



Universität Augsburg

Institut für
Mathematik

Christoph Gietl, Fabian P. Reffel

**Continuity of f-Projections and Applications to the Iterative
Proportional Fitting Procedure**

Preprint Nr. 13/2013 — 10. Juni 2013

Institut für Mathematik, Universitätsstraße, D-86135 Augsburg

<http://www.math.uni-augsburg.de/>

Impressum:

Herausgeber:

Institut für Mathematik

Universität Augsburg

86135 Augsburg

<http://www.math.uni-augsburg.de/de/forschung/preprints.html>

ViSdP:

Christoph Gietl

Institut für Mathematik

Universität Augsburg

86135 Augsburg

Preprint: Sämtliche Rechte verbleiben den Autoren © 2013

Continuity of f-projections and applications to the iterative proportional fitting procedure^{*}

Christoph Gietl[†] Fabian P. Reffel[‡]

June 7, 2013

Abstract This paper proves continuity of f-projections and the continuous dependence of the limit matrix of the iterative proportional fitting procedure (IPF procedure) on the given matrix as well as the given marginals under certain regularity constraints. For discrete spaces, the concept of f-projections of finite measures on a compact and convex set is introduced and continuity of f-projections is proven. This result is applied to the IPF procedure. Given a nonnegative matrix as well as row and column marginals the IPF procedure generates a sequence of matrices, called the IPF sequence, by alternately fitting rows and columns to match their respective marginals. If the IPF sequence converges, the application of the previous result yields the continuous dependence of the limit matrix on the given matrix. By generalized convex programming and under some constraints, it is shown that the limit matrix of the IPF sequence continuously depends not only on the given matrix but also on the marginals.

Keywords Continuity of f-projections · Iterative proportional fitting · Limit matrix · Dependence on input matrix and marginals · f-divergence · f-projection · I-divergence · I-projection

Mathematics Subject Classification (2000) 68W40 · 62B10 · 62H17 · 90C25

^{*}This research has been supported by the German National Academic Foundation, by the Elite Network of Bavaria through its graduate program TopMath and by the TUM Graduate School through its thematic graduate center TopMath.

[†]C. Gietl, Institut für Mathematik, Universität Augsburg, 86135 Augsburg, Germany;
Tel.: +49-821-5982202, Fax: +49-821-5982280, christoph.gietl@math.uni-augsburg.de

[‡]F. P. Reffel, Institut für Mathematik, Universität Augsburg, 86135 Augsburg, Germany;
Tel.: +49-821-5982202, Fax: +49-821-5982280, fabian.reffel@math.uni-augsburg.de

1 Introduction

The concept of f-divergence and f-projection plays a key role in statistics. Especially I-divergence, a special case of f-divergences, is helpful for the analysis of the iterative proportional fitting procedure (IPF procedure). The procedure aims to solve the following problem: Given a nonnegative $k \times \ell$ matrix A and given positive marginals, find a nonnegative $k \times \ell$ matrix B which fulfills the given marginals and is biproportional to A . To this end, the IPF procedure generates a sequence of matrices, called the IPF sequence, by alternately fitting rows and columns to match their respective marginals. The procedure is in use in many disciplines for problems such as calculating maximum likelihood estimators in graphical log-affine models (Lauritzen [13, Chapter 4.3.1]), ranking webpages (Knight [11]) or determining passenger flows (McCord et al. [15]).

In many applications, the entries of the matrix A and the given marginals are observations and may vary or have errors. If the IPF sequence converges to some limit matrix, then it is of interest, whether this limit matrix continuously depends on the input variables, that is the matrix A and the given marginals. The dependence on the matrix A fits into the more general question of continuity of f-projections.

The concept of f-divergence and f-projection has been introduced independently by Csiszár [6] and by Ali and Silvey [1]. An extensive overview of its properties can be found in the book by Liese and Vajda [14]. Among the family of f-divergences, I-divergence turns out to be fruitful for the analysis of the IPF procedure as shown by Csiszár [7], who proves necessary and sufficient conditions for convergence of the IPF procedure for nonnegative matrices. If these conditions are not fulfilled, Gietl and Reffel [9] have shown that the even-step IPF subsequence and the odd-step IPF subsequence still converge. The IPF procedure has been known since 1937 (Kruithof [12]) and has become popular due to the work of Deming and Stephan [8].

Section 2 introduces f-divergence and f-projections of finite measures on discrete spaces. The essential properties of f-divergence and f-projections are developed, leading to a short proof of continuity of f-projections. Section 3 defines the IPF procedure in detail, quotes necessary and sufficient criteria for its convergence and categorizes its limit matrices into direct and limit biproportional fits. The limit matrix is characterized as an I-projection, a special case of f-projection, and the section concludes with the continuous dependence of the limit biproportional fit on the given matrix A . To deal with varying marginals, the problem of finding the IPF limit matrix is put in the class of generalized convex programs, which is properly introduced in Section 4. An important corollary by Rockafellar [17] on the continuous dependence of the optimal value on the parameters is quoted. By restricting to direct biproportional fits, this corollary is applied to the IPF procedure in order to obtain the continuous dependence not only on the matrix A but also on the given marginals. Section 5 concludes the results and gives perspectives for further research.

In the sequel, all indices i belong to the set $\{1, \dots, k\}$ whereas all indices j belong to the set $\{1, \dots, \ell\}$. A $+$ as a subscript indicates the summation over the index that would otherwise appear in its place. A set as a subscript denotes the summation over all entries belonging to that set, i. e. $r_I = \sum_{i \in I} r_i$. The symbol \ll expresses dominance in the sense of measure theory. Thus, for two vectors s and t the statement $s \ll t$ is equivalent to

$t_i = 0$ implying $s_i = 0$ for all i . The symbol \approx expresses equivalence of two measures. Thus, two vectors s and t are equivalent when they have the same zero entries. The notation carries over for nonnegative matrices.

2 Continuity of f-projections

The family of f-divergences contains various popular distances and distance-like functions between measures, e. g. I-divergence, χ^2 -divergence, total variance or Hellinger distance. Liese and Vajda [14] present an abundance of properties of f-divergence for probability measures on arbitrary measurable spaces. For finite spaces, we extend some of their results to nonnegative vectors, which can be interpreted as finite measures. This leads to the section's main result, the continuity of f-projections on appropriate compact and convex sets. The outline of this section is partly based on Section 2 of Gietl and Reffel [9].

Let $f : [0; \infty) \rightarrow \mathbb{R}$ be some convex and (in 0) continuous function. We then define the function $\mathfrak{f} : [0; \infty)^2 \rightarrow \mathbb{R} \cup \{\infty\}$ by

$$\mathfrak{f}(v, w) := \begin{cases} w \cdot f\left(\frac{v}{w}\right) & \text{if } v \in [0; \infty), w \in (0; \infty), \\ v \cdot \lim_{x \rightarrow \infty} \frac{f(x)}{x} & \text{if } v \in (0; \infty), w = 0, \\ 0 & \text{if } v = 0, w = 0. \end{cases} \quad (2.1)$$

By Proposition A.24 of Liese and Vajda [14] the limit $\lim_{x \rightarrow \infty} f(x)/x$ exists in $\mathbb{R} \cup \{\infty\}$. The function \mathfrak{f} is lower semicontinuous for all $(v, w) \in [0; \infty)^2$ (see Liese and Vajda [14, Proposition A.35]). For fixed $v \geq 0$, the function \mathfrak{f} is continuous in the second argument. For $v = 0$, the claim is trivial since $\mathfrak{f}(0, w) = w \cdot f(0)$. For $v > 0$, only continuity in $w = 0$ has to be checked, because f is continuous on $[0; \infty)$. It holds

$$\lim_{w \searrow 0} \mathfrak{f}(v, w) = \lim_{w \searrow 0} w \cdot f\left(\frac{v}{w}\right) = \lim_{w \searrow 0} v \cdot \frac{f\left(\frac{v}{w}\right)}{\frac{v}{w}} = v \cdot \lim_{x \rightarrow \infty} \frac{f(x)}{x} = \mathfrak{f}(v, 0). \quad (2.2)$$

Obviously, for all $w > 0$ the function $\mathfrak{f}(v, w)$ is strictly convex in the first argument, if and only if f is strictly convex.

For two vectors $s, t \in \mathbb{R}_{\geq 0}^k$ the *f-divergence of s relative to t* is defined by

$$D_f(s | t) := \sum_i \mathfrak{f}(s_i, t_i). \quad (2.3)$$

The definition of f-divergence carries over for nonnegative matrices. It holds $D_f(s | t) < \infty$ if $s \ll t$. In most of the pertinent literature, properties of f-divergence are proven for probability measures only (Liese and Vajda [14]). Clearly, our vectors s and t can be seen as finite measures. We continue to work with finite measures in the form of nonnegative vectors and nonnegative matrices. The following theorem summarizes the properties of f-divergence inherited from the function \mathfrak{f} and needed in the sequel.

Theorem 2.1 (Continuity and convexity of f-divergence). *Let D_f be the f-divergence as defined by equation (2.3). Then the following three statements hold:*

- (i) The function $(s, t) \mapsto D_f(s|t)$ is lower semicontinuous for all $s, t \in \mathbb{R}_{\geq 0}^k$.
- (ii) For fixed $s \in \mathbb{R}_{\geq 0}^k$, the function $t \mapsto D_f(s|t)$ is continuous on $\mathbb{R}_{\geq 0}^k$.
- (iii) For fixed $t \in \mathbb{R}_{\geq 0}^k$, $t \neq 0$, the function $s \mapsto D_f(s|t)$ is strictly convex for all $s \in \mathbb{R}_{\geq 0}^k$ with $s \ll t$, if and only if f is strictly convex.

Proof. (i),(ii) Follows directly from the discussion of the function f above.

(iii) The constraints in (iii) imply all summands of $D_f(s|t)$ being finite and at least one component of s being positive within the convex combinations. \square

Let $t \in \mathbb{R}_{\geq 0}^k$ and $\mathcal{M} \subseteq \mathbb{R}_{\geq 0}^k$ be a compact and convex set. An element $s^* \in \mathcal{M}$ is an *f-projection of t on \mathcal{M}* when

$$D_f(s^*|t) = \inf_{s \in \mathcal{M}} D_f(s|t). \quad (2.4)$$

Lemma 2.2 (Existence and uniqueness of f-projections). *Let $t \in \mathbb{R}_{\geq 0}^k$ and $\mathcal{M} \subseteq \mathbb{R}_{\geq 0}^k$ be a compact and convex set. Then the following three statements hold:*

- (i) There exists an f-projection of t on \mathcal{M} .
- (ii) If f is strictly convex and $s \ll t$ holds for all $s \in \mathcal{M}$, the f-projection of t on \mathcal{M} is unique.
- (iii) If f is strictly convex and $\lim_{x \rightarrow \infty} f(x)/x = \infty$ and $s \ll t$ for some $s \in \mathcal{M}$, the f-projection of t on \mathcal{M} is unique.

Statements (i) and (iii) can also be found in Vajda and van der Meulen [22, Theorem 1]. Furthermore, their Theorem 1 generalizes our statement (ii). Nevertheless, we give a short self-contained proof.

Proof. (i) As shown in Theorem 2.1 (i), f-divergence is lower semicontinuous in both arguments. Then f-divergence is also lower semicontinuous in the first argument when keeping the second argument constant. On compact sets, each lower semicontinuous function attains its minimum (Bauschke and Combettes [3, Theorem 1.28]).

(ii) If f is strictly convex and $s \ll t$ holds for all $s \in \mathcal{M}$, the strict convexity shown in Theorem 2.1 (iii) together with the convexity of \mathcal{M} yields uniqueness.

(iii) With $\lim_{x \rightarrow \infty} f(x)/x = \infty$, it holds $f(s_i, 0) < \infty$ if and only if $s_i = 0$. Hence, the f-projection on \mathcal{M} coincides with the f-projection on $\{s \in \mathcal{M} \mid s \ll t\} \neq \emptyset$. Thus, the claim follows from statement (ii). \square

Example 2.3. We cannot expect uniqueness of f-projections if the function f is not strictly convex but only convex. Let the function f be given by $f(x) = |x - 1|$. This f-divergence is better known as the total variance, since

$$D_f(s|t) = \sum_i f(s_i, t_i) = \sum_i |s_i - t_i| =: \|s - t\|_1. \quad (2.5)$$

Let $t = (1, 1) \in \mathbb{R}_{\geq 0}^2$ and the convex and compact set \mathcal{M} be defined by

$$\mathcal{M} := \left\{ m \in \mathbb{R}_{\geq 0}^2 \mid m_+ = 1 \right\}. \quad (2.6)$$

It holds $m \in [0; 1]^2$ for all $m \in \mathcal{M}$, hence we have

$$\|m - t\|_1 = |m_1 - t_1| + |m_2 - t_2| = t_1 - m_1 + t_2 - m_2 = t_+ - m_+ = 1. \quad (2.7)$$

for all $m \in \mathcal{M}$. Thus, all elements of \mathcal{M} are f-projections of t on \mathcal{M} . \diamond

For all vectors $s \in \mathbb{R}_{\geq 0}^k$ we define the *support of s* by $\text{supp}(s) := \{i \mid s_i > 0\}$. Thus, the condition $s \ll t$ for some $t \in \mathbb{R}_{\geq 0}^k$ and all $s \in \mathcal{M} \subseteq \mathbb{R}_{\geq 0}^k$ will simply be written as $\mathcal{M} \subseteq \mathbb{R}_{\geq 0}^{\text{supp}(t)} := \left\{ s \in \mathbb{R}_{\geq 0}^k \mid \text{supp}(s) \subseteq \text{supp}(t) \right\}$. Now we are able to present this section's main result: f-projections on appropriate compact and convex sets are continuous.

Theorem 2.4 (Continuity of f-projections). *Let the f-divergence D_f be defined by a function f that is strictly convex on $[0; \infty)$. Let (t^n) , $t^n \in \mathbb{R}_{\geq 0}^k$, be a sequence converging to some $t \in \mathbb{R}_{\geq 0}^k$. Let $\mathcal{M} \subseteq \mathbb{R}_{\geq 0}^k$ be a compact and convex set. Then the following three statements hold:*

- (i) *If $\mathcal{M} \subseteq \mathbb{R}_{\geq 0}^{\text{supp}(t)}$, then for n large enough, the f-projection of t^n on \mathcal{M} and the f-projection of t on \mathcal{M} exist and are unique.*
- (ii) *If $\lim_{x \rightarrow \infty} f(x)/x = \infty$ and $\mathcal{M} \cap \mathbb{R}_{\geq 0}^{\text{supp}(t)} \neq \emptyset$, then for n large enough, the f-projection of t^n on \mathcal{M} and the f-projection of t on \mathcal{M} exist and are unique.*
- (iii) *Under the conditions of statement (i) or statement (ii), the f-projections $s^{n,*}$ of t^n on \mathcal{M} converge to the f-projection s^* of t on \mathcal{M} ,*

$$s^{n,*} \xrightarrow{n \rightarrow \infty} s^*. \quad (2.8)$$

Vajda and van der Meulen [22, Theorem 6] have proven a statement similar to statement (iii), which is restricted to f-divergences with $\lim_{x \rightarrow \infty} f(x)/x < \infty$ and sequences (t_n) along the line segment connecting the limit t and the uniform distribution.

Proof. (i) By Lemma 2.2 (i) the f-projections of t^n on \mathcal{M} and the f-projection of t on \mathcal{M} always exist. Since $t, t^n \in \mathbb{R}_{\geq 0}^k$ and $t^n \xrightarrow{n \rightarrow \infty} t$, there exists a constant $N \in \mathbb{N}$ such that $t \ll t^n$ for all $n \geq N$. From $\mathcal{M} \subseteq \mathbb{R}_{\geq 0}^{\text{supp}(t)}$ we conclude $\mathcal{M} \subseteq \mathbb{R}_{\geq 0}^{\text{supp}(t^n)}$ for all $n \geq N$. The strict convexity of the function f and application of Lemma 2.2 (ii) yield uniqueness of s^* and $s^{n,*}$ for $n \geq N$.

(ii) By Lemma 2.2 (i) the f-projections of t^n on \mathcal{M} and the f-projection of t on \mathcal{M} always exist. Since $t, t^n \in \mathbb{R}_{\geq 0}^k$ and $t^n \xrightarrow{n \rightarrow \infty} t$, there exists a constant $N \in \mathbb{N}$ such that $t \ll t^n$ for all $n \geq N$. From $\mathcal{M} \cap \mathbb{R}_{\geq 0}^{\text{supp}(t)} \neq \emptyset$ we conclude the existence of some $s \in \mathcal{M}$ such that $s \ll t$ and $s \ll t^n$ for all $n \geq N$. The strict convexity of the function f and application of Lemma 2.2 (iii) yield uniqueness of s^* and $s^{n,*}$ for $n \geq N$.

(iii) Compactness of \mathcal{M} guarantees the existence of a convergent subsequence $(s^{\eta_n, *})$ with limit $s^{**} \in \mathcal{M}$, say. Lower semicontinuity of f-divergence (Theorem 2.1 (i)), minimality of the f-projection (2.4) and continuity of f-divergence in the second argument (Theorem 2.1 (ii)) yield

$$D_f(s^{**} | t) = D_f\left(\lim_{n \rightarrow \infty} s^{\eta_n, *} \mid \lim_{n \rightarrow \infty} t^{\eta_n}\right) \quad (2.9)$$

$$\leq \liminf_{n \rightarrow \infty} D_f(s^{\eta_n, *} | t^{\eta_n}) \quad (2.10)$$

$$\leq \liminf_{n \rightarrow \infty} D_f(s^* | t^{\eta_n}) \quad (2.11)$$

$$= D_f(s^* | t). \quad (2.12)$$

Since s^* is the unique f-projection of t on \mathcal{M} , we conclude $s^{**} = s^*$. Thus, all accumulation points of the sequence $(s^{n, *})$ coincide and the sequence $(s^{n, *})$ converges to s^* . \square

3 Limit fitting problems with fixed marginals

We specify the IPF procedure in full detail. Thereafter, we quote necessary and sufficient criteria for its convergence and categorize its limit matrices B^* into direct and limit biproportional fits. Our notation sticks to Pukelsheim [16] as close as possible. After that, we introduce I-divergence as a member of the family of f-divergences and quote a statement by Csiszár [7] regarding the equivalence between limit fits and I-projections. We then use this equivalence statement to prove continuous dependence of the limit matrix B^* on the input matrix A in the case of fixed marginals c and r .

The *IPF procedure* takes as input an arbitrary nonnegative matrix $A = ((a_{ij})) \in \mathbb{R}_{\geq 0}^{k \times \ell}$ with positive row sums, $a_{i+} > 0$, and positive column sums, $a_{+j} > 0$, and two vectors with positive entries $r = (r_1, \dots, r_k) \in \mathbb{R}_{> 0}^k$ and $c = (c_1, \dots, c_\ell) \in \mathbb{R}_{> 0}^\ell$. The matrix A is referred to as the *input matrix*, whereas the vector r is called the *row marginals* and the vector c is called the *column marginals*.

The procedure is initialized by setting $A(0) := A$. Subsequently, the *IPF sequence* $(A(t))$ is calculated by iteratively repeating the following two steps:

- Odd steps $t + 1$ fit row sums to row marginals. To this end, all entries in the same row are multiplied by the same multiplier yielding

$$a_{ij}(t + 1) := \frac{r_i}{a_{i+}(t)} \cdot a_{ij}(t) \text{ for all entries } (i, j). \quad (3.1)$$

- Even steps $t + 2$ fit column sums to column marginals. To this end, all entries in the same column are multiplied by the same multiplier yielding

$$a_{ij}(t + 2) := \frac{c_j}{a_{+j}(t + 1)} \cdot a_{ij}(t + 1) \text{ for all entries } (i, j). \quad (3.2)$$

By induction, for all steps $t \geq 0$ the inequality $a_{ij}(t) > 0$ holds if and only if $a_{ij} > 0$ holds. Consequently, all row sums $a_{i+}(t)$ and all column sums $a_{+j}(t)$ always stay positive. Thus, the IPF procedure is well defined.

We say that the *IPF procedure converges*, when the IPF sequence $(A(t))$ converges. This does not depend on the exact entries of the input matrix A . Instead, it is sufficient to know which entries are positive and which entries equal 0. Knowing this, the set $J_A(I) := \{j \in \{1, \dots, \ell\} \mid \exists i \in I : a_{ij} > 0\}$ of all columns connected to the row subset I in A can be computed for all row subsets $I \subseteq \{1, \dots, k\}$. The input (A, c, r) can then be checked according to the following theorem.

Theorem 3.1 (Convergence of the IPF procedure). *Let A be an input matrix and let c and r be positive marginals. Then the IPF procedure converges, if and only if $c_+ = r_+$ holds and all row subsets $I \subseteq \{1, \dots, k\}$ fulfill the flow inequalities*

$$r_I \leq c_{J_A(I)}. \quad (3.3)$$

Proof. See Hershkowitz, Hoffman and Schneider [10, Theorem 3.2] or Pukelsheim [16, Theorem 5]. \square

In the case of convergence, the limit matrix $B^* := \lim_{t \rightarrow \infty} A(t)$ has three decisive properties. First, it fulfills column marginals, i. e. $b_{+j}^* = c_j$ holds for all columns j . Second, it fulfills row marginals, i. e. $b_{i+}^* = r_i$ holds for all rows i . Third, B^* is biproportional to A , i. e. there exist sequences of multipliers $\rho_i(t) > 0$ and $\sigma_j(t) > 0$ such that $b_{ij} = \lim_{t \rightarrow \infty} \rho_i(t) a_{ij} \sigma_j(t)$ holds for all entries (i, j) . These sequences can be computed as products of the row and column multipliers used in equations (3.1) and (3.2). We call each matrix matching those three properties a *limit biproportional fit of A to c and r* . Pukelsheim [16, Theorem 1], however, shows, that limit biproportional fits are unique. Moreover, if there exists a limit biproportional fit, the inequalities (3.3) hold. Hence, the inequalities (3.3) are a necessary and sufficient condition for the existence of the limit biproportional fit.

When the sequences $(\rho_i(t))$ and $(\sigma_j(t))$ can be replaced by constant multipliers $\rho_i > 0$ and $\sigma_j > 0$, the limit biproportional fit B^* is called the *direct biproportional fit of A to c and r* . Directness of B^* can be checked more easily if the input matrix A is connected. We call a nonnegative matrix C *connected* when it is not disconnected. A nonnegative matrix D is *disconnected* when there exists a permutation of rows and a permutation of columns such that D acquires block format,

$$D = \begin{array}{cc} & \begin{array}{cc} J & J' \end{array} \\ \begin{array}{c} I \\ I' \end{array} & \left(\begin{array}{cc} D^{(1)} & 0 \\ 0 & D^{(2)} \end{array} \right), \end{array} \quad (3.4)$$

where at least one of the subsets $I \subseteq \{1, \dots, k\}$ or $J \subseteq \{1, \dots, \ell\}$ is nonempty and proper. Here, the prime indicates the complement of a set. If the input matrix A is disconnected, it can be decomposed into several connected matrices and the marginals c and r can be decomposed accordingly. Running the IPF procedure on the resulting decomposed inputs yields the same results as running the procedure on the original input (A, c, r) . However, for the decomposed inputs, directness can be checked according to the following theorem.

Theorem 3.2 (Directness of biproportional fits). *Let A be a connected input matrix, let c and r be positive marginals and let B^* be the limit biproportional fit of A to c and r . Then B^* is a direct biproportional fit, if and only if the flow inequalities (3.3) are strict for all nonempty and proper row subsets I , $\emptyset \subsetneq I \subsetneq \{1, \dots, k\}$.*

Proof. See Rothblum and Schneider [18, Theorem 2] or Pukelsheim [16, Theorem 2]. \square

The input (A, c, r) of the IPF procedure as defined above is called a *fitting problem*. When there exists a limit biproportional fit, we call (A, c, r) a *limit fitting problem*. When there exists a direct biproportional fit and A is connected, we call (A, c, r) a *direct fitting problem*.

In the analysis of fitting problems, the f-divergence derived from the strictly convex and continuous function $f : [0, \infty) \rightarrow \mathbb{R}, x \mapsto x \log x$ with $f(0) := 0$, which fulfills $\lim_{x \rightarrow \infty} f(x)/x = \infty$, is prevalent (see Csiszár [7], Brown, Chase and Pittenger [4], Rüschemdorf [19] and Cramer [5]). We call this f-divergence the *information divergence* (*I-divergence*) and denote it by $I(\cdot | \cdot)$. The corresponding f-projections will be called *I-projections*. In the pertinent literature, I-divergence is also known as Kullback-Leibler divergence, relative entropy or information gain.

Let \mathcal{M} denote the set of all nonnegative matrices B fulfilling the marginals c and r ,

$$\mathcal{M} := \left\{ B = ((b_{ij})) \in \mathbb{R}_{\geq 0}^{k \times \ell} \mid \forall i : b_{i+} = r_i \text{ and } \forall j : b_{+j} = c_j \right\}. \quad (3.5)$$

This set \mathcal{M} is compact and convex. I-projections on \mathcal{M} turn out to be equivalent to biproportional fits to c and r .

Theorem 3.3 (Equivalence between biproportional fits and I-projections). *Let A be an input matrix, c and r be positive marginals and let the set \mathcal{M} be defined according to equation (3.5). Then B^* is the limit biproportional fit of A to c and r if and only if B^* is the unique I-projection of A to \mathcal{M} .*

Proof. See Csiszár [7, Theorem 3.2]. \square

Remark 3.4. Theorem 3.3 does not hold for f-projections in general. Let the direct fitting problem (A, c, r) be defined by

$$A = \begin{pmatrix} 1 & \frac{1}{2} \\ 1 & 1 \end{pmatrix}, \quad c = (1, 2) \text{ and } r = (2, 1). \quad (3.6)$$

Then, the biproportional fit B^* of A to c and r is given by

$$B^* = \begin{pmatrix} 3 - \sqrt{5} & \sqrt{5} - 1 \\ \sqrt{5} - 2 & 3 - \sqrt{5} \end{pmatrix}. \quad (3.7)$$

Taking the convex function $f : x \mapsto |x - 1|$ as in Example 2.3, f-divergence coincides with total variation $\|\cdot\|_1$. It holds

$$\|B^* - A\|_1 = 2\sqrt{5} - \frac{5}{2} \approx 1.972. \quad (3.8)$$

Yet, the matrix B^* is not the f-projection of A on \mathcal{M} since

$$D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathcal{M} \text{ and } \|D - A\|_1 = \frac{3}{2} < \|B^* - A\|_1. \quad (3.9)$$

When taking the strictly convex function $f : x \mapsto (x - 1)^2$, which also fulfills the equation $\lim_{x \rightarrow \infty} f(x)/x = \infty$, f-divergence D_f coincides with the χ^2 -divergence, since $f(v, w) = (v - w)^2/w$. It holds

$$D_f(B^* | A) = \frac{93}{2} - 20\sqrt{5} \approx 1.779 \quad (3.10)$$

Yet, the matrix B^* is not an f-projection of A on \mathcal{M} since

$$D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathcal{M} \text{ and } D_f(D | A) = \frac{3}{2} < D_f(B^* | A). \quad (3.11)$$

◇

Due to Theorem 3.3 we can now apply Theorem 2.4 to the IPF procedure and state our main result regarding limit fitting problems with fixed marginals.

Theorem 3.5 (Limit fitting problems with fixed marginals). *Let the triple (A, c, r) be a limit fitting problem, (A^n) a sequence of input matrices with $A^n \xrightarrow{n \rightarrow \infty} A$. Moreover, let B^* be the limit matrix of the IPF procedure applied to (A, c, r) . Then the following two statements hold:*

- (i) *For n large enough, the triple (A^n, c, r) is a limit fitting problem with the limit biproportional fit $B^{n,*}$.*
- (ii) *The limit biproportional fits $B^{n,*}$ of A^n to c and r converge to B^* ,*

$$B^{n,*} \xrightarrow{n \rightarrow \infty} B^*. \quad (3.12)$$

Proof. (i) Since the marginals c and r are fixed, convergence of the IPF procedure solely depends on the support of the input matrix A^n as stated in Theorem 3.1. Moreover, as the sequence (A^n) converges to A , we have $A \ll A^n$ for n large enough. This entails $c_{J_A(I)} \leq c_{J_{A^n}(I)}$ for all $I \subseteq \{1, \dots, k\}$ and for n large enough. By Theorem 3.1, (A, c, r) being a limit fitting problem then implies the triple (A^n, c, r) for n large enough being a limit fitting problem as well.

(ii) By Theorem 3.3, the limit biproportional fit $B^{n,*}$ of A^n to c and r is the I-projection of A^n on the compact and convex set $\mathcal{M} \subseteq \mathbb{R}_{\geq 0}^{k \times \ell}$ defined according to equation (3.5). The same holds true for the input matrix A . The limit biproportional fit B^* of A to c and r preserves the zero entries of A . Therefore, $\mathcal{M} \cap \mathbb{R}_{\geq 0}^{\text{supp}(A)} \neq \emptyset$ holds. Furthermore, for I-divergence it holds $\lim_{x \rightarrow \infty} f(x)/x = \infty$. Hence, application of Theorem 2.4 (iii) establishes the claim. □

Remark 3.6. For limit fitting problems (A, c, r) with square input matrices $A \in \mathbb{R}_{\geq 0}^{k \times k}$ and uniform marginals $c = r = (1, \dots, 1) \in \mathbb{R}_{> 0}^k$, Sinkhorn [21, Corollary 3] has already shown that the limit matrix B^* of the IPF procedure continuously depends on the input matrix A . However, his proof is based on results about doubly stochastic matrices and their permanents. For this reason, we cannot see how it could be generalized.

Example 3.7. The assumption of Theorem 3.5 that (A, c, r) is a limit fitting problem is essential. We present an example of a fitting problem for which the IPF procedure does not converge. This example can be verified easily. Let the sequence (A^n) be defined by

$$A^n = \begin{pmatrix} 1 & n^{-1} \\ 1 & 1 \end{pmatrix} \text{ for all } n \in \mathbb{N} \quad (3.13)$$

with fixed column marginals $c = (1, 2)$ and row marginals $r = (2, 1)$. Applying the IPF procedure to $(\lim_{n \rightarrow \infty} A^n, c, r)$ yields the two accumulation points (Gietl and Reffel [9, Theorem 5.3])

$$C^* = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{and} \quad R^* = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}. \quad (3.14)$$

The bottom left entries of C^* and R^* have to be 0, since the multipliers $r_2/a_{2+}(t)$ and $c_1/a_{+1}(t+1)$ used in scaling steps (3.1) and (3.2) are bounded from above by $1/2$.

When applying the IPF procedure to (A^n, c, r) , the IPF sequence converges to the I-projection

$$B^{n,*} = \begin{pmatrix} 2 - b^n & b^n \\ b^n - 1 & 2 - b^n \end{pmatrix}, \quad (3.15)$$

where

$$b^n = \frac{n - 4 + \sqrt{n^2 + 8n}}{2(n - 1)}. \quad (3.16)$$

This arises from the minimization of $I(B | A^n)$ over all $B \in \mathcal{M}$ in which the set \mathcal{M} has only one degree of freedom. It holds $b^n \xrightarrow{n \rightarrow \infty} 1$ and therefore

$$B^{n,*} \xrightarrow{n \rightarrow \infty} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (3.17)$$

Hence, neither $B^{n,*} \xrightarrow{n \rightarrow \infty} C^*$ nor $B^{n,*} \xrightarrow{n \rightarrow \infty} R^*$ holds. Another example is found in Balinski and Demange [2, Example 2]. Further examples for doubly stochastic matrices are found within the remarks of Sinkhorn [20]. \diamond

Varying the marginals c and r cannot be allowed in Theorem 3.5, since the set \mathcal{M} depends on c and r and, thus, the inequality $I(B^{\eta n,*} | A^{\eta n}) \leq I(B^* | A^{\eta n})$ cannot be guaranteed (see equations (2.10) and (2.11)).

4 Direct fitting problems with varying marginals

In order to examine the continuous dependence of the IPF limit matrix not only on the input matrix A , but also on the marginals c and r , a new concept is introduced. The equivalence between biproportional fits and I-projections (see Theorem 3.3) allows us to interpret each fitting problem (A, c, r) as a convex minimization problem where the objective function and the constraints depend on the parameters A , c and r . Therefore, we take a closer look at these so-called *generalized convex programs*, which are examined by Rockafellar [17, Section 29] using the notion of bifunctions. When restricting to direct fitting problems, this allows us to prove the continuous dependence of the limit matrix of the IPF procedure on the input matrix as well as the given marginals.

A *bifunction* from \mathbb{R}^p to \mathbb{R}^q is a mapping F assigning to each vector $u \in \mathbb{R}^p$ a function $Fu : \mathbb{R}^q \rightarrow [-\infty, +\infty]$. This function Fu maps each vector $x \in \mathbb{R}^q$ to a number $(Fu)(x)$. The bifunction F is called *convex*, when the function $(u, x) \mapsto (Fu)(x)$ is convex. The *effective domain* of F , $\text{dom } F$, is the set of all $u \in \mathbb{R}^p$, for which the function Fu is not the constant function ∞ , hence

$$\text{dom } F := \{u \in \mathbb{R}^p \mid \exists x \in \mathbb{R}^q : (Fu)(x) < \infty\}. \quad (4.1)$$

The function $(\inf F) : \mathbb{R}^p \rightarrow [-\infty, +\infty]$, $u \mapsto \inf_{x \in \mathbb{R}^q} (Fu)(x)$ mapping the parameters u to the optimal value $\inf_{x \in \mathbb{R}^q} (Fu)(x)$ is called the *perturbation function*. The operator int denotes the *interior* of a set. For an arbitrary set \mathcal{U} , we define

$$\text{int } \mathcal{U} := \{u \in \mathcal{U} \mid \exists \epsilon > 0 : u + B(0, \epsilon) \subseteq \mathcal{U}\}, \quad (4.2)$$

where $B(0, \epsilon) := \{x \mid \|x\| \leq \epsilon\}$. Rockafellar [17] examines continuity of the perturbation function $\inf F$.

Theorem 4.1 (Continuity of the perturbation function). *Let F be a convex bifunction from \mathbb{R}^p to \mathbb{R}^q with $0 \in \text{int}(\text{dom } F) \subseteq \mathbb{R}^p$. Then there exists an open convex neighborhood $\mathcal{U}(0)$ of $0 \in \mathbb{R}^p$, such that the perturbation function $(\inf F) : \mathbb{R}^p \rightarrow [-\infty, +\infty]$ is continuous on $\mathcal{U}(0)$.*

Proof. See Rockafellar [17, Corollary 29.1.5]. □

In order to apply Theorem 4.1, we restrict ourselves to sequences $((A^n, c^n, r^n))$ converging to a direct fitting problem (A, c, r) when n tends to ∞ . Furthermore, we assume $A \approx A^n$ and $c_+^n = r_+^n$ for all $n \in \mathbb{N}$. Hence, we lose one degree of freedom and vary only c^n and r_1^n, \dots, r_{k-1}^n while setting $r_k^n := c_+^n - (r_1^n + \dots + r_{k-1}^n)$. This section's main result then reads as follows.

Theorem 4.2 (Direct fitting problems with varying marginals). *Let (A, c, r) be a direct fitting problem, $((A^n, c^n, r^n))$ a sequence of fitting problems with $(A^n, c^n, r^n) \xrightarrow{n \rightarrow \infty} (A, c, r)$, $A^n \approx A$ and $c_+^n = r_+^n$ for all $n \in \mathbb{N}$. Let B^* be the limit matrix of the IPF procedure applied to (A, c, r) . Then the following two statements hold:*

(i) For n large enough, the triple (A^n, c^n, r^n) is a direct fitting problem with the limit matrix $B^{n,*}$.

(ii) The direct biproportional fits $B^{n,*}$ of A^n to c^n and r^n converge to B^* ,

$$B^{n,*} \xrightarrow{n \rightarrow \infty} B^*. \quad (4.3)$$

Proof. We proceed in six steps. Steps I to IV establish the applicability of Theorem 4.1. Statement (i) is proven within step V. Finally, step VI shows statement (ii).

I. We define m to be the minimal difference of the flow inequalities (3.3),

$$m := \min_{\emptyset \neq I \subsetneq \{1, \dots, k\}} \{c_{J_A(I)} - r_I\}. \quad (4.4)$$

By Theorem 3.2, m is positive.

II. Let $u = (s, v, w) \in \mathbb{R}^{\text{supp}(A) + (k-1) + \ell}$ and define the bifunction F from $\mathbb{R}^{\text{supp}(A) + (k-1) + \ell}$ to $\mathbb{R}^{k \times \ell}$ by

$$(u, x) \mapsto I(x | A + s) + \delta(x | S(u)), \quad (4.5)$$

where

$$\delta(x | S(u)) := \begin{cases} 0 & \text{if } x \in S(u), \\ \infty & \text{otherwise} \end{cases} \quad (4.6)$$

and

$$S(u) := \left\{ x \in \mathbb{R}_{\geq 0}^{k \times \ell} \mid x_{i+} = r_i + v_i \ \forall i \in \{1, \dots, k-1\}, \right. \quad (4.7)$$

$$\left. x_{+j} = c_j + w_j \ \forall j \in \{1, \dots, \ell\} \right\}. \quad (4.8)$$

The notion of bifunctions requires the function I-divergence $I(\cdot | \cdot)$ to be defined not only for the nonnegative matrices. Therefore, we set $I(x | A + s) = \infty$ whenever one of the two arguments has a negative entry. The set $S(u)$ depends only on the components v, w of the vector u . However, the notation $S(v, w)$ is a bit bulky for proving convexity. In the following, we show that F is convex (step III) and that $0 \in \text{int}(\text{dom } F)$ holds (step IV).

III. I-divergence is convex as a function of both arguments on the convex set $\mathbb{R}_{\geq 0}^{k \times \ell} \times \mathbb{R}_{\geq 0}^{k \times \ell}$ (Gietl and Reffel [9, Theorem 2.1 (i)]). Setting I-divergence to ∞ for negative entries does not change the function's epigraph and preserves convexity. Hence, the function F is convex, if $\delta(x | S(u))$ is a convex function in x and u . For all $u^1, u^2 \in \mathbb{R}^{\text{supp}(A) + (k-1) + \ell}$, all $x^1, x^2 \in \mathbb{R}^{k \times \ell}$ and all $\lambda \in (0; 1)$ we show

$$\delta(\lambda x^1 + (1 - \lambda)x^2 | S(\lambda u^1 + (1 - \lambda)u^2)) \leq \lambda \delta(x^1 | S(u^1)) + (1 - \lambda) \delta(x^2 | S(u^2)). \quad (4.9)$$

If $x^1 \notin S(u^1)$ or $x^2 \notin S(u^2)$ the inequality (4.9) is trivial. Thus, let $x^1 \in S(u^1)$ and $x^2 \in S(u^2)$. Then it holds

$$\lambda x_{i+}^1 + (1 - \lambda)x_{i+}^2 = \lambda(r_i + v_i^1) + (1 - \lambda)(r_i + v_i^2) \quad (4.10)$$

$$= r_i + (\lambda v_i^1 + (1 - \lambda)v_i^2) \quad (4.11)$$

for all $i \in \{1, \dots, k-1\}$ and

$$\lambda x_{+j}^1 + (1-\lambda)x_{+j}^2 = \lambda(c_j + w_j^1) + (1-\lambda)(c_j + w_j^2) \quad (4.12)$$

$$= c_j + (\lambda w_j^1 + (1-\lambda)w_j^2) \quad (4.13)$$

for all $j \in \{1, \dots, \ell\}$. Thus, $\lambda x^1 + (1-\lambda)x^2 \in S(\lambda u^1 + (1-\lambda)u^2)$ and the inequality (4.9) is fulfilled. Hence, the bifunction F is convex.

IV. It holds

$$\text{dom } F = \left\{ u \in \mathbb{R}^{\text{supp}(A)+(k-1)+\ell} \mid \exists x \in \mathbb{R}^{k \times \ell} : \text{I}(x \mid A + s) + \delta(x \mid S(u)) < \infty \right\} \quad (4.14)$$

$$= \left\{ u \in \mathbb{R}^{\text{supp}(A)+(k-1)+\ell} \mid \exists x \in S(u) : \text{I}(x \mid A + s) < \infty \right\}. \quad (4.15)$$

Restricting s to the set

$$\left(- \min_{(i,j) \in \text{supp}(A)} a_{ij}/2; \infty \right)^{\text{supp}(A)} \quad (4.16)$$

yields $\text{supp}(A + s) = \text{supp}(A)$. For such s we show that there exists a matrix $x \in S(u)$ with $\text{I}(x \mid A + s) < \infty$, if $|v_i| \leq m/(4(k-1)k)$ for all $i \in \{1, \dots, k-1\}$ and $|w_j| \leq m/(4k\ell)$ for all $j \in \{1, \dots, \ell\}$, where m is defined by equation (4.4). Defining

$$v_k := w_+ - \sum_{i=1}^{k-1} v_i \quad (4.17)$$

we have

$$|v_k| \leq |w_+| + \sum_{i=1}^{k-1} |v_i| \leq \ell \frac{m}{4k\ell} + (k-1) \frac{m}{4(k-1)k} = \frac{m}{2k}. \quad (4.18)$$

Consequently, it holds $|v_i| \leq m/(2k)$ for all $i \in \{1, \dots, k\}$ and $|w_j| \leq m/(2\ell)$ for all $j \in \{1, \dots, \ell\}$. From the definition (4.4) of m the inequality

$$v_I - w_{J_A(I)} \leq (k-1) \frac{m}{2k} + \ell \frac{m}{2\ell} = m - \frac{m}{2k} < c_{J_A(I)} - r_I \quad (4.19)$$

follows for all $I, \emptyset \neq I \subsetneq \{1, \dots, k\}$. The strict inequality (4.19) is equivalent to

$$r_I + v_I < c_{J_A(I)} + w_{J_A(I)} \quad (4.20)$$

for all $I, \emptyset \neq I \subsetneq \{1, \dots, k\}$. Thus, the marginals $r_i + v_i$, $i = 1, \dots, k$, and $c_j + w_j$, $j = 1, \dots, \ell$ fulfill the flow inequalities (3.3). By Theorem 3.1 the fitting problem $(A + s, r + v, c + w)$ is a limit fitting problem. With Theorem 3.3 there exists a nonnegative matrix x that fulfills the marginals, $x \in S(u)$, and is dominated by $A + s$, $x \ll A + s$. Therefore, it holds $\text{I}(x \mid A + s) < \infty$. Hence, we have

$$\text{dom } F = \left\{ u \in \mathbb{R}^{\text{supp}(A)+(k-1)+\ell} \mid \exists x \in S(u) : \text{I}(x \mid A + s) < \infty \right\} \quad (4.21)$$

$$\supseteq \left(- \min_{(i,j) \in \text{supp}(A)} a_{ij}/2; \infty \right)^{\text{supp}(A)} \quad (4.22)$$

$$\times \left[-\frac{m}{4(k-1)k}; \frac{m}{4(k-1)k} \right]^{k-1} \times \left[-\frac{m}{4k\ell}; \frac{m}{4k\ell} \right]^\ell. \quad (4.23)$$

It holds $0 \in \text{int}(\text{dom } F)$. Together with step III, the assumptions of Theorem 4.1 are fulfilled.

V. We apply Theorem 4.1 to the bifunction F . Consequently, the perturbation function

$$u \mapsto \inf_{x \in \mathbb{R}^{k \times \ell}} I(x | A + s) + \delta(x | S(u)) = \inf_{x \in S(u)} I(x | A + s) \quad (4.24)$$

is continuous in u on a neighborhood of 0. We define the sequence $(u^n) := ((s^n, v^n, w^n))$ by $s_{ij}^n := a_{ij}^n - a_{ij}$ for all $(i, j) \in \text{supp}(A)$ and $v_i^n := r_i^n - r_i$ for all $i \in \{1, \dots, k-1\}$ as well as $w^n := c^n - c$. Pursuant to step IV the problem (A^n, c^n, r^n) fulfills the flow inequalities (4.20) when n is large enough. By Theorem 3.2, the problems (A^n, c^n, r^n) are direct fitting problems when n is large enough. This proves statement (i). We define $B^{n,*}$ to be the limit matrix of the direct fitting problem (A^n, c^n, r^n) . Due to Theorem 3.3 it holds

$$\inf_{x \in S(u^n)} I(x | A + s^n) = I(B^{n,*} | A^n). \quad (4.25)$$

For n being large enough, the sequence (u^n) is in a neighborhood of 0. Therefore, we conclude

$$I(B^{n,*} | A^n) \xrightarrow{n \rightarrow \infty} I(B^* | A). \quad (4.26)$$

VI. This part is similar to part (iii) of Theorem 2.4 and therefore a bit skimmed. Because of $c_+^n \xrightarrow{n \rightarrow \infty} c_+ < \infty$, there exists an integer $N \in \mathbb{N}$ and a constant K such that $c_+^n \leq K$ for all $n \geq N$. Thus, the sequence $(B^{n,*})$ stays within a compact set, $B^{n,*} \in [0; K]^{k \times \ell}$. Hence, there exists a convergent subsequence (η_n) such that $B^{\eta_n,*} \xrightarrow{n \rightarrow \infty} B^{**} \in \mathcal{M}$. Applying Theorem 3.3, statement (4.26) and the lower semicontinuity of I-divergence (Theorem 2.1 (i)) we get

$$\inf_{B \in \mathcal{M}} I(B | A) = I(B^* | A) = \lim_{n \rightarrow \infty} I(B^{\eta_n,*} | A^{\eta_n}) \geq I(B^{**} | A). \quad (4.27)$$

Consequently, it holds $I(B^* | A) = I(B^{**} | A)$. The uniqueness of the I-projection (Lemma 2.2 (iii)) yields $B^* = B^{**}$. Thus, all accumulation points of the sequence $(B^{n,*})$ coincide and the sequence $(B^{n,*})$ converges to B^* . \square

Remark 4.3. Let (A, c, r) be a limit fitting problem with a positive input matrix $A \in \mathbb{R}_{>0}^{k \times \ell}$ and limit matrix B^* . If the sequence $((A^n, c^n, r^n))$ with $A^n \in \mathbb{R}_{>0}^{k \times \ell}$ and $c_+^n = r_+^n$ is converging to (A, c, r) , then Balinski and Demange [2, Corollary to Theorem 3] have shown that the limit matrices $B^{n,*}$ of the fitting problems (A^n, c^n, r^n) converge to B^* . Since $A^n, A \in \mathbb{R}_{>0}^{k \times \ell}$ imply strict flow inequalities (3.3), the triples (A, c, r) and (A^n, c^n, r^n) are direct fitting problems for all $n \in \mathbb{N}$. Hence, this case is included in Theorem 4.2.

When relaxing the restriction $A \in \mathbb{R}_{>0}^{k \times \ell}$ to $A \in \mathbb{R}_{\geq 0}^{k \times \ell}$ in the setting above, Balinski and Demange [2, Theorem 3] have proven that the limit matrices $B^{n,*}$ still converges to some $\tilde{B}^* \ll A$ fulfilling the given marginals. However, they have left open whether \tilde{B}^* equals the limit matrix of the limit fitting problem (A, c, r) . \diamond

5 Conclusion and perspectives

In this paper, we have proven that unique f-projections as defined by Liese and Vajda [14] are continuous (Theorem 2.4). Applying the special case of I-projections to the analysis of the IPF procedure, we have shown that for limit fitting problems (A, c, r) the limit matrix continuously depends on the input matrix A (Theorem 3.5). When restricting to direct fitting problems (A, c, r) , we were able to allow variation in the input matrix as well as in the given marginals c and r . Applying Rockafellar's [17] results on generalized convex programs we have proven that the limit matrix continuously depends on the input (A, c, r) (Theorem 4.2).

However, several questions remain open. Can the assumption $A \approx A^n$ for all $n \in \mathbb{N}$ in Theorem 4.2 be omitted? And is the assumption $c_+^n = r_+^n$ necessary? In the more general case of limit fitting problems, what happens when varying the marginals, that is $(A, c^n, r^n) \xrightarrow{n \rightarrow \infty} (A, c, r)$? In the even more general case, when the fitting problem (A, c, r) is not a limit fitting problem, the even-step IPF subsequence still converges (see Gietl and Reffel [9, Theorem 5.3]). In this case, what happens with the sequence of even-step limit points $(C^{n,*})$ of the fitting problem (A, c^n, r^n) when (A, c^n, r^n) is converging to a non limit fitting problem (A, c, r) ? Numerical examples suggest that in any above mentioned cases it holds $B^{*,n} \xrightarrow{n \rightarrow \infty} B^*$ respectively $C^{*,n} \xrightarrow{n \rightarrow \infty} C^*$. However, we do not see how our proof extends in that direction, since either the set $S(u)$ is not convex anymore or 0 is not in the interior of the domain of the bifunction.

Acknowledgements. *We are very grateful to our advisor Friedrich Pukelsheim for continuing support. Moreover, we would like to thank Friedrich Liese for helpful remarks on an earlier version of this paper.*

References

- [1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society. Series B (Methodological)*, 28:131–142, 1966.
- [2] M. L. Balinski and G. Demange. An axiomatic approach to proportionality between matrices. *Mathematics of Operations Research*, 14:700–719, 1989.
- [3] H. Bauschke and P. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, New York, 2011.
- [4] J. B. Brown, P. J. Chase, and A. O. Pittenger. Order independence and factor convergence in iterative scaling. *Linear Algebra and its Applications*, 190:1–38, 1993.
- [5] E. Cramer. Probability measures with given marginals and conditionals: I-projections and conditional iterative proportional fitting. *Statistics & Decisions*, 18:311–329, 2000.

- [6] I. Csiszár. Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, 8:85–108, 1963.
- [7] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3:146–158, 1975.
- [8] W. E. Deming and F. F. Stephan. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11:427–444, 1940.
- [9] C. Gietl and F. P. Reffel. Accumulation points of the iterative proportional fitting procedure. *Metrika*, 2012. doi:10.1007/s00184-012-0415-7.
- [10] D. Hershkowitz, A. J. Hoffman, and H. Schneider. On the existence of sequences and matrices with prescribed partial sums of elements. *Linear Algebra and its Applications*, 265:71–92, 1997.
- [11] P. A. Knight. The Sinkhorn–Knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, 30:261–275, 2008.
- [12] J. Kruithof. Telefoonverkeersrekening. *De Ingenieur*, 52:E15–E25, 1937.
- [13] S. L. Lauritzen. *Graphical models*. Clarendon Press, Oxford, 1996.
- [14] F. Liese and I. Vajda. *Convex Statistical Distances*. Teubner, Leipzig, 1987.
- [15] M. R. McCord, R. G. Mishalani, P. Goel, and B. Strohl. Iterative proportional fitting procedure to determine bus route passenger origin–destination flows. *Transportation Research Record: Journal of the Transportation Research Board*, 2145:59–65, 2010.
- [16] F. Pukelsheim. Biproportional matrix scaling and the iterative proportional fitting procedure. Preprint, Institut für Mathematik, Universität Augsburg, 2013. URL: <http://opus.bibliothek.uni-augsburg.de/opus4/frontdoor/index/index/docId/2306>.
- [17] R. T. Rockafellar. *Convex analysis*. Princeton University Press, 1972.
- [18] U. G. Rothblum and H. Schneider. Scalings of matrices which have prespecified row sums and column sums via optimization. *Linear Algebra and its Applications*, 114:737–764, 1989.
- [19] L. Rüschendorf. Convergence of the iterative proportional fitting procedure. *Annals of Statistics*, 23:1160–1174, 1995.
- [20] R. Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *Annals of Mathematical Statistics*, 35:876–879, 1964.
- [21] R. Sinkhorn. Continuous dependence on A in the D_1AD_2 theorems. In *Proceedings of the American Mathematical Society*, volume 32, pages 395–398, 1972.

- [22] I. Vajda and E. C. van der Meulen. On minimum divergence adaptation of discrete bivariate distributions to given marginals. *IEEE Transactions on Information Theory*, 51:313–320, 2005.