \mathbf{c} = (7, 8, 5).$$

Any matrix in $R^0(\mathbf{p}(t), \boldsymbol{\sigma})$ is of the following form:

$$\begin{bmatrix} 7 - \alpha & 2 + \alpha & 0 \\ \alpha & 6 - \alpha & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

The problems are all reducible at $(\{1,2\}, \{1,2\})$ forcing f_{31} and f_{32} to zero. Since a fair share has exactly the same zeros as $\mathbf{p}$, one exists here only for $t = 0$. This leads to the following definition:

A problem is *decomposable* at (I, J) if it is reducible at (I, J) and $p_{\bar{I}J} = 0$.

If a problem is reducible then each of its subproblems is either irreducible or may itself be reduced. In the latter case reduce each reducible subproblem and iterate until a set of irreducible independent subproblems is obtained. If each reduction is in fact a decomposition then the problem is said to be *decomposable* into independent irreducible subproblems. A decomposable problem admits a partition of rows $M = I_1 \cup I_2 \cup \cdots \cup I_k$ and $N = J_1 \cup J_2 \cup \cdots \cup J_k$ with $P_{I_i J_j} = 0$ if $i \neq j$.

THEOREM 4. *The problem* $(\mathbf{p}, \sigma)$ *has a fair share matrix if and only if it is decomposable into independent irreducible subproblems.*

PROOF. *Sufficiency.* Suppose that $(\mathbf{p}, \sigma)$ is decomposable into independent irreducible subproblems. If the problem itself is irreducible the result is in hand by Lemma 4, so assume inductively the existence of a fair share matrix if the problem is decomposable into not more than d independent irreducible subproblems, and consider a problem with $d + 1$ such subproblems. Let the subproblem on $I \times J$ be one of the irreducible subproblems and $\mathbf{f}_{I \times J}$ be its fair share matrix, and let $\mathbf{f}_{\bar{I} \times \bar{J}}$ be the fair share matrix of the subproblem on $\bar{I} \times \bar{J}$. The multipliers within each subproblem satisfy the conditions (2) and the problem is weakly irreducible so it only needs to be shown that the conditions (2) hold for multipliers corresponding to different subproblems.

If $c_J^+ = r_I^-$ the sum over every row of $\mathbf{f}_{I \times J}$ equals its lower bound and over every column its upper bound. So one can scale the multipliers of rows in I up and the multipliers of columns in J down so that each multiplier of a row in I is larger than all those of $\mathbf{f}_{\bar{I} \times \bar{J}}$ and each multiplier of a column in J is smaller than all those of $\mathbf{f}_{\bar{I} \times \bar{J}}$: conditions (2) are then met. If $r_I^+ = c_J^-$, a similar scaling of the multipliers of $\mathbf{f}_{\bar{I} \times \bar{J}}$ does the trick. If $c_J^- + r_I^- = h$ then scaling the row multipliers of $\mathbf{f}_{I \times J}$ up and the column multipliers of $\mathbf{f}_{\bar{I} \times \bar{J}}$ up assures conditions (2) as well and similarly if $h = c_J^+ + r_I^+$.

Necessity. If the fair share matrix $\mathbf{f}$ exists for a problem $(\mathbf{p}, \sigma)$ then $f_{ij} = 0$ if and only if $p_{ij} = 0$, implying $\mathbf{f} \in R^0(\mathbf{p}, \sigma)$, so the inequalities (3) must hold. If they are all satisfied strictly then $(\mathbf{p}, \sigma)$ is irreducible. Otherwise, equality holds for some (I, J). But, then, $f_{\bar{I}J} = f_{I\bar{J}} = 0$ implying $p_{\bar{I}J} = p_{I\bar{J}} = 0$, so that $(\mathbf{p}, \sigma)$ is decomposable into two independent subproblems. The same argument may be repeated for each subproblem until all subproblems are irreducible. ∎

The following subset is the set of allocations with the same zeros as $\mathbf{p}$:

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

COROLLARY. *The problem* $(\mathbf{p}, \sigma)$ *has a fair share matrix if and only if* $R^+(\mathbf{p}, \sigma)$ *is nonempty.*

Clearly, if a fair share exists, $R^+(\mathbf{p}, \sigma)$ is nonempty. Conversely if $R^+(\mathbf{p}, \sigma) \neq \varnothing$ then $R^0(\mathbf{p}, \sigma) \neq \varnothing$ and the conditions (3) are satisfied. If they are all satisfied strictly then the problem is irreducible and a fair share exists by Theorem 5. Otherwise, the problem is reducible at some (I, J). But then, if $\mathbf{f} \in R^+(\mathbf{p}, \sigma)$, $f_{\bar{I}J} = 0$ implying $p_{\bar{I}J} = 0$, so the problem is decomposable into two independent subproblems whose corresponding sets R^+ are nonempty. By an induction argument these subproblems are decomposable into independent irreducible subproblems. Thus the problem itself is decomposable and admits a fair share by Theorem 4. ∎

II. Proportionality in integers.

II.1. *Definitions and axioms.* As before, a *problem* is a pair $(\mathbf{p}, \sigma)$, but we are now interested in *apportionments* in $R(\sigma)$, that is, allocations all of whose components are integer valued. The central question of this part of the paper is: what does it mean to say that an apportionment $\mathbf{a} \in R(\sigma)$ is proportional to $\mathbf{p}$ and when do such apportionments exist?

It is assumed that, as before, $\mathbf{p}$ contains no row or column of zeros, and also that all of the components of σ are integer valued. The set $R^0(\mathbf{p}, \sigma)$ (defined (in I.4) plays a central role so the reader is reminded that $R^0(\mathbf{p}, \sigma) \neq \varnothing$ for σ integer valued if and only if it contains integer valued solutions (by a classical result of network flows [13]).

A *method of apportionment* A is a rule or mapping that assigns at least one apportionment to every problem: $A(\mathbf{p}, \sigma)$ is a nonempty integer valued subset of $R(\sigma)$. The possibility of multiple apportionments is once again not excluded. The axioms that follow are directly inspired by those used for a method of allocation, and are discussed below.

AXIOM 1' (Exactness).   *If* $\mathbf{f} = F^*(\mathbf{p}, \sigma)$ *is integer in all components, then* $A(\mathbf{p}, \sigma) = \mathbf{f}$.

AXIOM 2' (Relevance).   *If* $A(\mathbf{p}, \sigma) \cap R(\hat{\sigma}) \neq \varnothing$ *and* $R(\hat{\sigma}) \subset R(\sigma)$ *then* $A(\mathbf{p}, \hat{\sigma}) = A(\mathbf{p}, \sigma) \cap R(\hat{\sigma})$.

AXIOM 3' (Uniformity).   *If* $\mathbf{a} \in A(\mathbf{p}, \sigma)$ *then the same statements hold as do in Axiom 3 with A replacing F and $\mathbf{a}$ replacing $\mathbf{f}$.*

AXIOM 4' (Monotonicity).   *If* $\mathbf{a} \in A(\mathbf{p}, \sigma)$, $\mathbf{a}' \in A(\mathbf{p}', \sigma)$ *and $\mathbf{p}'$ is equal to $\mathbf{p}$ except that $p_{kl} < p'_{kl}$ then $a_{kl} \leqslant a'_{kl}$.*

Given $\mathbf{a}$ in $R(\sigma)$, define $I^- = \{i \in M : a_{iN} = r_i^-\}$ and $I^+ = \{i \in M : a_{iN} = r_i^+\}$ and define J^- and J^+ similarly.

AXIOM 5' (Homogeneity).   *If* $\mathbf{a} \in A(\mathbf{p}, \sigma)$ *and* $\delta > 0$, $\alpha > 0$ *and* $\beta > 0$ *are such that*

$\alpha_i > 1$ *implies* $i \in I^+$, *and* $\alpha_i < 1$ *implies* $i \in I^-$,
$\beta_j > 1$ *implies* $j \in J^+$, *and* $\beta_j < 1$ *implies* $j \in J^-$,
then $\mathbf{a} \in A(\delta \alpha \mathbf{p} \beta, \sigma)$.

AXIOM 6' (Completeness).   *If* $\mathbf{p}^s \to \mathbf{p}$ *and* $\mathbf{a} \in A(\mathbf{p}^s, \sigma)$ *for every s, then* $\mathbf{a} \in A(\mathbf{p}, \sigma)$.

An apportionment is simply an allocation with the additional requirement that each component be integer valued, so if the (unique) acceptable allocation is in integers then it is the unique solution. Axiom 2' is a slightly strengthened form of Axiom 2 which is justified identically. Axioms 3' and 4' are direct transcriptions of the real counterparts. Homogeneity as defined in Axiom 5' is a property that is true for fair share allocations satisfying Axioms 1 through 5: one can therefore not expect anything less to hold for apportionments. Completeness is a continuity axiom necessary in the case of apportionment problems that handles a difficulty that does not arise in the case of allocation problems.

II.2. *Divisor methods of apportionments.* A *divisor function* d is a strictly monotone real function defined over the nonnegative integers satisfying $a \leqslant d(a) \leqslant a + 1$, and $d(a) = a$ and $d(b) = b + 1$ for no integers $a \geqslant 1$ and $b \geqslant 0$. A *d-rounding* of $x = 0$ is 0 and of $x > 0$ is defined by

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

$$= a \text{ or } a + 1 \quad \text{if } x = d(a),$$

so a d-rounding is unique unless $x = d(a)$, in which case it is a or $a + 1$. Accordingly we will sometimes write $a \in [x]_d$. In effect, $d(a) \in [a, a + 1]$ is the threshold below which x is rounded down, above which it is rounded up and at which x is rounded either up or down. The additional caveat on the definition of a divisor function excludes the possibility that in one open interval $(a, a + 1)$ with $a \geqslant 0$ integer every x is rounded up whereas in another interval every x is rounded down.

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

$$(4) \qquad\qquad \mathbf{a} = (a_{ij}) = \left(\left[\delta \lambda_i p_{ij} \mu_j \right]_d \right), \quad \mathbf{a} \in R(\boldsymbol{\sigma})$$

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

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

$\mu_j > 1$ implies $a_{Mj} = c_j$ and $\mu_j < 1$ implies $a_{Mj} = c_j^+$.

The multipliers $(\delta, \boldsymbol{\lambda}, \boldsymbol{\mu})$ satisfying (4) will be said to be *proper* for **a** in $A^d(\mathbf{p}, \boldsymbol{\sigma})$.

The set of divisor method apportionments will take on the role of the fair share allocations: the integer requirement introduces some indecision on the meaning of "proportionality" that cannot be avoided.

Divisor methods first arose in the study of the vector apportionment problem [6] and have a particularly simple realization in that case. To obtain all divisor method apportionments based on d for the vector problem with $m = 1$: determine $\delta > 0$ so that $\sum_j \text{mid}(c_j^-, [\delta p_j]_d, c_j^+) = h$ and let $a_j = \text{mid}(c_j^-, [\delta p_j]_d, c_j^+)$. Several particular methods are used for the apportionment of seats in legislatures to regions or to political parties in one or another nation, notably: Adams's $d(a) = a$; Webster's, $d(a) = a + 1/2$; Jefferson's, $d(a) = a + 1$; and Hill's, $d(a) = \sqrt{a(a + 1)}$.

LEMMA 5. *Suppose* $\mathbf{a} \in A^d(\mathbf{p}, \boldsymbol{\sigma})$ *with proper multipliers* $(\delta, \boldsymbol{\lambda}, \boldsymbol{\mu})$ *and* $\mathbf{a}' \in A^d(\mathbf{p}', \boldsymbol{\sigma})$. *If* $p_{kl} \leqslant p'_{kl}$ *for some* (k, l), $p_{ij} = p'_{ij}$ *for* $(i, j) \neq (kl)$ *and* $a_{kl} > a'_{kl}$ *then* $p_{kl} = p'_{kl}$ *and* $(\delta, \boldsymbol{\lambda}, \boldsymbol{\mu})$ *is proper for* $\mathbf{a}'$ *in* $A^d(\mathbf{p}, \boldsymbol{\sigma})$.

PROOF. The argument parallels that given for Lemma 2, but with several accommodations for the fact that d-roundings are not unique.

A divisor method is said to be *unique up to ties* if the same multipliers $(\delta, \boldsymbol{\lambda}, \boldsymbol{\mu})$ are proper for all $\mathbf{a} \in A^d(\mathbf{p}, \boldsymbol{\sigma})$. Lemma 5 shows that divisor methods are unique up to ties and that this means $a_{ij} \in [\delta \lambda_i p_{ij} \mu_j]_d$ for *any* $\mathbf{a} \in A^d(\mathbf{p}, \boldsymbol{\sigma})$.

In the case of apportionments it is perfectly acceptable to have $a_{ij} = 0$ when $p_{ij} > 0$. Nevertheless, a divisor method A^d based on a divisor function satisfying $d(0) = 0$ implies that $a_{ij} \geqslant 1$ if $p_{ij} > 0$. Accordingly, the subset of $R^0(\mathbf{p}, \boldsymbol{\sigma})$ defined by

$$R^1(\mathbf{p}, \boldsymbol{\sigma}) = \left\{ \mathbf{f} \in R^0(\mathbf{p}, \boldsymbol{\sigma}): p_{ij} > 0 \text{ implies } f_{ij} \geqslant 1 \right\}$$

will have to be considered.

THEOREM 5. *Let d be a divisor function. If $d(0) > 0$, $A^d(\mathbf{p}, \boldsymbol{\sigma})$ is nonempty if and only if $R^0(\mathbf{p}, \boldsymbol{\sigma})$ is nonempty. If $d(0) = 0$, $A^d(\mathbf{p}, \boldsymbol{\sigma})$ is nonempty if and only if $R^1(\mathbf{p}, \boldsymbol{\sigma})$ is nonempty. When A^d-apportionments exist, they are unique up to ties.*

Existence is proven constructively by an algorithm (given in [5]) that either provides an A^d-apportionment or shows that $R^0(\mathbf{p}, \boldsymbol{\sigma})$ or $R^1(\mathbf{p}, \boldsymbol{\sigma})$, depending upon the d in hand, is empty. The algorithm terminates in a finite number of steps, each step consisting in changes in multipliers δ, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$ done very much in the spirit of the out-of-kilter method for network flows [13]. Uniqueness up to ties is an immediate consequence of Lemma 5.

Let $e_{IJ} = \{$the number of $p_{ij} > 0: i \in I, j \in J\}$.

LEMMA 6. *The set $R^1(\mathbf{p}, \sigma)$ is nonempty if and only if $c_J^+ \geqslant r_I^- + e_{\bar{I}J}$, $r_I^+ \geqslant c_J^- + e_{\bar{I}J}$ 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 $p_{I\bar{J}} = 0$.*

PROOF. Consider the same network as in Lemma 3 but impose a lower capacity of 1 on any existing arc (i, j). A feasible flow exists in this network if and only if $R^1(\mathbf{p}, \sigma) \neq \varnothing$. ∎

LEMMA 7. *If $R(\sigma) = R(\sigma')$ then $A^d(\mathbf{p}, \sigma) = A^d(\mathbf{p}, \sigma')$.*

PROOF. Define $\hat{\sigma}$ by $\hat{r}_i^- = \min\{a_{iN}: \mathbf{a} \in R(\sigma)\}$, $\hat{r}_i^+ = \max\{a_{iN}: \mathbf{a} \in R(\sigma)\}$, and $\hat{c}^-$ and $\hat{c}^+$ similarly. Then $R(\hat{\sigma}) = R(\sigma)$, and the vector $\hat{\sigma}$ depends only on $R(\sigma)$.

Now suppose $\mathbf{a} \in A^d(\mathbf{p}, \sigma)$, with proper multipliers (δ, λ, μ). Then $\lambda_i > 1$ implies $a_{iN} = r_i^- = \hat{r}_i^-$, and similarly for the other multipliers, so (δ, σ, μ) is proper for $\mathbf{a} \in A^d(\mathbf{p}, \hat{\sigma})$. If $\mathbf{b} \in A^d(\mathbf{p}, \hat{\sigma})$ then, by uniqueness up to ties, (δ, λ, μ) is proper for $\mathbf{b}$ in $A^d(\mathbf{p}, \hat{\sigma})$, so $\lambda_i > 1$ implies $b_{iN} = \hat{r}_i^- = r_i^-$, etc., and thus is proper for $\mathbf{b} \in A^d(\mathbf{p}, \sigma)$. ∎

II.3. *Characterization.* A problem $(\mathbf{p}, \sigma)$ is *strongly positive* if $\mathbf{p} > \mathbf{0}$ and $R^1(\mathbf{p}, \sigma) \neq \varnothing$.

THEOREM 6. *A method of apportionment satisfies the 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 $d(0) > 0$.*

PROOF. *Sufficiency.* We begin by checking that any divisor method A^d satisfies the axioms.

By Theorem 5, A^d is defined over the set of strongly positive problems for any d, and over the positive problems if $d(0) > 0$. To verify exactness, suppose $\mathbf{f} = F^*(\mathbf{p}, \sigma)$ is all integer. Since $f_{ij} \in [f_{ij}]_d$ then $\mathbf{f} \in A^d(\mathbf{p}, \sigma)$ and only uniqueness must be proven. So, suppose $\mathbf{f} \neq \mathbf{a} \in A^d(\mathbf{p}, \sigma)$. Then, by uniqueness up to ties, there must exist a pair (i, j) and (k, l) with $a_{ij} = f_{ij} - 1$, $a_{ij} \in [f_{ij}]_d$ and also $a_{kl} = f_{kl} + 1$, $a_{kl} \in [f_{kl}]_d$, contradicting the definition of a divisor function.

Before turning to relevancy, remark that if $R(\hat{\sigma}) \subset R(\sigma)$ we may assume $\hat{\mathbf{r}}^- \geqslant \mathbf{r}^-$, $\mathbf{r}^+ \geqslant \hat{\mathbf{r}}^+$ and the same for the $\mathbf{c}$. Indeed, if, for example $r_i^- > \hat{r}_i^-$, the region $R(\hat{\sigma})$ does not change when $\hat{r}_i^-$ is replaced by r_i^- and by Lemma 7 the A^d-apportionments are not changed either.

We can now verify relevancy. Suppose $\mathbf{a} \in A^d(\mathbf{p}, \sigma) \cap R(\hat{\sigma})$ for $R(\hat{\sigma}) \subset R(\sigma)$. Let (δ, λ, μ) be proper multipliers for $\mathbf{a}$ in $A^d(\mathbf{p}, \sigma)$. Then $\lambda_i > 1$ implies $a_{iN} = r_i^-$, whereas $\mathbf{a} \in R(\hat{\sigma})$ and $R(\hat{\sigma}) \subset R(\sigma)$ imply $r_i^- = \hat{r}_i^-$, so $a_{iN} = \hat{r}_i^-$. The remaining conditions (4) are checked similarly, showing (δ, λ, μ) is proper for $\mathbf{a} \in A^d(\mathbf{p}, \hat{\sigma})$. To prove the reverse inclusion, it must be shown that if $\mathbf{a} \in A^d(\mathbf{p}, \sigma) \cap R(\hat{\sigma})$ then any $\mathbf{b} \in A^d(\mathbf{p}, \hat{\sigma})$ belongs also to $A^d(\mathbf{p}, \sigma)$. We have just seen that if (δ, λ, μ) is proper for $\mathbf{a}$ in $A^d(\mathbf{p}, \sigma)$ it is also proper for $\mathbf{a} \in A^d(\mathbf{p}, \hat{\sigma})$. By uniqueness up to ties, (δ, λ, μ) is proper for $\mathbf{b} \in A^d(\mathbf{p}, \hat{\sigma})$ also, so $\lambda_i > 1$ implies $b_{iN} = \hat{r}_i^-$ and $a_{iN} = \hat{r}_i^- = r_i^-$ thus $b_{iN} = r_i^-$, and similarly for the remaining conditions (4), so $\mathbf{b} \in A^d(\mathbf{p}, \sigma)$.

To see that uniformity holds, note first that the restriction of $\mathbf{a}$ in $A^d(\mathbf{p}, \sigma)$ is clearly a solution of the corresponding subproblem on $I \times J$. Conversely, if $\mathbf{b}_{I \times J}$ is another apportionment of the subproblem, then uniqueness up to ties shows that if the sum over a row or column of the subproblem apportionments differ then the corresponding multiplier must be 1, so substitution indeed produces an alternate apportionment for the whole.

Monotonicity follows directly from Lemma 5, and homogeneity is immediate.

Finally, we verify completeness in the case $d(0) > 0$ (an analogous proof works for the case $d(0) = 0$). Suppose $\mathbf{p}^s$ converges to $\mathbf{p} > \mathbf{0}$ with $\mathbf{a} \in A^d(\mathbf{p}^s, \sigma)$. All we need to show is that proper multipliers $\delta^s, \lambda^s, \mu^s$ for the problems $(\mathbf{p}^s, \sigma)$ can be chosen in a

compact set since then any accumulation point will be proper for $\mathbf{a}$ in $A^d(\mathbf{p}, \sigma)$. Since $\mathbf{p} > \mathbf{0}$ we may assume there is a $W > 0$ satisfying $1 \leqslant p_{ij}^s \leqslant W$ and $d(a_{ij}) \leqslant W$ for all (i, j). Consider any problem $(\mathbf{q}, \sigma)$ having $\mathbf{a} \in A^d(\mathbf{q}, \sigma)$ where $1 \leqslant q_{ij} \leqslant W$, all (i, j). If (δ, λ, μ) is proper we may assume $\max \lambda_i \geqslant 1$ and $\max \mu_j \geqslant 1$: otherwise $\lambda_i < 1$ all i, implying $a_{iN} = r_i^+$ and dividing all λ_i by $\max \lambda_i$ and multiplying δ by the same quantity gives a proper set of multipliers.

From $\delta \lambda_i q_{ij} \mu_j \leqslant W$ for all (i, j) and $q_{ij} \geqslant 1$, we deduce $\delta \leqslant W$ and δ is in a bounded set.

Suppose that the following condition held: for some (i, j), $\lambda_i \leqslant 1$, $\delta \lambda_i q_{ij} \mu_j \geqslant d(0)$. Then $\delta \mu_j \geqslant d(0)/W$ and for every row k in M the inequality $\delta \lambda_k q_{kj} \mu_j \leqslant d(a_{kj}) \leqslant W$ implies $\lambda_k \leqslant W^2/d(0)$, so the λ may be taken in a bounded set.

If the condition does not hold, define $K = \{ k \in M: \lambda_k \leqslant 1 \}$. If $k \in K$, $\delta \lambda_k q_{kj} \mu_j < d(0)$ for every j, so $a_{kN} = 0$. Choose $\alpha > 1$ as large as possible satisfying $\alpha \delta \lambda_k q_{kj} \mu_j \leqslant d(0)$ for $k \in K$ and $j \in N$, and $\lambda_k/\alpha \geqslant 1$ for $k \notin K$. δ may be replaced by $\alpha \delta$ and λ_k by λ_k/α for $k \notin K$. If the condition still does not hold then the set K is larger and the construction repeated. At some point the condition must obtain, for otherwise $K = M$ and $\mathbf{a} = \mathbf{0}$, contradicting $a_{MN} = h > 0$. The same argument applies to columns, and ends the proof.

Necessity (The free case). It is first shown that any method of apportionment for a *free problem* (no constraints except $a_{MN} = h$) satisfying the axioms must be a divisor method. This part of the proof is similar to the main characterization result for the vector case [6], but is somewhat simpler and uses an altered set of axioms.

A free problem may simply be referred to as $(\mathbf{p}, h)$. Consider a method A that satisfies the axioms. Let

$$\mathbb{P}_a = \{ p > 0: (x, y) \in A((p, 1), a + 2) \text{ for some } x \geqslant a + 1 \text{ and } y \leqslant 1 \}.$$

By exactness $a + 1 \in \mathbb{P}_a$ for any integer $a \geqslant 0$, so $\mathbb{P}_a \neq \varnothing$ and $d(a) = \inf \mathbb{P}_a$ is well defined. It is shown that A is the divisor method A^d. The key fact is

(5) $p/q > d(a)/d(b)$, $(x, y) \in A((p, q), h)$ implies $x \geqslant a + 1$ or $y \leqslant b$.

For suppose (5) is true, that d is a divisor function and that $\mathbf{a} \in A^d(\mathbf{p}, h)$ and $\mathbf{x} \in A(\mathbf{p}, h)$. Then $[p_{ij}\delta]_d = a_{ij}$ all i, j so

(6) $p_{ij}/d(a_{ij} - 1) \geqslant p_{kl}/d(a_{kl})$ for any i, j and k, l and $a_{ij} \geqslant 1$.

Suppose that all of these inequalities are satisfied strictly but $\mathbf{x} \neq \mathbf{a}$. Then, for some (i, j), $a_{ij} > x_{ij} \geqslant 0$, so uniformity and (5) and (6) imply $x_{il} < a_{il}$ for any l in N and $x_{kj} < a_{kj}$ for any k in M; repeating the argument from x_{il} or x_{kj} yields $x_{kl} < a_{kl}$ for any (k, l) and $x_{MN} < h$, a contradiction; therefore for any problem $(\mathbf{p}, h)$ where the inequalities (6) are strict, the methods A and A^d coincide. Completeness shows they coincide on all problems. Notice, however, that if A is defined over all positive problems, necessarily $d(0) > 0$.

To prove (5) we need to establish a number of facts. Since uniformity holds and the problem is free it suffices to consider two-population problems. We first show that:

(7)   If $(b, c) \in A((p, p), h)$,   then $|b - c| \leqslant 1$.

To see this consider the two cases, h even and h odd.

If $h = 2a$, by exactness $(a, a) = A((p, p), 2a)$. If $h = 2a + 1$ first compare $(b, c) \in A((p, p), 2a + 1)$ with $(a + 1, a) = A(((a + 1)p/a, p), 2a + 1)$, then with $(a, a + 1)$

$= A((ap/(a + 1), p), 2a + 1)$. By monotonicity, $a + 1 \geqslant b$ and $b \geqslant a$, b is either a or $a + 1$ and the same for c.

Next,

(8) $\mathbf{a} \in A(\mathbf{p}, h)$ implies there exists $\mathbf{a}' \in A(\mathbf{p}, h + 1)$ satisfying $\mathbf{a}' \geqslant \mathbf{a}$.

Suppose the contrary, namely, that $(a, b) \in A((p, q), h)$, $(a', b') \in A((p, q), h + 1)$ with $a > a'$. Consider $(c, d, c', d') \in A((p, q, p, q), 2h + 1)$. By (7) it may be assumed that $c + d = h + 1$ and $c' + d' = h$, so by uniformity (a, b, a', b') is also a solution. But, then, $b < b' + 1$ contradicts (7).

(9)
$$\begin{aligned} &\text{Suppose } (x', y') \in A((p', q'), h'), \quad x' \geqslant \quad a + 1 \quad \text{and} \\ &y' \leqslant b. \text{ Then } p/q > p'/q' \text{ and } (x, y) \in A((p, q), h) \\ &\text{implies } x \geqslant a + 1 \text{ or } y \leqslant b \text{ for any } h. \end{aligned}$$

Let $(\bar{x}, \bar{y}) \in A((p, q), h')$. By monotonicity and homogeneity, $\bar{x} \geqslant x' \geqslant a + 1$ so $\bar{y} \leqslant b$. Consider, then, $(x, y) \in A((p, q), h)$. If $h > h'$, $x \geqslant \bar{x} \geqslant a + 1$ by (8), whereas if $h < h'$, $y \leqslant b$ for the same reason, proving (9).

By (9) and the definition of $\mathbb{P}_a$

(10) $p > d(a)$, $(x, y) \in A((p, 1), h)$ implies either $x \geqslant a + 1$ or $y \leqslant 1$,   and

(11) $q < d(b)$, $(x, y) \in A((q, 1), h)$ implies either $x \leqslant b$ or $y \geqslant 2$.

Now, to prove (5), suppose $p/q > d(a)/d(b)$, choose (p', q') so that $p/q > p'/q'$ and $p' > d(a)$, $q' < d(b)$, and let $(x', y', z') \in A((p', q', 1), a + b + 2)$. We claim $x' \geqslant a + 1$ *and* $y' < b$. There are two cases to consider.

Case 1: $z' \leqslant 1$. By uniformity, $(y', z') \in A((q', 1), y' + z')$ and (11) implies $y' \leqslant b$, so $y' + z' \leqslant b + 1$ and thus $x' > a + 1$.

Case 2: $z' \geqslant 2$. By uniformity $(x', y') \in A((p', 1), x' + y')$ and (10) implies $x' \geqslant a + 1$, so $x' + z' \geqslant a + 3$ and thus $y' < b$. Now x, y and $p/q > p'/q'$ satisfy the hypotheses of (9), and the result establishes (5).

It remains to verify that d is a divisor function. Remark first, that $d(1) = 1$: $(2, 1)$ and $(1, 2)$ both belong to $A((1, 1); 3)$, whereas by monotonicity $(x, y) \in A((p, 1), 3)$ for $p < 1$ must satisfy $x \leqslant 1$. Second, $d(\mathbf{a}) \geqslant 1$, if $a \geqslant 2$, because $1 \notin \mathbb{P}_a$ by (7). Now we prove $d(b)/(b + 1) < d(a)/a$ for all $a \geqslant 1$ and $b \geqslant 0$. For suppose the contrary, $d(b)/(b + 1) \geqslant d(a)/a$ for some a, b. $a \geqslant 1$ implies $d(a) > 0$ so $d(b) > 0$. Let (p'', q'') be a sequence converging to $(a, b + 1)$ with $p''/q'' > a/(b + 1)$ and $(x'', y'') \in A((p'', q''); a + b + 1)$. By (5) $x'' \geqslant a + 1$ or $y'' \leqslant b$, so $x'' \geqslant a + 1$ and $y'' \leqslant b$. A complete means (x'', y'') converges to $(x, y) \in A((a, b + 1), a + b + 1)$, but $x \geqslant a + 1$ and $y \leqslant b$ contradicts exactness.

Therefore, there exists $\lambda > 0$ such that $\sup_{b \geqslant 0} d(b)/(b + 1) \leqslant \lambda \leqslant \inf_{a \geqslant 1} d(a)/a$, and the function $d^* = d/\lambda$ satisfies $a \leqslant d^*(a) \leqslant a + 1$ and defines a divisor method equivalent to A (and to A^d). This completes the proof that any A satisfying the axioms must be a divisor method for some d over all free problems.

Using the same argument as for the real case shows that A and A^d coincide on all vector problems.

The general case. Suppose, now, that A satisfies the Axioms 1' through 6' over the set of positive problems and let d be the divisor function such that $A = A^d$ over the set of free problems. We show that $A(\mathbf{p}, \boldsymbol{\sigma}) = A^d(\mathbf{p}, \boldsymbol{\sigma})$.

Let δ, λ, μ be proper for $\mathbf{a} \in A^d(\mathbf{p}, \boldsymbol{\sigma})$. Then $\mathbf{a} \in A^d(\delta\lambda\mathbf{p}\mu, h) \cap R(\boldsymbol{\sigma})$ and, by what we know about free problems, $\mathbf{a} \in A(\delta\lambda\mathbf{p}\mu, h) \cap R(\boldsymbol{\sigma})$. Since A is relevant, we

conclude $\mathbf{a} \in A(\delta\lambda\mathbf{p}\mu, \sigma)$, with $\lambda_i > 1$ implying $a_{iN} = r_i^-$ and the remaining conditions (4). But A homogeneous implies $\mathbf{a} \in A(\mathbf{p}, \sigma)$, so $A^d \subset A$.

To prove $A \subset A^d$ is more involved. To begin, suppose $\mathbf{b} \in A(\mathbf{p}, \sigma)$ and define $\hat{\sigma}$ by $\hat{r}_i^- = \min\{a_{iN}, b_{iN}\} \geqslant r_i^-$, $\hat{r}_i^+ = \max\{a_{iN}, b_{iN}\} \leqslant r_i^+$ and similarly for $\hat{c}^-$ and $\hat{c}^+$. Notice that

$\lambda_i > 1$ implies $a_{iN} = r_i^-$ so $b_{iN} = \hat{r}_i^+$ and
$\lambda_i < 1$ implies $a_{iN} = r_i^+$ so $b_{iN} = \hat{r}_i^-$,

and similarly for the μ_j. By homogeneity, $\mathbf{b} \in A(\delta\lambda\mathbf{p}\mu, \hat{\sigma})$. But $\mathbf{a} \in A^d(\delta\lambda\mathbf{p}\mu, h) \cap R(\hat{\sigma}) \subset A(\delta\lambda\mathbf{p}\mu, h) \cap R(\hat{\sigma}) \neq \varnothing$ so by relevance of A, $A(\delta\lambda\mathbf{p}\mu, \hat{\sigma}) = A^d(\delta\lambda\mathbf{p}\mu, h) \cap R(\hat{\sigma})$. Thus $\mathbf{b}$ is also a d-rounding of $\delta\lambda\mathbf{p}\mu$. It remains to show that conditions (4) are satisfied for $\mathbf{b}$. Consider the "replicated" $m \times 2n$ problem $((\mathbf{p}, \mathbf{p}\mu), (2\mathbf{r}^-, 2\mathbf{r}^+, \mathbf{c}^-, \mathbf{c}^-, \mathbf{c}^+, \mathbf{c}^+, 2h))$. The apportionment $(\mathbf{a}, \mathbf{a})$ is an A^d-solution so also an A-solution of this problem. By uniformity of A, $(\mathbf{b}, \mathbf{a})$ is also an A-apportionment. But if $\mu_j > 1$ and $b_{Mj} > a_{Mj} = c_j^-$, there is i in M with $b_{ij} > a_{ij}$ and (b_{ij}, a_{ij}) is not an A^d-solution of the corresponding two-vector problem $(p_{ij}, p_{ij}\mu_j)$: this ends the proof. ∎

II.4. *Applications.* The set of proportional solutions in reals, $F(\mathbf{p}, \sigma)$, is the unique fair share matrix, whereas the set of proportional solutions in integers admits any divisor method solution $A^d(\mathbf{p}, \sigma)$. So there are an infinite number of proportional methods in integers, and any particular divisor method can contain several solutions in case of "ties". Although the number of different divisor method apportionments to any one problem is small and the apportionments themselves are closely related, to be useful in practice the choice must somehow be limited to a particular method. But which one?

The application of the methods of Adams, Webster and Jefferson can yield very different solutions, as witness the problem reported in Table 1. The data come from the French Parliamentary election of March 1986. The 577 seats of metropolitan France have been apportioned by the three methods on the basis of the matrix $\mathbf{p} = (p_{ij})$, where p_{ij} is the number of votes obtained in department i by political party j, and σ corresponds to asking only that each department receives a minimum of two deputies. The last row gives the apportionment determined by law and found as follows. Each department is apportioned a minimum of 2 seats found by the method of Adams on the basis of the populations of the departments. Then Jefferson's method is used to apportion the seats of each department among the parties on the basis of their departmental vote totals.

Thus some criteria or operational principles that are basic to the context of the practical problem that is at hand must be invoked to choose a particular method.

	Adams	Webster	Jefferson	law
Socialist	153	201	219	216
RPR (right)	130	162	171	162
UDF (right)	117	112	104	129
Communist	90	55	42	35
National Front (extreme right)	87	47	41	35

TABLE 1

Consider first the context of apportionment. The following summarizes some of the salient results obtained in the case of vector problems [6]. Take an apportionment method A by which a political party with p votes gets a seats and another with q votes gets b seats. If these two parties form a coalition having $p + q$ votes, whereas every other party's vote does not change, how many seats will the coalition receive by A? A *encourages coalitions* if it has a solution that gives at least $a + b$ seats to the coalition, and it *encourages schisms* if it has a solution that gives at most $a + b$ seats to the coalition. *Jefferson's method* is the unique divisor method that encourages coalitions, and *Adams's method* is the unique divisor method that encourages schisms. In the context of proportional representation this characterization is a strong argument in favor of Jefferson's method.

A seemingly natural consideration is this. Let $\lfloor x \rfloor$ be the greatest integer less than or equal to x, and $\lceil x \rceil$ the least integer greater than or equal to x. An apportionment method A satisfies *lower fair share* if it has solutions with $a_i \geq \lfloor f_i \rfloor$, it *satisfies upper fair share* if it has solutions with $a_i \leq \lceil f_i \rceil$, and it *satisfies fair share* if both of the prior conditions are met, that is, $\lfloor f_i \rfloor \leq a_i \leq \lceil f_i \rceil$. For vector problems there exists *no divisor method* that satisfies fair share, *Jefferson's* is the unique divisor method that satisfies lower fair share, and *Adams's* is the unique divisor method that satisfies upper fair share. The underlying concept of satisfying fair share is simply not a "proportional" idea. The intuitive reason for this is clear: forcing a large party's or state's apportionment to satisfy fair share is in the context of proportionality a much more restrictive demand than forcing a small party's or state's apportionment to satisfy fair share. A weakened notion for vector problems can be met: A is *near fair share* if it has solutions for which no transfer of a seat from one party or state to another brings both apportionments nearer to their fair shares. *Webster's* is the unique divisor method that is near fair share. In the context of regional representation and "one man, one vote", this characterization provides a strong argument in favor of Webster's method. Added to this is the statistical observation that for all practical intents and purposes Webster apportionments (almost) always satisfy fair share.

A method has *no bias* if the average of the fair shares is the same as the average of its apportionments over many problems. Simple inspection of several Adams and Jefferson apportionments convinces one that the method of Adams is significantly biased in favor of small states in that the average of the apportionments of a small state is significantly higher than the average of its fair shares whereas the opposite occurs for large states, and the method of Jefferson is significantly biased in favor of the large states. It has been shown that for vector problems the method of Webster is the only unbiased method [6] over a particular sample space. Statistical analysis of apportionment data drawn from United States history [6] and simulations concerning projected 1990 and 2000 census figures [8] confirm this conclusion. In light of these results it is surprising that Hill's method, $d(a) = \sqrt{a(a + 1)}$ currently used in the United States was adopted in 1941 on the (false) grounds it is unbiased! In the case of the constrained matrix problem we have no mathematical results. The only divisor method that has the possibility of being unbiased over all matrix problems—free or not free—is of course, Webster's, and preliminary experimental computation indicates that Webster's method is unbiased for equality constrainted matrix problems as well.

Consider now the problem of reporting census data [12]. In Canada, as well as most countries, census data are gathered with a guarantee of anonymity [19]. Given, however, arrays of many tables containing cross classifications concerning a community that is small in numbers, it may be possible to deduce the identity of certain subjects, and so fail the anonymity guarantee. To foil this possibility the counts in Canada are recorded as multiples of 5. If one wishes the data of the tables to add precisely to given row and column totals and be multiples of 5, the same apportion-

ment problem is faced. Indeed the problem of apportionment may be said to concern how to round data so that all sums are precise: given the data p_{ij}, what a_{ij} multiples of 5, *should* replace the corresponding p_{ij}? Again the answer seems to be, proportional numbers. But then which divisor method should be used? In this context, the paramount concern is to produce proportional matrices that are "unbiased". Two alternative approaches commend themselves. If the sums of the rounded entries in each row (each column) is to be an unbiased rounding of $\Sigma_j p_{ij}$ (of $\Sigma_i p_{ij}$), then Webster's method should be used to determine $\mathbf{r} = (r_i)$, $\Sigma r_i = h$ (to determine $\mathbf{c} = (c_j)$, $\Sigma c_j = h$) and then Webster's method to determine a proportional integer matrix with the corresponding equation constraints. If, however, the precise sums in rows and columns are not of paramount interest, then an allowable margin of error could be applied to first determine lower and upper constraints on rows, $\mathbf{r}^-$ and $\mathbf{r}^+$, and on columns $\mathbf{c}^-$ and $\mathbf{c}^+$, then solving the corresponding matrix problem by Webster's method.

The results so obtained will definitely be different than those obtained by the approach heretofore considered [12]. It begins by calculating the *quota* of every piece of data: $q'_{ij} = p_{ij} h / p_{MN}$, $r'_i = \Sigma_j q'_{ij}$, $c'_j = \Sigma_i q'_{ij}$, where $h = \Sigma_i r'_i = \Sigma c'_j$. Then, by a classical result of flow theory [13], the polyhedron R'

$$R' = \left\{ \mathbf{a}: \lfloor q_{ij} \rfloor \leqslant a_{ij} \leqslant \lceil q_{ij} \rceil, \lfloor r'_i \rfloor \leqslant a_{iN} \leqslant \lceil r'_i \rceil, \lfloor c'_j \rfloor \leqslant a_{Mj} \leqslant \lceil c'_j \rceil, a_{MN} = h \right\}$$

contains integer solutions. The current approach chooses an $\mathbf{a} \in R'$ that minimizes some measure of distance of $\mathbf{a}$ from the respective quotas. The difficulty here is that these solutions are *not* proportional. Even the determination $\lfloor r'_i \rfloor \leqslant a_{iN} \leqslant \lceil r'_i \rceil$ results in a vector of row sums (a_{iN}) that is not proportional to (p_{iN}). This choice is analogous to using Hamilton's method or some method that "satisfies fair share" and this, we have already seen, cannot be realized by a method that is proportional.

References

[1]   *Annales de l'INSEE*. (1976). Analyse et ajustement de tableaux statistiques. 22–23, Avril–Septembre.

[2]   Anthonisse, J. M. (1984). Proportional Representation in a Regional Council. *CWI Newsletter* (Centrum voor Wiskunde en Informatica), 5 (December) 22–29.

[3]   Aumann, R. J. and Mashler, M. (1985). Game Theoretic Analysis of a Bankruptcy Problem from the Talmud. *J. Economic Theory*, 36, 2 195–213.

[4]   Bacharach, M. (1970). *Biproportional Matrices and Input–Output Change*. Cambridge University Press, Cambridge, England.

[5]   Balinski, M. L. and Demange, G. (1987). Algorithms for Proportional Matrices in Reals and Integers. *Cahier du Laboratoire d'Econométrie*, Ecole Polytechnique, Avril. To appear in *Math. Programming, Ser. B*.

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

[7]   _____ and _____ (1978). Stability, Coalitions and Schisms in Proportional Representation Systems. *Amer. Political Sci. Rev.* 72, 3 848–858.

[8]   _____ and _____ (1984). Evaluation of Apportionment Methods. Report prepared for Congressional Research Service, U.S. Library of Congress, September.

[9]   Brualdi, Richard A. (1968). Convex Sets of Non-Negative Matrices. *Canad. J. Math.* 20 144–157.

[10] Brualdi, R. A., Parter, S. V. and Schneider, H. (1966). The Diagonal Equivalence of a Nonnegative Matrix to a Stochastic Matrix. *J. Math. Anal. Appl.* 16, 1 31–50.

[11] Censor, Y. and Lent, A. (1981). An Interactive Row-Action Method for Interval Convex Programming. *J. Optim. Theory Appl.* 34 321–353.

[12] Cox, Lawrence H. and Ernst, Lawrence R. (1982). Controlled Rounding. *INFOR* 20 423–432.

[13] Ford, Jr., L. R. and Fulkerson, D. R. (1962). *Flows in Networks*. Princeton University Press, Princeton, NJ.

[14] Ireland, C. T. and Kullback, S. (1968). Contingency Tables with Given Marginals. *Biometrika* **55** 179–188.

[15] Nagy, Andras (1979). *Methods of Structural Analysis and Projection of International Trade*, Study 13, Hungarian Academy of Sciences, Institute of Economics, Budapest.

[16] Potts, R. B. and Oliver, R. M. (1972). *Flows in Transportation Networks*. Academic Press, New York.

[17] Pyatt, G. and Round, J. J. (Eds.) (1985). *Social Accounting Matrices: A Basis for Planning*. World Bank, Washington D.C.

[18] Sinkhorn, R. (1967). Diagonal Equivalence to Matrices with Prescribed Row and Column Sums. *Amer. Math. Monthly* **74** 402–405.

[19] _____ and Knopp, Paul (1967). Concerning Nonnegative Matrices and Doubly Stochastic Matrices. *Pacific J. Math.* **21** 343–348.

[20] Statistics Canada (1982). *1981 Census of Canada—Population*, Minister of Supply and Services Canada, September.

[21] Young, H. Peyton (1987). On Dividing an Amount According to Individual Claims or Liabilities. *Math. Oper. Res.* **12** 398–414.

LABORATOIRE D'ECONOMETRIE, ECOLE POLYTECHNIQUE, 1 RUE DESCARTES, 75230, PARIS, FRANCE