Biproportional scaling of matrices and the iterative proportional fitting procedure

Friedrich Pukelsheim

Springer

Springer

Biproportional scaling of matrices and the iterative proportional fitting procedure

Friedrich Pukelsheim

Abstract A short proof is given of the necessary and sufficient conditions for the convergence of the Iterative Proportional Fitting procedure. The input consists of a nonnegative matrix and of positive target marginals for row sums and for column sums. The output is a sequence of scaled matrices to approximate the biproportional fit, that is, the scaling of the input matrix by means of row and column divisors in order to fit row and column sums to target marginals. Generally it is shown that certain structural properties of a biproportional scaling do not depend on the particular sequence used to approximate it. Specifically, the sequence that emerges from the Iterative Proportional Fitting procedure is analyzed by means of the L_1-error that measures how current row and column sums compare to their target marginals. As a new result a formula for the limiting L_1-error is obtained. The formula is in terms of partial sums of the target marginals, and easily yields the other well-known convergence characterizations.

Keywords Alternating scaling algorithm · Biproportional fitting · Matrix scaling · RAS procedure

1 Introduction

A novel L_1-based analysis of the Iterative Proportional Fitting (IPF) procedure is presented. The IPF procedure is an algorithm for scaling rows and columns of an input $k \times \ell$ weight matrix $A = ((a_{ij}))$ so that the output matrix $B = ((b_{ij}))$ achieves row sums equal to a prespecified vector of row marginals, $r = (r_1, \ldots, r_k)$, and column sums equal to a prespecified vector of column marginals, $s = (s_1, \ldots, s_\ell)$. We assume that all weights are nonnegative, $a_{ij} \geq 0$, and that at least one entry in each row and column of A is positive. All marginals are taken to be positive, $r_i > 0$ and $s_j > 0$.

Dedicato alla memoria di Bruno Simeone (1945–2010).

F. Pukelsheim (✉)
Institut für Mathematik, Universität Augsburg, 86135 Augsburg, Germany
e-mail: Pukelsheim@Math.Uni-Augsburg.De

The problem family has a continuous member, the biproportional fitting problem, and a discrete member, the biproportional apportionment problem. The continuous case allows the entries of B to be nonnegative, $b_{ij} \in [0, \infty)$. The output B is called the biproportional fit of the weight matrix A to the target marginals r and s. The IPF procedure produces scaled matrices $A(t) = ((a_{ij}(t)))$ with matching row sums for odd steps, $a_{i+}(t + 1) = r_i$ for all $i \leq k$, and matching column sums for even steps, $a_{+j}(t + 2) = s_j$ for all $j \leq \ell$. Throughout this paper a subscript plus-sign indicates summation over the index concerned, such as $a_{i+} = \sum_{j \leq \ell} a_{ij}$ etc. If the biproportional fit B exists, the sequence of scaled matrices $A(t)$, $t \geq 0$, converges to B.

In the discrete case the entries of B are restricted to be natural numbers, $b_{ij} \in \{0, 1, 2, \ldots\}$, as are the marginal components r_i and s_j. Now the output B is called a biproportional apportionment, for the weight matrix A and the target marginals r and s. The discrete problem may be solved using the Alternating Scaling (AS) algorithm. It produces scaled matrices $A(t)$, but rounds its entries $a_{ij}(t)$ to integers. There are rare instances when the biproportional apportionment B exists while the AS algorithm fails to converge to it. An example is given by Gaffke and Pukelsheim (2008, p. 157).

The present paper focuses on the continuous fitting problem where the relevant terms $r_i, s_j, a_{ij}(t), b_{ij}$ are not restricted to be integers. Yet our major tool, the L_1-error function

$$f\big(A(t)\big) = \sum_{i \leq k} \big|a_{i+}(t) - r_i\big| + \sum_{j \leq \ell} \big|a_{+j}(t) - s_j\big|,$$

is borrowed from Balinski and Demange's (1989a, 1989b) inquiry into the discrete apportionment problem. In the discrete case the error function is quite suggestive, simply counting how many units are wrongly allocated in step t. For the continuous problem the L_1-error is, at first glance, just one out of many ways to assess a lack of fit. At second glance it is a most appropriate way, as this paper endeavors to show.

1.1 The literature on biproportional fitting

The continuous biproportional fitting problem is the senior member of the problem family. It has created an enormous body of literature; we review only the papers that influenced the present research. The term *IPF procedure* prevails in Statistics, see Fienberg and Meyer (2006), or Speed (2005). Some Statisticians speak of *matrix raking*, such as Fagan and Greenberg (1987). In Operations Research and Econometrics the label *RAS method* is popular, pointing to a (diagonal) matrix R of row multipliers, the weight matrix A, and a (diagonal) matrix S of column multipliers, as mentioned already by Bacharach (1965, 1970). Computer scientists prefer the term *matrix scaling*, as in Rote and Zachariasen (2007).

The IPF procedure became popular through Deming and Stephan (1940), see Fienberg and Meyer (2006). Deming and Stephan (1940, p. 440) recommend terminating iterations when *the table reproduces itself*. This closeness is what is measured by the L_1-error function $f(A(t))$, see the display preceding Lemma 2 in Sect. 3. While successfully advocating the merits of the algorithm, Deming and Stephan were somewhat led astray in its analysis, as communicated by Stephan (1942).

Brown (1959) proposes a convergence proof that Ireland and Kullback (1968) criticize to lack rigor. The latter authors establish convergence by relating the IPF procedure to the minimum entropy solution. Csiszár (1975, p. 155) notes that their argument is flawed, and that the generalization to measure-spaces by Kullback (1968) suffers from a similar deficiency. Csiszár (1975) salvages the entropy approach, and Rüschendorf (1995) extends it to

general measure-spaces. Rüschendorf and Thomsen (1993, 1997) rectify a technical detail that escaped Csiszár's (1975) attention.

Ireland and Kullback (1968, Eqs. (4·32) and (4·33)) ultimately substitute convergence of entropy by convergence in L_1, referring to a result of Kullback (1966). Also Bregman (1967) starts out with entropy, and then uses the L_1-error function. Here we focus on the L_1-approach from start to finish. Ireland and Kullback (1968, p. 184) prove that the entropy criterion decreases monotonically, as does the likelihood function of Bishop et al. (1975, p. 86), and the L_1-error function, see Bregman (1967, p. 197). Marshall and Olkin (1968) and Macgill (1977) minimize a quadratic objective function. Pretzel (1980) uses a geometric matrix-mean and makes do with the arithmetic-geometric-mean inequality. The computational complexity of the IPF procedure is investigated by Kalantari et al. (2008).

The existence question is studied by many authors, such as Brualdi et al. (1966), Rothblum and Schneider (1989), Schneider (1990), Brown et al. (1993), Hershkowitz et al. (1997). Some papers employ network and graph theory by interpreting the problem as a transportation problem, as reviewed by Pukelsheim et al. (2012). Since this author was unable to smooth the arguments used here by employing graph theoretic tools the present paper's language is that of linear algebra.

Fienberg (1970) opens up a different route by embedding the IPF procedure into the geometry of the manifold of constant interaction in a $(k\ell - 1)$-dimensional simplex of reference. The author works with the assumption that all input weights are positive, $a_{ij} > 0$. He points out (p. 915) that the extension to problems involving zero weights is *quite complex*, which is attested to by much of the literature. Ireland and Kullback's (1968, p. 182) plea of assuming positive weights in order *to simplify the argument* is a friendly understatement, unless it is meant to be the utter truth.

Yet another approach, staying close to calculus and linear algebra, is due to Bacharach (1965, 1970), and Sinkhorn (1964, 1966, 1967, 1972, 1974) and Sinkhorn and Knopp (1967). Our viewpoint follows their lead. Michael Owen Leslie Bacharach (b. 1936, d. 2002) was an Oxford econometrician. In 1965 he earned a PhD degree in Mathematics from Cambridge. His thesis was published as Bacharach (1965), and became Sect. 4 of Bacharach (1970). Richard Dennis Sinkhorn (b. 1934, d. 1995) received his Mathematics PhD in 1962 from the University of Wisconsin–Madison, with a thesis entitled *On Two Problems Concerning Doubly Stochastic Matrices*. Throughout his career he served as a Mathematics professor with the University of Houston. Though contemporaries, neither of the two ever quoted the other.

1.2 The literature on biproportional apportionment

The discrete biproportional apportionment problem is the junior member of the problem family. It was put forward first by Balinski and Demange (1989a, 1989b), see also Simeone and Pukelsheim (2006). The rounding of scaled quantities to integers is appropriate for the statistical analysis of frequency tables, as noted by Wainer (1998) and Pukelsheim (1998). It disposes of any disclaimer that the adjusted figures are *rounded off, hence when summed may occasionally disagree a unit or so*, as warned in Table I of Deming and Stephan (1940, p. 433). When calculating percentages, as in Table 3.6-4 of Bishop et al. (1975, p. 99), the method finishes off with 100 percent and does not stop short with 99 percent. Yet Balinski's motivation was not contingency table analysis in statistics, but proportional representation systems for parliamentary elections.

The task of allocating seats of a parliamentary body to political parties does not tolerate any disclaimer excusing residual rounding errors. Methods must account for every seat.

This is achieved by biproportional methods. In 2003, the Swiss Canton of Zurich adopted a doubly proportional system, the biproportional divisor method with standard rounding, see Pukelsheim and Schuhmacher (2004, 2011), and Balinski and Pukelsheim (2006). The method may be attractive also for other countries as studied by Pennisi (2006) for Italy, Zachariassen and Zachariasen (2006) for the Farœ Islands, Ramírez et al. (2008) for Spain, and Oelbermann and Pukelsheim (2011) for the European Union. A computer program for the calculations is provided at www.uni-augsburg.de/bazi, see Pukelsheim (2004), Joas (2005), Maier (2009). The user may choose to run the IPF procedure, the AS algorithm, the Tie-and-Transfer (TT) algorithm of Balinski and Demange (1989b), or various hybrid combinations. The performance of these algorithms is studied by Maier et al. (2010).

In the electoral application the entries a_{ij} in the input weight matrix A signify vote counts. When a party j does not campaign in district i it enters into the final evaluations with vote count $a_{ij} = 0$. Therefore weights that are zero must be properly dealt with, even if the labor entailed becomes quite complex. It is no longer appropriate to assume all weights to be positive in order to simplify the argument.

1.3 Section overview

A brief overview of the paper is as follows. Section 2 investigates biproportional scalings of a given weight matrix A. A biproportional scaling reweighs rows and columns of A moderately enough so that none of the rows nor columns is annihilated. If two scalings share the same row and column sums, then they coincide (Theorem 1). A scaling is called direct when a passage to the limit, though allowed by definition, becomes superfluous. In such a case the input matrix A and the output matrix B decompose accordingly (Lemma 1), see also Balinski and Demange (1989a), Gietl (2009). Theorem 2 establishes necessary and sufficient conditions to check for directness.

Section 3 turns to the IPF procedure by adjoining the goal to match prespecified row marginals r and prespecified column marginals s. In each step t a scaled weight matrix $A(t)$ is produced that has matching rows for odd t, and matching columns for even t. The fit of $A(t)$ is measured by the L_1-error function mentioned above. Lemma 2 ascertains that the L_1-error is non-increasing, and admits a finite set of lower bounds. The largest lower bound turns out to be equal to the L_1-error limit. The L_1-error limit formula is new. Its proof makes use of the result that the IPF even-step subsequence is always convergent, as shown by Gietl and Reffel (2013) on the basis of the information divergence approach of Csiszár and Tusnády (1984). The result replaces the IPF conjecture on which a previous version of the present paper relied, see Pukelsheim (2013). Based on the limiting L_1-errror formula Theorem 3 offers a brief proof of the necessary and sufficient conditions for the IPF sequence to converge to the biproportional fit.

2 Biproportional scalings

Let $A = ((a_{ij}))$ be a given $k \times \ell$ *weight matrix*, that is, A is assumed to have nonnegative entries and no zero row nor zero column. Whether rows $i \leq k$ or columns $j \leq \ell$ do not vanish is conveniently read off from their component sums, $a_{i+} > 0$ and $a_{+j} > 0$. Another weight matrix B is said to *preserve the zeros of* A when all zeros of A are zeros also of B, $a_{ij} = 0 \Rightarrow b_{ij} = 0$. Two matrices A and B *have the same zeros* when $a_{ij} = 0 \Leftrightarrow b_{ij} = 0$. We concentrate on true matrix problems, $k \geq 2$ and $\ell \geq 2$.

Definition 1 A $k \times \ell$ weight matrix $B = ((b_{ij}))$ is called a *biproportional scaling of A* when for all rows $i \le k$ and for all columns $j \le \ell$ there are sequences of positive *row divisors* $\rho_i(1), \rho_i(2), \ldots$ and of positive *column divisors* $\sigma_j(1), \sigma_j(2), \ldots$ such that

$$b_{ij} = \lim_{n \to \infty} \frac{a_{ij}}{\rho_i(n) \sigma_j(n)}.$$

A biproportional scaling B is termed *direct* when its associated divisor sequences can be chosen to be constant, that is, for all rows $i \le k$ and for all columns $j \le \ell$ there are positive divisors μ_i and ν_j such that $b_{ij} = a_{ij}/(\mu_i \nu_j)$.

The requirement that B is a weight matrix necessitates positive row and column sums and imposes a certain balance on row and column divisors. Neither can grow unboundedly of an order dwarfing the other lest a vanishing row or column invalidates the definition. Furthermore the requirement implies uniqueness of biproportional scalings that share the same marginals.

Theorem 1 (Uniqueness) *If two biproportional scalings B and C of a weight matrix A share the same row and column sums, $b_{i+} = c_{i+}$ for all rows $i \le k$ and $b_{+j} = c_{+j}$ for all columns $j \le \ell$, then they coincide, $B = C$.*

Proof The proof is by contraposition. Assuming the two scalings to be distinct, $B \ne C$, the difference $B - C$ is nonzero, but has row and column sums vanishing. We construct a cycle of cells

$$(i_1, j_1), (i_1, j_2), (i_2, j_2), (i_2, j_3), \ldots, (i_{q-1}, j_{q-1}), (i_{q-1}, j_q), (i_q, j_q), (i_q, j_1) \qquad \text{(CC)}$$

along which the entries in $B - C$ alternate in sign. First we assemble a "long list" of cells $(i_1, j_1), (i_1, j_2), (i_2, j_2), \ldots, (i_Q, j_Q), (i_Q, j_1)$, as follows. We start with a cell (i_1, j_1) where $b_{i_1 j_1} > c_{i_1 j_1}$. In row i_1 there is a cell (i_1, j_2) with $b_{i_1 j_2} < c_{i_1 j_2}$. Next we search in column j_2 a row i_2 where $b_{i_2 j_2} > c_{i_2 j_2}$. Then we look for a column j_3 such that $b_{i_2 j_3} < c_{i_2 j_3}$. The long list terminates when encountering a column j_Q already listed, that is, when for some $P > Q$ we find $j_P = j_Q$. The initial $P - 1$ cells are discarded, and the remaining "short list" is relabeled as in (CC).

A *cyclic ratio* is a ratio having the entries along a given cell cycle alternately appear in the numerator and in the denominator. Since $a_{ij} = 0$ implies $b_{ij} = c_{ij} = 0$, the cell cycle [CC] involves only positive entries of the weight matrix A. Let $\rho_i(n)$ and $\sigma_j(n)$ be the divisor sequences for B, and $\mu_i(n)$ and $\nu_j(n)$ for C. Since biproportionality preserves cyclic ratios, the cyclic ratios in A, B, and C are seen to be equal,

$$\prod_{p \le q} \frac{a_{i_p j_p}}{a_{i_p j_{p+1}}} = \prod_{p \le q} \frac{\frac{a_{i_p j_p}}{\rho_{i_p}(n) \sigma_{j_p}(n)}}{\frac{a_{i_p j_{p+1}}}{\rho_{i_p}(n) \sigma_{j_{p+1}}(n)}} = \prod_{p \le q} \frac{b_{i_p j_p}}{b_{i_p j_{p+1}}} = \prod_{p \le q} \frac{\frac{a_{i_p j_p}}{\mu_{i_p}(n) \nu_{j_p}(n)}}{\frac{a_{i_p j_{p+1}}}{\mu_{i_p}(n) \nu_{j_{p+1}}(n)}} = \prod_{p \le q} \frac{c_{i_p j_p}}{c_{i_p j_{p+1}}},$$

where $j_{q+1} = j_1$. The first and third equation signs are obvious. The fourth equality involves a passage to the limit as n tends to ∞ and holds since the limiting denominator is positive, $c_{i_p j_{p+1}} > b_{i_p j_{p+1}} \ge 0$. As the quotient is positive, the numerator must be positive, too, $c_{i_p j_p} > 0$. A similar argument establishes the second equality.

But the cycle (CC) precludes equality, $\prod_{p \leq q} b_{i_p j_p} / b_{i_p j_{p+1}} > \prod_{p \leq q} c_{i_p j_p} / c_{i_p j_{p+1}}$. Hence the assumption $B \neq C$ is untenable and uniqueness obtains, $B = C$. □

Directness of a biproportional scaling is closely related to the notion of connectedness. However, disconnectedness is defined first. A nonzero matrix D is called *disconnected* when a suitable permutation of rows and a suitable permutation of columns give rise to a row subset I and a column subset J such that D acquires block structure,

$$D = \begin{array}{c} \\ I \\ I' \end{array} \overset{\textstyle \begin{array}{cc} J & J' \end{array}}{\begin{pmatrix} D^{(1)} & 0 \\ 0 & D^{(2)} \end{pmatrix}},$$

where at least one of the subsets I or J is non-empty and proper, $\emptyset \subsetneq I \subsetneq \{1, \ldots, k\}$ or $\emptyset \subsetneq J \subsetneq \{1, \ldots, \ell\}$. Throughout this paper a prime denotes a set complement, such as $I' = \{1, \ldots, k\} \setminus I$ etc. In most applications both subsets I and J are non-empty and proper. A nonzero matrix C is said to be *connected* when it is not disconnected.

For keeping track of the nonzero entries in a weight matrix A we associate with every row subset $I \subseteq \{1, \ldots, k\}$ the *set of columns connected in A with I*,

$$J_A(I) = \{ j \leq \ell \mid a_{ij} > 0 \text{ for some } i \in I \}.$$

The complement $J_A(I)'$ embraces the columns j with entries $a_{ij} = 0$ for all $i \in I$. Hence the $I \times J_A(I)'$ submatrix of A vanishes and the sum of its entries is zero, $a_{I \times J_A(I)'} = 0$. The extreme settings provide simple examples. If we choose $I = \{1, \ldots, k\}$ then we get $J_A(I) = \{1, \ldots, \ell\}$, since no row nor column of A vanishes. If $I = \emptyset$ then $J_A(I) = \emptyset$.

The term "connectness" is used in the same way that is familiar from graph theory. The matrix A is connected, in the sense of linear algebra as defined above, if and only if the bipartite graph with rows and columns as vertices and with non-zero entries in A as edges is connected, in the sense of graph theory.

When the input weight matrix A is disconnected, the calculation of a biproportional scaling B decomposes into several separate instances. Hence there is no loss of generality of assuming A to be connected. An ensuing biproportional scaling B may still turn out to be disconnected. If so, the pattern of zeros in the output matrix B has repercussions on the pattern of zeros in the input matrix A, as follows.

Lemma 1 (Joint decomposition) *For every connected weight matrix A and for every disconnected biproportional scaling B of A there exists a non-empty and proper subset I of rows, $\emptyset \subsetneq I \subsetneq \{1, \ldots, k\}$, such that A acquires block-triangular structure and B acquires block-diagonal structure,*

$$A = \begin{array}{c} \\ I \\ I' \end{array} \overset{\textstyle \begin{array}{cc} J_A(I) & J_A(I)' \end{array}}{\begin{pmatrix} A^{(1)} & 0 \\ A^{(2,1)} & A^{(2)} \end{pmatrix}}, \qquad B = \begin{array}{c} \\ I \\ I' \end{array} \overset{\textstyle \begin{array}{cc} J_B(I) & J_B(I)' \end{array}}{\begin{pmatrix} B^{(1)} & 0 \\ 0 & B^{(2)} \end{pmatrix}},$$

where the sets of columns connected in A or B with I are equal, $J_A(I) = J_B(I)$.

Proof Without loss of generality we assume the largest accumulation point of the column divisors to be positive, $\limsup_{n \to \infty} \sigma_{\max}(n) = M \in (0, \infty)$, where $\sigma_{\max}(n) = \max\{\sigma_1(n), \ldots, \sigma_\ell(n)\}$. If need be, we would adjust the divisors according to $\widetilde{\rho}_i(n) = \rho_i(n)\sigma_{\max}(n)$ and $\widetilde{\sigma}_j(n) = \sigma_j(n)/\sigma_{\max}(n) \leq 1$, and use $M = 1$. Let an arrow $\to$ indicate a passage to the limit

as n tends to infinity. The set I is defined to contain the rows with divisors not degenerating, in the sense of not diverging to infinity,

$$I = \{i \le k \mid \rho_i(n) \nrightarrow \infty\}, \qquad I' = \{i \le k \mid \rho_i(n) \to \infty\}.$$

Likewise the columns connected in B with I will turn out to have their divisors not degenerating, but now in the sense of not converging to zero,

$$J_B(I) = \{j \le \ell \mid \sigma_j(n) \nrightarrow 0\}, \qquad J_B(I)' = \{j \le \ell \mid \sigma_j(n) \to 0\}.$$

With A connected and B disconnected there exists a cell (i, j) that is fading, $a_{ij} > 0 = b_{ij}$. This implies $\lim_{n \to \infty} \rho_i(n) \sigma_j(n) = \infty$ and, since column divisors stay bounded by assumption, $\lim_{n \to \infty} \rho_i(n) = \infty$. Hence I' is not empty, nor is $J_B(I')$—which in the end will turn out to coincide with $J_B(I)'$. By definition of $J_B(I')$, the $I' \times J_B(I')'$ block of B is zero.

The columns $j \in J_B(I')$ have their divisors converge to zero, $\lim_{t \to \infty} \sigma_j(n) = 0$. Indeed, there exists a row $i \in I'$ with $b_{ij} > 0$. In this cell we have $\lim_{n \to \infty} \rho_i(n) \sigma_j(n) = a_{ij}/b_{ij} < \infty$. Since the divisors of row $i \in I'$ diverge to infinity, column j has its divisors converge to zero.

By assumption the column divisors satisfy $\sigma_{\max}(n) > M/2$ infinitely often. Thus there exists a column j with divisors fulfilling $\sigma_j(n) > M/2$ again and again and not converging to zero. Hence $J_B(I')'$ is not empty, nor is I.

Now every column $j \in J_B(I')'$ has its divisors not converging to zero, $\limsup_{n \to \infty} \sigma_j(n) > 0$. Indeed, there is a row $i \in I$ with $b_{ij} > 0$. In this cell we have $\lim_{n \to \infty} \rho_i(n) \sigma_j(n) = a_{ij}/b_{ij} > 0$. Since column j admits a divisor subsequence bounded away from zero, the row divisors that go along cannot diverge to infinity.

The $I \times J_B(I')$ top right block of A has $a_{ij} = 0$. Indeed, the case $a_{ij} > 0$ would admit a row divisor subsequence bounded from above, in the presence of column divisors converging to zero. But $a_{ij} > 0$ and $\lim_{n \to \infty} \rho_i(n) \sigma_j(n) = 0$ lead to the contradiction $b_{ij} = \infty$. The top right block of B inherits the zeros of A. The structures of B and A entail $J_B(I')' = J_B(I) = J_A(I)$. □

Given a biproportional scaling B, there are various ways to check for directness.

Theorem 2 (Directness) *For every connected weight matrix A and for every biproportional scaling B of A the following five statements are equivalent*:

(a) *The biproportional scaling B is direct.*
(b) *The matrices A and B have the same zeros.*
(c) *There exists a weight matrix C sharing the same zeros with A and the same row and column sums with B.*
(d) *For every non-empty and proper subset I of rows, $\emptyset \subsetneq I \subsetneq \{1, \dots, k\}$, partial row and column sums of B fulfill $\sum_{i \in I} b_{i+} < \sum_{j \in J_A(I)} b_{+j}$.*
(e) *The matrix B is connected.*

Proof

(a) $\Rightarrow$ (b). A direct scaling, $b_{ij} = a_{ij}/(\mu_i \nu_j)$, has the same zeros as has A.

(b) $\Rightarrow$ (c). The scaling B, sharing all zeros with A, is of the type asked for in (c).

(c) $\Rightarrow$ (d). For every row subset I we have $a_{I \times J_A(I)'} = 0$, and hence $c_{I \times J_A(I)'} = 0$. If I is non-empty and proper then $c_{I' \times J_A(I)} > 0$, as otherwise C were disconnected and so would be A. We get $\sum_{i \in I} b_{i+} = c_{I \times J_A(I)} < c_{I \times J_A(I)} + c_{I' \times J_A(I)} = \sum_{j \in J_A(I)} b_{+j}$.

(d) $\Rightarrow$ (e). The proof is by contraposition. If B is disconnected, then Lemma 1 provides a non-empty and proper row set I fulfilling $\sum_{i \in I} b_{i+} = b_{I \times J_A(I)} = \sum_{j \in J_A(I)} b_{+j}$.

(e) $\Rightarrow$ (a). Row divisors μ_i and column divisors ν_j for B are constructed in the course of a scanning process. The process is initialized by standardizing the given divisor sequences according to $\widetilde{\rho}_i(n) = \rho_i(n)/\rho_1(n)$ and $\widetilde{\sigma}_j(n) = \rho_1(n)\sigma_j(n)$, thus equipping the first row with constant divisor unity, $\widetilde{\rho}_1(n) = 1 = \mu_1$, $n \geq 1$. Then the process scans all columns j with $b_{1j} > 0$, and sets

$$0 < \nu_j = \frac{a_{1j}}{\mu_1 b_{1j}} = \frac{\lim_{n\to\infty} \widetilde{\rho}_1(n)\widetilde{\sigma}_j(n)}{\lim_{n\to\infty} \widetilde{\rho}_1(n)} = \lim_{n\to\infty} \widetilde{\sigma}_j(n), \quad \text{whence } b_{1j} = \frac{a_{1j}}{\mu_1 \nu_j}.$$

Next all unscanned rows i with $b_{ij} > 0$ for some scanned column j are scanned, setting

$$0 < \mu_i = \frac{a_{ij}}{b_{ij}\nu_j} = \frac{\lim_{n\to\infty} \widetilde{\rho}_i(n)\widetilde{\sigma}_j(n)}{\lim_{n\to\infty} \widetilde{\sigma}_j(n)} = \lim_{n\to\infty} \widetilde{\rho}_i(n), \quad \text{whence } b_{ij} = \frac{a_{ij}}{\mu_i \nu_j}.$$

Thereafter the process turns to columns again, then rows. In this fashion it keeps enlarging the scanned sets of rows and columns, terminating after at most $k + \ell$ steps. The terminal scanned row set I and column set J enforce a block structure upon B,

$$B = \begin{matrix} & \overset{J}{} & \overset{J'}{} \\ \begin{matrix} I \\ I' \end{matrix} & \begin{pmatrix} B^{(1)} & 0 \\ 0 & B^{(2)} \end{pmatrix} \end{matrix}.$$

Connectedness of B lets the scanned sets be exhaustive, $I = \{1, \ldots, k\}$ and $J = \{1, \ldots, \ell\}$. All rows and all columns having constant divisors, the scaling is direct. $\qquad\square$

We now turn to the Iterative Proportional Fitting (IPF) procedure proper.

3 The IPF procedure

The IPF procedure seeks to calculate a biproportional scaling B that achieves pre-specified row marginals r and pre-specified column marginals s. Let A be a given $k \times \ell$ weight matrix. Furthermore let $r = (r_1, \ldots, r_k)$ and $s = (s_1, \ldots, s_\ell)$ be given vectors with positive entries, called *target row marginals* and *target column marginals*.

Definition A $k \times \ell$ nonnegative matrix $B = ((b_{ij}))$ is said to *match the target marginals r and s* when its row sums are equal to r and its column sums are equal to s, that is, $b_{i+} = r_i$ for all $i \leq k$ and $b_{+j} = s_j$ for all $j \leq \ell$. A weight matrix B is called a *biproportional fit of the weight matrix A to the target marginals r and s* when B is a biproportional scaling of A and B matches the target marginals r and s.

With this terminology the IPF procedure aims to determine a biproportional fit of the weight matrix A to the target marginals r and s. If a fit exists then it is unique, by Theorem 1. An existing fit B necessitates equal totals of the marginals, $r_+ = b_{++} = s_+$. We do not assume equality of the totals, though, since the IPF procedure is well-defined without this assumption and since distinct marginal subtotals may evolve when the procedure leads to accumulation points that are disconnected.

The IPF procedure starts by scaling A into a matrix $A(0)$ with column sums equal to target column marginals. That is, with column divisors $\beta_j(0) = a_{+j}/s_j$ the entries of $A(0)$ are

defined to be $a_{ij}(0) = a_{ij}/\beta_j(0)$, for all $i \leq k$ and $j \leq \ell$. Thereafter the procedure advances in pairs of an odd step $t + 1$ and an even step $t + 2$, for $t = 0, 2, \ldots$:

- Odd steps $t + 1$ fit row sums to target row marginals by calculating row divisors $\alpha_i(t + 1)$ from the preceding even step t, and scaled weights $a_{ij}(t + 1)$:

$$\alpha_i(t + 1) = \frac{a_{i+}(t)}{r_i}, \tag{IPF1}$$

$$a_{ij}(t + 1) = \frac{a_{ij}(t)}{\alpha_i(t + 1)}, \tag{IPF2}$$

for all rows $i \leq k$ and for all columns $j \leq \ell$.
- Even steps $t + 2$ fit column sums to target column marginals by calculating column divisors $\beta_j(t + 2)$ from the preceding odd step $t + 1$, and scaled weights $a_{ij}(t + 2)$:

$$\beta_j(t + 2) = \frac{a_{+j}(t + 1)}{s_j}, \tag{IPF3}$$

$$a_{ij}(t + 2) = \frac{a_{ij}(t + 1)}{\beta_j(t + 2)}, \tag{IPF4}$$

for all columns $j \leq \ell$ and for all rows $i \leq k$.

The *incremental row divisors* $\alpha_i(1), \alpha_i(3), \ldots$ from [IPF1] give rise to *cumulative row divisors* ρ_i, and the *incremental column divisors* $\beta_j(2), \beta_j(4), \ldots$ from [IPF3] generate *cumulative column divisors* σ_j. They are defined, for steps $t = 0, 2, \ldots$, through

$$\alpha_i(1) \alpha_i(3) \cdots \alpha_i(t + 1) = \rho_i(t + 1) = \rho_i(t + 2),$$

$$\beta_j(0) \beta_j(2) \beta_j(4) \cdots \beta_j(t + 2) = \sigma_j(t + 2) = \sigma_j(t + 3).$$

Adjoining $\rho_i(0) = 1$ and $\sigma_j(0) = \sigma_j(1) = \beta_j(0)$, cumulative divisors are defined for all steps $t \geq 0$. The scaled weights take the form $a_{ij}(t) = a_{ij}/(\rho_i(t)\sigma_j(t))$.

The *scaled weight matrices* $A(t) = ((a_{ij}(t)))$, $t \geq 0$, constitute the *IPF sequence*, for the fitting of the weight matrix A to the target marginals r and s. An L_1-*error function* f is employed to assess the goodness-of-fit of a scaled weight matrix $A(t)$. It checks whether any row is *underfitted*, $a_{i+}(t) < r_i$, or *overfitted*, $a_{i+}(t) > r_i$, as well as whether any column is under- or overfitted, and then totals the absolute deviations between current sums and target marginals,

$$f\big(A(t)\big) = \sum_{i \leq k} |a_{i+}(t) - r_i| + \sum_{j \leq \ell} |a_{+j}(t) - s_j|.$$

Odd steps t have row sums matching their target marginals, whence the L_1-error $f(A(t))$ is equal to the (second) column-error sum. For even steps t, the (first) row-error sum is decisive.

The L_1-error coincides with the L_1-distance between a scaled weight matrix and its successor. To see this for an even step t, we substitute $r_i = a_{i+}(t)/\alpha_i(t + 1)$ from (IPF1) and $a_{ij}(t)/\alpha_i(t + 1) = a_{ij}(t + 1)$ from (IPF2) to obtain

$$f\big(A(t)\big) = \sum_{i \leq k} \left| 1 - \frac{1}{\alpha_i(t + 1)} \right| a_{i+}(t) = \sum_{i \leq k} \sum_{j \leq \ell} |a_{ij}(t) - a_{ij}(t + 1)|.$$

Definitions (IPF3) and (IPF4) confirm the result for odd steps t. Deming and Stephan (1940, p. 440) recommend that the IPF procedure *is continued until the table reproduces itself*. This is exactly what is captured by the error function f: The table reproduces itself, $A(t) = A(t+1)$, if and only if the L_1-error is zero, $f(A(t)) = 0$.

Lemma 2 shows that the L_1-error is non-increasing in t, admits a finite set of lower bounds, and converges to the largest of these bound as $t \to \infty$. The new limit formula in part (c) greatly simplifies the proof of the Convergence Theorem 3 below.

Lemma 2 (L_1-error) *Let $A(t)$, $t \geq 0$, be the IPF sequence for the fitting of the weight matrix A to the target marginals r and s. Then we have for all steps $t \geq 0$:*

(a) *The L_1-error is non-increasing, $f(A(t)) \geq f(A(t+1))$.*
(b) *The L_1-error is bounded according to $f(A(t)) \geq r_I - s_{J_A(I)} + s_{J_A(I)'} - r_{I'}$, for all row subsets $I \subseteq \{1, \ldots, k\}$.*
(c) *The limiting L_1-error is given by*

$$\lim_{t \to \infty} f(A(t)) = \max_{I \subseteq \{1, \ldots, k\}} (r_I - s_{J_A(I)} + s_{J_A(I)'} - r_{I'}).$$

Proof

(a) Let step $t \geq 0$ be even, whence $A(t)$ has fitted columns and its L_1-error originates from rows. From (IPF1) and $a_{i+}(t) = \alpha_i(t+1)r_i$ we get $f(A(t)) = \sum_{i \leq k} |a_{i+}(t) - r_i| = \sum_{i \leq k} |\alpha_i(t+1) - 1| r_i$. Insertion of $r_i = \sum_{j \leq \ell} a_{ij}(t+1)$ allows us to apply the triangle inequality within each column j,

$$\sum_{j \leq \ell} \sum_{i \leq k} |1 - \alpha_i(t+1)| a_{ij}(t+1) \geq \sum_{j \leq \ell} \left| \sum_{i \leq k} (1 - \alpha_i(t+1)) a_{ij}(t+1) \right|.$$

From (IPF2) and $\sum_{i \leq k} \alpha_i(t+1) a_{ij}(t+1) = a_{+j}(t) = s_j$ we obtain $\sum_{j \leq \ell} |a_{+j}(t+1) - s_j| = f(A(t+1))$. With Definitions (IPF3) and (IPF4) the argument carries over to odd steps $t \geq 1$. Thus monotonicity is established.

(b) Because of monotonicity we may assume step t to be even. Let I be a subset of rows. Since $\sum_{i \in I} (a_{i+}(t) - r_i) + \sum_{i \in I'} (a_{i+}(t) - r_i) = s_+ - r_+$, the complement I' satisfies $\sum_{i \in I'} (a_{i+}(t) - r_i) = s_+ - r_+ + r_I - \sum_{i \in I} a_{i+}(t)$. Hence we obtain

$$f(A(t)) \geq \sum_{i \in I} (r_i - a_{i+}(t)) + \sum_{i \in I'} (a_{i+}(t) - r_i) = s_+ - r_+ + 2r_I - 2 \sum_{i \in I} a_{i+}(t).$$

From $\sum_{i \in I} a_{i+}(t) = \sum_{i \in I} \sum_{j \in J_A(I)} a_{ij}(t) \leq \sum_{j \in J_A(I)} \sum_{i \leq k} a_{ij}(t) = s_{J_A(I)}$ we get

$$s_+ - r_+ + 2r_I - 2 \sum_{i \in I} a_{i+}(t) \geq s_+ - r_+ + 2r_I - 2s_{J_A(I)} = r_I - s_{J_A(I)} + s_{J_A(I)'} - r_{I'}.$$

This establishes the lower bounds in (b).

(c) Theorem 5.3 of Gietl and Reffel (2013) states that the IPF even-step subsequence is convergent, $\lim_{t=0,2,4,\ldots} A(t) = B$ say. In view of part (i) every accumulation point minimizes the L_1-error, whence $\lim_{t \to \infty} f(A(t)) = \lim_{t=0,2,4,\ldots} f(A(t)) = f(B)$. We claim that $U = \{i \leq k \mid b_{i+} < r_i\}$, the set of underfitted rows in B, fulfills

$$f(B) = r_U - s_{J_A(U)} + s_{J_A(U)'} - r_{U'}. \tag{1}$$

This and the boundedness part (b) establish the limit formula in part (c). If the set U is empty, $U = \emptyset$, then $f(B) = s_+ - r_+$ verifies (1). If the set U is exhaustive, $U = \{1, \ldots, k\}$, then $f(B) = r_+ - s_+$ verifies (1). Generally (1) is verified in four steps.

The first step claims that all row sums of B are positive, $b_{i+} > 0$ for all $i \leq k$. Indeed, Definition (IPF3) implies a uniform upper bound for the column divisors, $\beta_j(t+2) = a_{+j}(t+1)/s_j \leq a_{++}(t+1)/s_{\min} = r_+/s_{\min}$, where $s_{\min} = \min\{s_1, \ldots, s_k\}$. Now definition (IPF4) yields a lower bound,

$$a_{i+}(t+2) = \sum_{j \leq \ell} \frac{a_{ij}(t+1)}{\beta_j(t+2)} \geq \frac{s_{\min}}{r_+} a_{i+}(t+1) = \frac{s_{\min}}{r_+} r_i.$$

Thus the limiting row sums remain positive, $b_{i+} \geq s_{\min} r_i / r_+ > 0$, as claimed.

Hence the accumulation point B is a weight matrix itself. Let $\gamma_i = b_{i+}/r_i > 0$ denote the row divisor to scale row i of B so as to meet the target marginal r_i. Upon defining the matrix $B(1)$ with entries $b_{ij}(1) = b_{ij}/\gamma_i$, let $\delta_j = b_{+j}(1)/s_j$ denote the column divisor to fit column j to the target marginal s_j.

The second step establishes a relation between the divisors γ_i and δ_j,

$$b_{ij} > 0 \quad \Rightarrow \quad \gamma_i \delta_j = 1 \qquad \text{for all } i \leq k \text{ and } j \leq \ell. \tag{2}$$

To see this we note that γ_i is the limit of the incremental row divisors $\alpha_i(t+1)$,

$$\gamma_i = \frac{b_{i+}}{r_i} = \lim_{t=0,2,4,\ldots} \frac{a_{i+}(t)}{r_i} = \lim_{t=0,2,4,\ldots} \alpha_i(t+1), \tag{3}$$

as follows from (IPF1). Now (3), (IPF2), and (IPF3) yield

$$\delta_j = \frac{b_{+j}(1)}{s_j} = \sum_{i \leq k} \frac{b_{ij}}{\gamma_i s_j} = \lim_{t=0,2,4,\ldots} \sum_{i \leq k} \frac{a_{ij}(t)}{\alpha_i(t+1)s_j} = \lim_{t=0,2,4,\ldots} \beta_j(t+2). \tag{4}$$

From (IPF2) and (IPF4) we get $a_{ij}(t) = \alpha_i(t+1)a_{ij}(t+1) = \alpha_i(t+1)\beta_j(t+2)a_{ij}(t+2)$. Hence the L_1-distance between the scaled weight matrices $A(t)$ and $A(t+2)$ is

$$\sum_{i \leq k} \sum_{j \leq \ell} |a_{ij}(t) - a_{ij}(t+2)| = \sum_{i \leq k} \sum_{j \leq \ell} |\alpha_i(t+1)\beta_j(t+2) - 1| a_{ij}(t+2).$$

Convergence of the IPF even-step subsequence forces the L_1-distance to vanish in the limit. Together with (3) and (4) we obtain $0 = \sum_{i \leq k} \sum_{j \leq \ell} |\gamma_i \delta_j - 1| b_{ij}$, that is, (2).

The third step depicts the structure of B as follows:

$$B = \begin{array}{c} \\ U \\ U' \end{array} \begin{array}{cc} J_B(U) & J_B(U)' \\ \begin{pmatrix} B^{(1)} & 0 \\ 0 & B^{(2)} \end{pmatrix} & \begin{array}{c} \gamma_i < 1 \\ \gamma_i \geq 1 \end{array} \\ \delta_j > 1 \quad \delta_j \leq 1 \end{array}$$

Indeed, the top right block is zero because $J_B(U)$ contains all columns connected in B with U. Being underfitted rows $i \in U$ satisfy $\gamma_i < 1$. Complementary rows $i \in U'$ fulfill $\gamma_i \geq 1$. Columns $j \in J_B(U)$ are connected in B with U and have $\delta_j > 1$, by (2). Columns $j \in J_B(U)' \subseteq J_B(U')$, connected in B with U', have $\delta_j \leq 1$, also by (2). Finally the bottom

left block is seen to be zero, since $i \in U'$ and $j \in J_B(U)$ have $\gamma_i \delta_j > 1$ while $b_{ij} > 0$ would imply $\gamma_i \delta_j = 1$, by (2). Now the structure of B entails $\sum_{i \in U} b_{i+} = s_{J_B(U)}$ and $\sum_{i \in U'} b_{i+} = s_{J_B(U)'}$. The L_1-error of B, depending on rows only, turns into

$$f(B) = \sum_{i \in U}(r_i - b_{i+}) + \sum_{i \in U'}(b_{i+} - r_i) = r_U - s_{J_B(U)} + s_{J_B(U)'} - r_{U'}. \tag{5}$$

The fourth step establishes $J_B(U) = J_A(U)$. The inclusion $J_B(U) \subseteq J_A(U)$ holds true since $b_{ij} > 0$ implies $a_{ij} > 0$. As for the converse inclusion we note that the top right block $U \times J_B(U)'$ has $\gamma_i \delta_j < 1$. Since the products $\alpha_i(t+1)\beta_j(t+2)$ converge to $\gamma_i \delta_j < 1$, by (3) and (4), the cumulative divisors tend to zero, $\lim_{t \to \infty} \rho_i(t)\sigma_j(t) = 0$. Now $\lim_{t \to \infty} a_{ij}/(\rho_i(t)\sigma_j(t)) = b_{ij} = 0$ forces $a_{ij} = 0$. But $a_{ij} = 0$ on $U \times J_B(U)'$ implies $J_A(U) \subseteq J_B(U)$. This proves $J_A(U) = J_B(U)$, and turns (5) into (1). □

Lemma 2 allows a remarkably brief proof of the well-known necessary and sufficient conditions for the convergence of the IPF procedure. The arrangement of conditions runs parallel to the characterizations of directness in Theorem 2.

Theorem 3 (Convergence) *For the fitting of a weight matrix A to the target marginals r and s the following five statements are equivalent*:

(a) *The IPF sequence $A(t)$, $t \geq 0$, is convergent.*
(b) *The biproportional fit of A to r and s exists.*
(c) *There exists a weight matrix D preserving the zeros of A and matching the target marginals r and s.*
(d) *Marginal totals are equal, $r_+ = s_+$ and marginal partial row and column sums fulfill $r_I \leq s_{J_A(I)}$, for every row subset $I \subseteq \{1, \ldots, k\}$.*
(e) *The L_1-errors of the IPF sequence $A(t)$, $t \geq 0$, tend to zero, $\lim_{t \to \infty} f(A(t)) = 0$.*

Proof
(a) ⇒ (b). If the IPF sequence converges then its limit B is a biproportional scaling of A. It inherits matching row sums along odd steps and matching column sums along even steps. By Theorem 1, B is the unique biproportional fit.
(b) ⇒ (c). The biproportional fit clearly qualifies for a matrix D asked for in (c).
(c) ⇒ (d). As in the proof of Theorem 2, the definition of $J_A(I)$ yields $d_{I \times J_A(I)'} = 0 \leq d_{I' \times J_A(I)}$, and $r_I = d_{I \times J_A(I)} + d_{I \times J_A(I)'} \leq d_{I \times J_A(I)} + d_{I' \times J_A(I)} = s_{J_A(I)}$.
(d) ⇒ (e). Equal marginal totals entail $r_I - s_{J_A(I)} + s_{J_A(I)'} - r_{I'} = 2(r_I - s_{J_A(I)})$. With Lemma 2(c) we get $0 \leq \lim_{t \to \infty} f(A(t)) = 2 \max_{I \subseteq \{1, \ldots, k\}} (r_I - s_{J_A(I)}) \leq 0$.
(e) ⇒ (a). Since the entries of $A(t)$ are nonnegative and sum to r_+ in odd steps and to s_+ in even steps, the IPF sequence $A(t)$ stays in the compact set $[0, \max\{r_+, s_+\}]^{k \times \ell}$. Let B be an accumulation point along a subsequence $A(t_n)$, $n \geq 1$. From (e) we get $f(B) = \lim_{n \to \infty} f(A(t_n)) = \lim_{t \to \infty} f(A(t)) = 0$. With row and column sums fitted, B is a biproportional fit. By Theorem 1 there exists but one. Hence the IPF sequence $A(t)$, $t \geq 0$, has B for its unique accumulation point, and converges. □

If all weights are positive, $a_{ij} > 0$, and marginal totals coincide, $r_+ = s_+$, life simplifies dramatically. Statement (c) is fulfilled by the matrix D with entries $d_{ij} = r_i s_j / r_+ > 0$. Statement (d) holds true since non-empty row subsets I satisfy $J_A(I) = \{1, \ldots, \ell\}$ and hence $r_I \leq s_{J_A(I)} = r_+$. Either way the IPF sequence is seen to converge (Theorem 3), and the resulting biproportional fit transpires to be direct (Theorem 2).

The results disclose their power if some weight is zero. In the general case it is not easy to decide whether the IPF sequence converges or not. However, the equivalence of statements (c) and (d) is known as the Feasible Distribution Theorem of network theory, see Gale (1957, p. 1075) or Rockafellar (1984, p. 69). It is well-known that statement (d) may be quickly checked by means of the Max-flow Min-cut Theorem.

References

Bacharach, M. (1965). Estimating nonnegative matrices from marginal data. *International Economic Review (Osaka)*, *6*, 294–310.

Bacharach, M. (1970). *Biproportional matrices & input-output change*. Cambridge: Cambridge University Press.

Balinski, M. L., & Demange, G. (1989a). An axiomatic approach to proportionality between matrices. *Mathematics of Operations Research*, *14*, 700–719.

Balinski, M. L., & Demange, G. (1989b). Algorithms for proportional matrices in reals and integers. *Mathematical Programming*, *45*, 193–210.

Balinski, M. L., & Pukelsheim, F. (2006). Matrices and politics. In E. Liski, J. Isotalo, S. Puntanen, & G. P. H. Styan (Eds.), *Festschrift for Tarmo Pukkila on his 60th birthday* (pp. 233–242). University of Tampere: Department of Mathematics, Statistics, and Philosophy.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. Cambridge: MIT Press.

Bregman, L. M. (1967). Proof of the convergence of Sheleikhovskii's method for a problem with transportation constraints. *U.S.S.R. Computational Mathematics and Mathematical Physics*, *7*, 191–204.

Brown, D. T. (1959). A note on approximations to discrete probability distributions. *Information and Control*, *2*, 386–392.

Brown, J. B., Chase, P. J., & Pittenger, A. O. (1993). Order independence and factor convergence in iterative scaling. *Linear Algebra and Its Applications*, *190*, 1–38.

Brualdi, R. A., Parter, S. V., & Schneider, H. (1966). The diagonal equivalence of a nonnegative matrix to a stochastic matrix. *Journal of Mathematical Analysis and Applications*, *16*, 31–50.

Csiszár, I. (1975). *I*-Divergence geometry of probability distributions and minimization problems. *Annals of Probability*, *3*, 146–158.

Csiszár, I., & Tusnády, G. (1984). Information geometry and alternating minimization procedures. *Statistics & Decisions (Supplement Issue)*, *1*, 205–237.

Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, *11*, 427–444.

Fagan, J. T., & Greenberg, B. (1987). Making tables additive in the presence of zeros. *American Journal of Mathematical and Management Sciences*, *7*, 359–383.

Fienberg, S. E. (1970). An iterative procedure for estimation in contingency tables. *The Annals of Mathematical Statistics*, *41*, 907–917.

Fienberg, S. E., & Meyer, M. M. (2006). Iterative proportional fitting. In *In Encyclopedia of Statistical Sciences* (Vol. 6, pp. 3723–3726).

Gaffke, N., & Pukelsheim, F. (2008). Vector and matrix apportionment problems and separable convex integer optimization. *Mathematical Methods of Operations Research*, *67*, 133–159.

Gale, D. (1957). A theorem on flows in networks. *Pacific Journal of Mathematics*, *7*, 1073–1082.

Gietl, C. (2009). *Struktur der Häufungspunkte beim iterativen proportionalen Anpassungsverfahren*. Bachelor thesis, Institut für Mathematik, Universität Augsburg.

Gietl, C., & Reffel, F. P. (2013). Accumulation points of the iterative proportional fitting procedure. *Metrika*, *76*, 783–798.

Hershkowitz, D., Hoffman, A. J., & Schneider, H. (1997). On the existence of sequences and matrices with prescribed partial sums of elements. *Linear Algebra and Its Applications*, *265*, 71–92.

Ireland, C. T., & Kullback, S. (1968). Contingency tables with given marginals. *Biometrika*, *55*, 179–188.

Joas, B. (2005). *A Graph-theoretic Solvability Check for Biproportional Multiplier Methods*. Thesis, Institut für Mathematik, Universität Augsburg.

Kalantari, B., Lari, I., Ricca, F., & Simeone, B. (2008). On the complexity of general matrix scaling and entropy minimization via the RAS algorithm. *Mathematical Programming Series A*, *112*, 371–401.

Kullback, S. (1966). An information-theoretic derivation of certain limit relations for a stationary Markov chain. *SIAM Journal on Control*, *4*, 454–459.

Kullback, S. (1968). Probability densities with given marginals. *The Annals of Mathematical Statistics*, *39*, 1236–1243.

Macgill, S. M. (1977). Theoretical properties of biproportional matrix adjustments. *Environment & Planning A*, *9*, 687–701.

Maier, S. (2009). *Biproportional apportionment methods: constraints, algorithms, and simulation*. München: Dr. Hut-Verlag.

Maier, S., Zachariassen, P., & Zachariasen, M. (2010). Divisor-based biproportional apportionment in electoral systems: a real-life benchmark study. *Management Science*, *56*, 373–387.

Marshall, A. W., & Olkin, I. (1968). Scaling of matrices to achieve specified row and column sums. *Numerische Mathematik*, *12*, 83–90.

Oelbermann, K.-F., & Pukelsheim, F. (2011). Future European Parliament elections: ten steps towards uniform procedures. *Zeitschrift für Staats- und Europawissenschaften/Journal for Comparative Government and European Policy*, *9*, 9–28.

Pennisi, A. (2006). The Italian bug: a flawed procedure for bi-proportional seat allocation. In *Mathematics and democracy—recent advances in voting systems and collective choice*. Berlin: Springer. pp. 151–166

Pretzel, O. (1980). Convergence of the iterative scaling procedure for non-negative matrices. *Journal of the London Mathematical Society*, *21*, 379–384.

Pukelsheim, F. (1998). Rounding tables on my bicycle. *Chance*, *11*, 57–58.

Pukelsheim, F. (2004). BAZI—a Java program for proportional representation. *Oberwolfach Reports*, *1*, 735–737.

Pukelsheim, F. (2013). *Biproportional matrix scaling and the iterative proportional fitting procedure*. Preprint Nr. 7/2013, Institut für Mathematik, Universität Augsburg.

Pukelsheim, F., & Schuhmacher, C. (2004). Das neue Zürcher Zuteilungsverfahren für Parlamentswahlen. *Aktuelle Juristische Praxis (Pratique Juridique Actuelle)*, *13*, 505–522.

Pukelsheim, F., & Schuhmacher, C. (2011). Doppelproporz bei Parlamentswahlen – ein Rück- und Ausblick. *Aktuelle Juristische Praxis (Pratique Juridique Actuelle)*, *20*, 1581–1599.

Pukelsheim, F., Ricca, F., Simeone, B., Scozzari, A., & Serafini, P. (2012). Network flow methods for electoral systems. *Networks*, *59*, 73–88.

Ramírez, V., Pukelsheim, F., Palomares, A., & Martínez, J. (2008). The bi-proportional method applied to the Spanish Congress. *Mathematical and Computer Modelling*, *48*, 1461–1467.

Rockafellar, R. T. (1984). *Network, flows, and monotropic optimization*. New York: Wiley.

Rote, G., & Zachariasen, M. (2007). Matrix scaling by network flow. In *Proceedings in applied mathematics: Vol. 125. Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms*, pp. 848–854.

Rothblum, U. G., & Schneider, H. (1989). Scalings with matrices which have prespecified row sums and column sums via optimization. *Linear Algebra and Its Applications*, *114*, 737–764.

Rüschendorf, L. (1995). Convergence of the iterative proportional fitting procedure. *The Annals of Statistics*, *23*, 1160–1174.

Rüschendorf, L., & Thomsen, W. (1993). Note on the Schrödinger equation and *I*-projections. *Statistics & Probability Letters*, *17*, 369–375.

Rüschendorf, L., & Thomsen, W. (1997). Closedness of sum spaces and the generalized Schrödinger problem. *Theory of Probability and Its Applications*, *42*, 483–494.

Schneider, M. H. (1990). Matrix scaling, entropy minimization, and conjugate duality (II): the dual problem. *Mathematical Programming*, *48*, 103–124.

Simeone, B., & Pukelsheim, F. (2006). *Mathematics and democracy—recent advances in voting systems and collective choice*. Berlin: Springer.

Sinkhorn, R. (1964). A relationship between arbitrary positive matrices and doubly stochastic matrices. *The Annals of Mathematical Statistics*, *35*, 877–879.

Sinkhorn, R. (1966). A relationship between arbitrary positive matrices and stochastic matrices. *Canadian Journal of Mathematics*, *18*, 303–306.

Sinkhorn, R. (1967). Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, *74*, 403–405.

Sinkhorn, R. (1972). Continuous dependence on A in the $D_1 A D_2$ theorems. *Proceedings of the American Mathematical Society*, *32*, 395–398.

Sinkhorn, R. (1974). Diagonal equivalence to matrices with prescribed row and column sums, II. *Proceedings of the American Mathematical Society*, *45*, 195–198.

Sinkhorn, R., & Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, *21*, 343–348.

Speed, T. P. (2005). Iterative proportional fitting. In *Encyclopedia of Biostatistics* (Vol. 7, pp. 2646–2650).

Stephan, F. F. (1942). An iterative method of adjusting sample frequency tables when expected marginal totals are known. *The Annals of Mathematical Statistics*, *13*, 166–178.

Wainer, H. (1998). Visual revelations: rounding tables. *Chance*, *11*, 46–50.

Zachariassen, P., & Zachariasen, M. (2006). A comparison of electoral formulae for the Farœse Parliament. In *Mathematics and democracy—recent advances in voting systems and collective choice* (pp. 235–252). Berlin: Springer.