Symmetric Iterative Proportional Fitting
— Supplementary Material —

This supplement consists of several parts that refer directly to specific topics in the paper:

A - Proof of Equation (2)

B - Proof of Lemma 3.1 (Symmetric biproportional fit)

C - Technical details on why "local affinity" is sufficient in Section 4.1

D - Proof of Theorem 4.2 (Convergence of PSIPF)

E - Proof of Lemma 4.4 (L_1-monotony)

F - Proof of Lemma 4.5 (Volume bounds)

G - Proof of Lemma 4.6 (Limit points)

H - Proof of Lemma 4.7 (Strong convergence)

I - Proof (sketch) of Proposition 5.2 (Strictly positive feasibility)

A Proof of Equation (2)

Equation (2) claims that for all $A, W \in \Omega$ with $E(A) \subseteq E(W)$, $\mathbf{r} \in \mathbb{R}^m_{>0}$, and $\mathcal{R} := \Omega(\mathbf{r}, \cdot, W)$, it holds that $\mathcal{P}_\mathcal{R}(A) = \operatorname{diag}(\mathbf{r}) \cdot \operatorname{diag}(A\mathbf{1})^{-1} \cdot A$. In order to keep notation simple, we first prove the following weaker statement (by assuming $A = W$) and finally generalize the result to any A with $E(A) \subseteq E(W)$

For $W \in \Omega$, $\mathbf{r} \in \mathbb{R}^m_{>0}$ and $\mathcal{R} := \Omega(\mathbf{r}, \cdot, W) = \{X \in \mathbb{R}^{m\times n}_{\geq 0} \mid X\mathbf{1} = \mathbf{r},\ E(X) \subseteq E(W)\}$ it holds that

$$\mathcal{P}_\mathcal{R}(W) \;=\; \operatorname{diag}(\mathbf{r})\operatorname{diag}(W\mathbf{1})^{-1} \cdot W$$

Proof. From $w_{ij} = 0 \Rightarrow x_{ij} = 0$ and $0 \cdot \log(0/w_{ij}) := 0$ it follows by definition of RE that

$$\varrho(X) \;:=\; RE(X\|W) \;=\; \sum_{\substack{i=1\dots m,\\ j=1\dots n}} x_{ij}\log(x_{ij}/w_{ij}) - x_{ij} + w_{ij} \;=\; \sum_{ij\in E(W)} x_{ij}\log(x_{ij}/w_{ij}) - x_{ij} + w_{ij}$$

From the definitions we get that

$$\begin{aligned}\mathcal{P}_\mathcal{R}(W) \;&=\; \operatorname{arg\,min}_{X\in\mathbb{R}^{m\times n}_{\geq 0}} \{\ \varrho(X) \quad | \quad X\mathbf{1} = \mathbf{r},\ E(X) \subseteq E(W)\ \} \\ &=\; \operatorname{arg\,min}_{X\in\mathbb{R}^{m\times n}_{\geq 0}} \Big\{ \textstyle\sum_{ij\in E(W)} x_{ij}\log(x_{ij}/w_{ij}) - x_{ij} \quad | \quad X\mathbf{1} = \mathbf{r},\ E(X) \subseteq E(W) \Big\}.\end{aligned} \qquad (\star)$$

It follows from strict convexity of ϱ and convexity of the constraints that the minimizer is unique, if existing (※).

Comment: Note that ($\star$) describes a non-smooth optimization problem because the partial derivative "$\nabla_{X_{ij}}\rho(X) = \log(x_{ij}/w_{ij})$", does not exist whenever $ij \in E(W) \setminus E(X)$. A general strategy to avoid this problem is by first determining $E(\mathcal{P}_\mathcal{R}(W))$ separately and then skipping all summands with $ij \notin E(\mathcal{P}_\mathcal{R}(W))$ from ($\star$), before finally taking the Lagrangian approach. For the projection to $\mathcal{R}$ we show in the following that $E(\mathcal{P}_\mathcal{R}(W)) = E(W)$.

We re-state ($\star$) as the minimization over $2^{|E(W)|}$ many individual minimization problems, one for each $E \subseteq E(W)$, and each providing a unique (if existing) "inner minimizer" $R^*(E)$ that achieves its respective minimum $\varrho^*(E)$:

$$\mathcal{P}_{\mathcal{R}}(W) \;=\; R^*(E^*) \qquad \text{for} \qquad E^* \;:=\; \underset{E \subseteq E(W)}{\arg\min}\, \{\varrho^*(E)\}$$

$$\text{and} \qquad \varrho^*(E) := \begin{cases} \min\limits_{X \in \mathbb{R}^{m\times n}_{\geq 0}} \{\; \varrho(X) \mid X\mathbf{1} = \mathbf{r},\; E(X) = E \;\} & \text{with unique minimizer } R^*(E) \\ \infty & \text{if not existing} \end{cases}.$$

If $\mathcal{P}_{\mathcal{R}}(W)$ exists, then E^* is well-defined because at least one inner minimizer exists: for $E = E(\mathcal{P}_{\mathcal{R}}(W))$ there trivially exists $R^*(E) = \mathcal{P}_{\mathcal{R}}(W)$. We now show that further E^* is maximal in the sense that $E(\mathcal{P}_{\mathcal{R}}(W)) \supseteq E$ for all $E \subseteq E(W)$ for which $R^*(E)$ exists.

For any $E_1, E_2 \subseteq E(W)$ for which $R^*(E_1)$ and $R^*(E_2)$ exist let $R_\alpha := \alpha R^*(E_1) + (1-\alpha)R^*(E_2)$ denote their convex combination for $\alpha \in (0,1)$. Then $E(R_\alpha) = E_1 \cup E_2 =: E_{12}$. In particular $E_{12} \neq \emptyset$, thus $R^*(E_{12})$ exists. By continuity of ϱ we get that $\lim_{\alpha\to 0} \varrho(R_\alpha) = \varrho(R^*(E_1)) = \varrho^*(E_1)$ and similarly that $\lim_{\alpha\to 1} \varrho(R_\alpha) = \varrho(R^*(E_2)) = \varrho^*(E_2)$. Since further $\varrho^*(E_{12}) \leq \varrho(R_\alpha)$ we get that $\varrho^*(E_{12}) \leq \min\{\varrho^*(E_1), \varrho^*(E_2)\}$. Thus ϱ^* cannot increase when considering the union of any two sets. This implies that with $\xi := \{E \subseteq E(W) \mid \varrho^*(E) < \infty\}$ and $E^* := \bigcup_{E\in\xi} E$ that $\varrho^*(E^*) \leq \varrho^*(E)$ for all $E \subseteq E(W)$. Hence $R^*(E^*)$ must be *some* global minimizer. It is further unique over all subsets $E \subseteq E(W)$ by the separate uniqueness argument (※) above. Hence $\mathcal{P}_{\mathcal{R}}(W) = R^*(E^*)$.

Here, we even get that $E^* = E(W)$ because there exists the feasible solution

$$W^* := \operatorname{diag}(\mathbf{r})\operatorname{diag}(W\mathbf{1})^{-1} \cdot W$$

of maximum possible non-zeros $E(W^*) = E(W)$. It remains to show that $\mathcal{P}_{\mathcal{R}}(W)$ indeed equals W^*.

With the knowledge that $E(\mathcal{P}_{\mathcal{R}}(W)) = E(W)$ we can re-state the minimization problem ($\star$) as

$$\mathcal{P}_{\mathcal{R}}(X) \quad = \quad \arg\min_{X \in \mathbb{R}^{m\times n}_{\geq 0}} \Big\{ \textstyle\sum_{ij\in E(W)} x_{ij} \log(x_{ij}/w_{ij}) - x_{ij} \quad \Big| \quad X\mathbf{1} = \mathbf{r},\; E(X) = E(W) \Big\}\;.$$

Now we can take the standard Lagrangian approach, that is to find a global minimum of the Lagrangian function

$$\Lambda(X, \boldsymbol{\mu}) = \varrho(X) - \boldsymbol{\mu}^T(X\mathbf{1} - \mathbf{r}) = \sum_{ij\in E(W)} x_{ij}\log(x_{ij}/w_{ij}) - x_{ij} - \sum_{i=1}^{m} \mu_i \left(\Big(\sum_{j \text{ s.t. } ij\in E(W)} x_{ij}\Big) - r_i \right)$$

over $X \in \mathbb{R}^{E(W)}_{>0}$, by which we handle $x_{ij} := 0$ for all $ij \notin E(W)$ as constants. From $w_{ij} \neq 0 \Leftrightarrow x_{ij} \neq 0$ we get that $\nabla_{x_{ij}} \Lambda(X, \boldsymbol{\mu}) = \log(x_{ij}/w_{ij}) - \mu_i$ for all $ij \in E(W)$. Setting all derivatives to zero gives after entry-wise exponentiation that

$$x_{ij} = \begin{cases} \exp(\mu_i) \cdot w_{ij} & \text{for } ij \in E(W) \\ 0 & \text{for } ij \notin E(W) \end{cases},$$

that is in matrix notation $X = \operatorname{diag}(\exp(\boldsymbol{\mu})) \cdot W$. Multiplying $\mathbf{1}$ from the right gives that $\mathbf{r} = X\mathbf{1} = \operatorname{diag}(\exp(\boldsymbol{\mu})) \cdot W\mathbf{1}$, hence $\operatorname{diag}(\exp(\boldsymbol{\mu})) = \operatorname{diag}(\mathbf{r})\operatorname{diag}(W\mathbf{1})^{-1}$. Thus $X = W^*$ is indeed the (only) critical point. Further, the Hessian matrix is the diagonal matrix with the all-positive diagonal $\nabla_{x_{ij}}\nabla_{x_{ij}}\Lambda(X, \boldsymbol{\mu}) = 1/x_{ij}$, which is positive definite. Thus W^* is the (unique global) minimum.

□

The proof easily generalizes to $\mathcal{P}_{\mathcal{R}}(A)$ for any $A \in \Omega$ with $E(A) = E(W)$ just by flipping the notation from $W = [w_{ij}]$ to $A = [a_{ij}]$ whenever W appears as the argument of $\mathcal{P}_{\mathcal{R}}(\cdot)$. Further, it generalizes to any $A \in \Omega$ with $E(A) \subseteq E(W)$ by the convention $x_{ij} \cdot \log(x_{ij}/0) := \infty$ for $x_{ij} > 0$, which immediately implies that $E(\mathcal{P}_{\mathcal{R}}(A)) \subseteq E(A)$, hence by maximality $E(\mathcal{P}_{\mathcal{R}}(A)) = E(A)$.

The proof for $\mathcal{P}_{\mathcal{C}}(W)$ is fully analogous.

B Proof of Lemma 3.1

Lemma 3.1 (Symmetric biproportional fits). *Let B denote the biproportional fit of $W \in \mathcal{S}$ to row and column marginals $\mathbf{f} \in \mathbb{R}^n_{>0}$. Then*
(i) $B = B^T$ is symmetric
(ii) $B = \lim_{k\to\infty} W_k$ for a sequence of $W_k \in \Psi(W)$
(iii) $B \in \Psi(W)$ if and only if B is direct

Proof. (i) Let $B = \lim_{k\to\infty} R^{(k)}WS^{(k)}$ denote the biproportional fit of W to $\mathbf{f}$. Then $B = \lim_{k\to\infty} R^{(k)}W^TS^{(k)}$ is the same biproportional fit of $W^T = W$ to $\mathbf{f}$. Hence $B^T = \lim_{k\to\infty} S^{(k)}WR^{(k)}$ is another biproportional fit of W to $\mathbf{f}$, so by uniqueness, $B^T = B$.

(ii) Now define $T^{(k)} := \sqrt{R^{(k)}S^{(k)}}$ by its diagonal elements $t_i := (r_i^{(k)}s_i^{(k)})^{1/2}$. For all i, j with $w_{ij} = w_{ji} \neq 0$, we get from $\lim_{k\to\infty} w_{ij}\ r_i^{(k)}s_j^{(k)} = b_{ij} = b_{ji} = \lim_{k\to\infty} w_{ij}\ r_j^{(k)}s_i^{(k)}$ that $\lim_{k\to\infty}(r_i^{(k)}s_j^{(k)})^{1/2} = \lim_{k\to\infty}(r_j^{(k)}s_i^{(k)})^{1/2}$. Hence,

$$\lim_{k\to\infty} w_{ij}\ t_i^{(k)}t_j^{(k)} = \lim_{k\to\infty} w_{ij}\ (r_i^{(k)}s_j^{(k)}r_j^{(k)}s_i^{(k)})^{1/2} = w_{ij}\cdot(\lim_{k\to\infty}\sqrt{r_i^{(k)}s_j^{(k)}})^2 = \lim_{k\to\infty} w_{ij}\ r_i^{(k)}s_j^{(k)} = b_{ij}.$$

For $b_{ij} \neq 0$ this implies with $b_{ij} \neq 0 \Rightarrow w_{ij} \neq 0$ that $\lim_{k\to\infty} w_{ij}\ t_i^{(k)}t_j^{(k)} = b_{ij}$. If $b_{ij} = 0$, then either $w_{ij} \neq 0$, thus again $\lim_{k\to\infty} w_{ij}\ t_i^{(k)}t_j^{(k)} = b_{ij}$, or $w_{ij} = 0$, in which case $\lim_{k\to\infty} w_{ij}\ t_i^{(k)}t_j^{(k)} = \lim_{k\to\infty} 0t_i^{(k)}t_j^{(k)} = 0 = b_{ij}$.

Thus, $\lim_{k\to\infty} T^{(k)}WT^{(k)} = B$, which proves (ii) for $W_k := T^{(k)}WT^{(k)} \in \Psi(W)$.

(iii) The implication "$\Rightarrow$" follows from the definitions, so it remains to show "$\Leftarrow$". In case of direct $B = RWS$, we get from (ii) with $T^{(k)} := T := \sqrt{RS}$ that $B = TWT \in \Psi(W)$. □

C Technical details on why "local affinity" is sufficient in Section 4.1

In this section we refer to the recursion (4) as the non-reflected sequence $(\tilde{x}_k)$ defined for some $\tilde{x}_0 \in \mathbb{R}^N$ by

$$\tilde{x}_k \quad := \quad \mathcal{P}_{[k]}(\tilde{x}_{k-1}) \quad ,$$

and to the recursion (5) as the sequence (x_k) defined for $x_0 = \tilde{x}_0$ by

$$\begin{aligned} x_k \quad &:= \quad (\mathcal{P}^h_{[k]} \circ \nabla h^*)(\nabla h(x_{k-1}) + r_{k-\ell}) \\ \text{and} \quad r_k \quad &:= \quad \nabla h(x_{k-1}) + r_{k-\ell} - \nabla h(x_k) \quad . \end{aligned}$$

We say that a function $h : \mathbb{R}^N \to \mathbb{R}$ is "suitable", if it is Legendre, co-finite, and very strictly convex. Please refer to B&L (Bauschke and Lewis, 2000) for detailed definitions. Beside others, orthogonal projections and *RE*-projections are induced by a suitable function h as a Bregman divergence, so all results hold for both of them.

In the following, $\langle\cdot,\cdot\rangle$ denotes an inner product on $\mathbb{R}^N$. For any $A \subseteq \mathbb{R}^N$ let $\mathcal{B}_A(x_0, \epsilon) := \{x \in A \mid \|x - x_0\|_2 \leq \epsilon\}$ denote the ball of radius ϵ around x_0 in A. For any subset $\mathcal{C}$ of an affine subspace $\mathcal{A}$ we denote $c \in \mathcal{C}$ as being *locally affine* if there exists $\epsilon > 0$ such that $\mathcal{B}_A(c, \epsilon) \subseteq \mathcal{C}$. The sequence (4) is locally affine if all its points for $k \geq 1$ are locally affine.

All that we need, and which is implied by local affinity, is sloppy spoken that there "exists a basis of $\mathcal{A}$ and its negate within a small neighborhood in $\mathcal{C}$ around the point c". Let us give some intuition on what this means and why this can be helpful: if for a linear subspace $\mathcal{L} \subseteq \mathbb{R}^N$ and some $q \in \mathbb{R}^N$ it holds that $\langle x, q\rangle \leq 0$ for all $x \in \mathcal{L}$ then it follows that $\langle x, q\rangle = 0$. This is easy to see from $\langle b, q\rangle \leq 0$ and $\langle -b, q\rangle \leq 0$ for every basis vector b of $\mathcal{L}$. It is already sufficient if the inequality holds everywhere within an arbitrary small ball around $\mathbf{0}$, since it already contains a smaller-scaled basis of $\mathcal{L}$. This insight is stated in the following lemma more general by taking $\mathcal{L}$ as the difference space of an affine subspace $\mathcal{A} \subseteq \mathbb{R}^N$.

Lemma C.1. *Let $\mathcal{A} \subseteq \mathbb{R}^N$ denote an affine subspace. Then it holds for all $q \in \mathbb{R}^N$ and $a \in \mathcal{A}$:*

$$\forall x \in \mathcal{A} \ : \ \langle x - a, q \rangle = 0 \qquad \Leftrightarrow \qquad \exists \epsilon > 0 : \forall x \in \mathcal{B}_{\mathcal{A}}(a, \epsilon) \ : \ \langle x - a, q \rangle \leq 0.$$

The same holds true for $\geq$ instead of $\leq$.

Proof. The direction "$\Rightarrow$" is trivial, so we only have to prove "$\Leftarrow$": let $\mathcal{L}$ denote the linear subspace that underlies $\mathcal{A}$, thus, $\forall x \in \mathcal{A} : a - x \in \mathcal{L}$. Let $b_1, \ldots, b_c$ denote a basis of $\mathcal{L}$. With $b_i' := \epsilon \cdot b_i / (\max_i \{\|b_i\|\})$ for $i = 1, \ldots, c$ we get that $b_1', \ldots, b_c'$ is also a basis of $\mathcal{L}$, and that $a \pm b_i' \in \mathcal{B}_{\mathcal{A}}(a, \epsilon)$. Thus, $\langle b_i', q \rangle \leq 0$ and $\langle -b_i', q \rangle \leq 0$, which implies that $\langle b_i', q \rangle = 0$ for all i. Thus, we get for all $x \in \mathcal{A}$ from the representation $x = a + \sum_{i=1}^{c} \alpha_i b_i'$ for some $\alpha_1, \ldots, \alpha_c \in \mathbb{R}$ that $\langle x - a, q \rangle = \sum_{i=1}^{c} \alpha_i \langle b_i', q \rangle = 0$. □

Note that this lemma does *not* state that $q \perp x$ for $x \in \mathcal{A}$. The orthogonality refers to the difference space, that is $q \perp (x - a) \in \mathcal{L}$. Lemma C.1 is the tool that we need in order to generalize B&L (Theorem 4.3) to local affinity, as formulated in the following theorem.

Theorem C.2 (Convergence of locally affine sequences). *Let $\mathcal{F} = \mathcal{C}_1 \cap \ldots \cap \mathcal{C}_\ell$ be the non-empty intersection of ℓ closed convex sets $\mathcal{C}_i \subseteq \mathbb{R}^N$, where each $\mathcal{C}_i$ is a subset of an affine subspace $\mathcal{A}_i \subseteq \mathbb{R}^N$. For $\tilde{x}_0 \in \mathbb{R}^N$ consider the recursion*

$$\tilde{x}_k := \mathcal{P}^h_{[k]}(\tilde{x}_{k-1})$$

for $k \geq 1$. If the sequence $(\tilde{x}_0, \tilde{x}_1, \ldots)$ is locally affine, then $\lim_{k\to\infty} \tilde{x}_k = P^h_{\mathcal{F}}(\tilde{x}_0)$.

Proof. The proof shows by induction that $x_k = \tilde{x}_k$ for all $k \geq 0$. Then $\lim_{k\to\infty} \tilde{x}_k = P^h_{\mathcal{F}}(\tilde{x}_0)$ follows from the fact that it is also the limit of (x_k). For all $k \leq \ell$ it holds that $x_k = \tilde{x}_k$ because $r_{k-\ell} = 0$ by definition, and $\nabla h^* \circ \nabla h = \mathrm{id}$. Now fix $k > \ell$ with by induction $x_i = \tilde{x}_i$ for all $i < k$. Note that $\mathcal{C}_{[k]} = \mathcal{C}_{[k-\ell]}$ and $\mathcal{A}_{[k]} = \mathcal{A}_{[k-\ell]}$. Now we incorporate our local affinity assumption. Hence there exists some $\epsilon_{k-\ell} > 0$ such that $\mathcal{B}_{\mathcal{A}_{[k]}}(\tilde{x}_{k-\ell}, \epsilon_{k-\ell}) \subseteq \mathcal{C}_{[k]}$. Because of $x_{k-\ell} = \tilde{x}_{k-\ell}$ we have that $\mathcal{B}_{\mathcal{A}_{[k]}}(x_{k-\ell}, \epsilon_{k-\ell}) \subseteq \mathcal{C}_{[k]}$. From B&L (Equation 1 in Theorem 3.2) we get that for all $c \in \mathcal{C}_{[k]}$ it holds that $\langle c - x_{k-\ell}, r_{k-\ell} \rangle \leq 0$. So we can apply Lemma C.1 in order to get that $\langle a - x_{k-\ell}, r_k \rangle = 0$ for all $a \in \mathcal{A}_{[k]}$. For the underlying difference space $\mathcal{L}_{[k]} = \mathcal{A}_{[k]} - x_{k-\ell}$ we are free to choose any other element $z \in \mathcal{A}_{[k]}$ for its representation, that is $\mathcal{A}_{[k]} - x_{k-\ell} = \mathcal{A}_{[k]} - z$. Choosing $z = x_k$ gives that $\langle a - x_k, r_k \rangle = 0$ for all $a \in \mathcal{A}_{[k]} \supseteq \mathcal{C}_{[k]}$. The remaining part follows the ideas of the original proof, but we also fix a small flaw in it. We have that $x_{k-1} = \tilde{x}_{k-1}$, and we claim that

$$\tilde{x}_k = \mathcal{P}^h_{[k]}(\tilde{x}_{k-1}) \quad \overset{!}{=} \quad (\mathcal{P}^h_{[k]} \circ \nabla h^*)(\nabla h(x_{k-1}) + r_{k-\ell}) = x_k. \tag{$\star$}$$

Applying the characterization

$$a' = \mathcal{P}^h_{[k]}(a) \quad \Leftrightarrow \quad \forall c \in \mathcal{C}_{[k]} \ : \ \langle c - a', \nabla h(a) - \nabla h(a') \rangle \leq 0 \tag{※}$$

for $a := \nabla h^*(\nabla h(x_{k-1}) + r_{k-\ell})$ and corresponding[1] $a' := x_k$ in the direction of "$\Rightarrow$" to the right side in ($\star$) gives:

$$\begin{aligned}
0 &\geq \langle c - x_k, \nabla h(\nabla h^*(\nabla h(x_{k-1}) + r_{k-\ell})) - \nabla h(x_k) \rangle & & \\
&= \langle c - x_k, \nabla h(x_{k-1}) + r_{k-\ell} - \nabla h(x_k) \rangle & & \nabla h \circ \nabla h^* = \mathrm{id} \\
&= \langle c - x_k, \nabla h(\tilde{x}_{k-1}) - \nabla h(x_k) \rangle + \langle c - x_k, r_{k-\ell} \rangle & & x_{k-1} = \tilde{x}_{k-1} \\
&= \langle c - x_k, \nabla h(\tilde{x}_{k-1}) - \nabla h(x_k) \rangle & & \langle c - x_k, r_{k-\ell} \rangle = 0
\end{aligned}$$

From (※) again, now for $a := \tilde{x}_{k-1}$ and $a' = x_k$ in direction "$\Leftarrow$", we get that $x_k = \mathcal{P}^h_{[k]}(\tilde{x}_{k-1})$, hence $\tilde{x}_k = x_k$.

Thus, we can simply drop all reflection terms without affecting the sequence *at all*, nor its limit. Hence in this case the sequence generated by (5) coincides with (4). □

Note that this theorem can be formulated in more generality by allowing to drop the reflection terms for individual $\mathcal{C}_i$'s: for any $i \in \{1, \ldots, \ell\}$ such that x_k is locally affine for all $[k] = i$, we can drop all reflection terms r_k for $[k] = i$, while keeping all other required reflection terms. This does not affect the sequence at all.

[1] here is a small flaw in B&L: they apply a together with $a' = \tilde{x}_k$ and later refer to $\langle c - \tilde{x}_k, r_{k-\ell} \rangle = 0$. However, it must be $a' = x_k$ and the orthogonality must be shifted to $\langle c - x_k, r_{k-\ell} \rangle = 0$.

D Proof of Theorem 4.2

Theorem 4.2 (Convergence of PSIPF). *Let $W \in \mathcal{S}$ and $\mathbf{f} \in \mathbb{R}^n_{>0}$ such that $\mathcal{S}(\mathbf{f}, W) \neq \emptyset$. Then the PSIPF-sequence $(W^{\langle\langle k\rangle\rangle})$ converges to $\mathcal{P}_{\mathcal{S}(\mathbf{f},W)}(W)$. Further, $W^{\langle\langle k\rangle\rangle} \in \Psi(W)$ for all $k \geq 0$.*

Proof. Let $B = \lim_{k\to\infty} W^{\langle k\rangle} = \lim_{k\to\infty} Y^{\langle k\rangle} W Z^{\langle k\rangle}$ denote the limit of the IPF-sequence. Lemma 3.1 gives that $B = B^T$. We have to show that $\lim_{k\to\infty} w_{ij}^{\langle\langle k\rangle\rangle} = b_{ij}$ for all i, j. Whenever $w_{ij} = w_{ji} \neq 0$, we get from $\lim_{k\to\infty} w_{ij}\, y_i^{\langle k\rangle} z_j^{\langle k\rangle} = b_{ij} = b_{ji} = \lim_{k\to\infty} w_{ij}\, y_j^{\langle k\rangle} z_i^{\langle k\rangle}$ that $\lim_{k\to\infty}(y_i^{\langle k\rangle} z_j^{\langle k\rangle})^{1/2} = \lim_{k\to\infty}(y_j^{\langle k\rangle} z_i^{\langle k\rangle})^{1/2}$. Hence $\lim_{k\to\infty} w_{ij}^{\langle\langle k\rangle\rangle} = \lim_{k\to\infty} w_{ij}(y_i^{\langle k\rangle} z_j^{\langle k\rangle} y_j^{\langle k\rangle} z_i^{\langle k\rangle})^{1/2} = \lim_{k\to\infty} w_{ij} y_i^{\langle k\rangle} z_j^{\langle k\rangle} = b_{ij}$. For $b_{ij} \neq 0$ this implies with $b_{ij} \neq 0 \Rightarrow w_{ij} \neq 0$ that $\lim_{k\to\infty} w_{ij}^{\langle\langle k\rangle\rangle} = b_{ij}$. Otherwise (i.e., $b_{ij} = 0$) it either holds that $w_{ij} \neq 0$, thus again $\lim_{k\to\infty} w_{ij}^{\langle\langle k\rangle\rangle} = b_{ij}$, or $w_{ij} = 0$, in which case $\lim_{k\to\infty} w_{ij}^{\langle\langle k\rangle\rangle} = \lim_{k\to\infty} 0 (y_i^{\langle k\rangle} z_j^{\langle k\rangle} y_j^{\langle k\rangle} z_i^{\langle k\rangle})^{1/2} = 0 = b_{ij}$. □

E Proof of Lemma 4.4 (L_1-monotony)

Lemma 4.4 (L_1-monotony). *For any $W \in \mathcal{S}$ and any mean function m, the m-sequence of W implies that $\|\mathbf{f} - \mathbf{d}^{(k)}\|_1$ is monotonously decreasing.*

Proof. For $k \geq 0$ we get with $s_i := f_i / d_i^{(k)}$ that

$$
\begin{aligned}
& \|\mathbf{f} - \mathbf{d}^{(k+1)}\|_1 \;=\; \textstyle\sum_i |f_i - d_i^{(k+1)}| \\
= \;& \textstyle\sum_i |\sum_j w_{ij}^{(k)} s_i - \sum_j w_{ij}^{(k)} m(s_i, s_j)| \\
\leq \;& \textstyle\sum_i \sum_{j>i} w_{ij}^{(k)} \left(|s_i - m(s_i, s_j)| + |s_j - m(s_j, s_i)|\right) \\
\overset{(\star)}{=} \;& \textstyle\sum_i \sum_{j>i} w_{ij}^{(k)} |s_i - s_j| \\
\leq \;& \textstyle\sum_i \sum_{j>i} w_{ij}^{(k)} \left(|s_i - 1| + |s_j - 1|\right) \\
\leq \;& \textstyle\sum_i |s_i - 1| \cdot d_i^{(k)} \;=\; \sum_i |f_i - d_i^{(k)}| \;=\; \|\mathbf{f} - \mathbf{d}^{(k)}\|_1 \quad ,
\end{aligned}
$$

where equality $(\star)$ holds true because of $m(s_i, s_j) = m(s_j, s_i) \in [\min(s_i, s_j), \max(s_i, s_j)]$. □

F Proof of Lemma 4.5 (Volume bounds)

Lemma 4.5 (Volume bounds). *For any $W \in \mathcal{S}$ and any mean function m, the m-sequence of W satisfies for all $k \geq 1$ that*

$$
\begin{array}{rlcll}
(i) & \|\mathbf{d}^{(k)}\|_1 & \leq & \|\mathbf{f}\|_1 & \text{if } m \text{ is sub-arithmetic} \\
(ii) & \|\mathbf{d}^{(k)}\|_1 & = & \|\mathbf{f}\|_1 & \text{if } m = m_A \\
(iii) & \|\mathbf{d}^{(k)}\|_1 & \geq & \|\mathbf{f}\|_1 & \text{if } m \text{ is super-arithmetic}
\end{array}
$$

If m is strict in (i) or (iii), then equality holds if and only if $f_i / d_i^{(k)} = f_j / d_j^{(k)}$ for all $w_{ij} \neq 0$.

Proof. (i) For $k \geq 0$ we get with $s_i := f_i / d_i^{(k)}$ that

$$
\begin{aligned}
& \|\mathbf{f}\|_1 - \|\mathbf{d}^{(k+1)}\|_1 = \textstyle\sum_i (\sum_j w_{ij}^{(k)} s_i - \sum_j w_{ij}^{(k)} m(s_i, s_j)) \\
& = \textstyle\sum_i \sum_{j>i} w_{ij}^{(k)} \underbrace{(s_i - m(s_i, s_j) + s_j - m(s_j, s_i))}_{\geq 0}
\end{aligned}
$$

where non-negativity of each summand follows from $x + y - 2m(x, y) \geq x + y - 2(x + y)/2 = 0$ $(\star)$ for sub-arithmetic m. If m is strict, then the inequality in $(\star)$ holds with equality iff $x = y$. This implies that

$\|\mathbf{f}\|_1 - \|\mathbf{d}^{(k+1)}\|_1 \geq 0$ holds with equality iff $s_i = s_j$ whenever $w_{ij}^{(k)} \neq 0$, that is $w_{ij} \neq 0$. (ii) follows from (i) and the fact that ($\star$) always holds with equality for $m = m_A$. (iii) equals (i) with all inequalities flipped. $\square$

G Proof of Lemma 4.6 (Limit points)

Lemma 4.6 (Limit points). *Every m-sequence is bounded and has at least one limit point W^*. If $\|\mathbf{f} - \mathbf{d}^{(k)}\|_1 \to 0$, then every limit point W^* satisfies $W^*\mathbf{1} = \mathbf{f}$. If further $m = m_G$, then W^* is the (unique) biproportional fit of W to row and column marginals $\mathbf{f}$, and it holds that $W^{(k)} \to W^*$.*

Proof. Lemma 4.4 implies that there exists $L > 0$ such that $w_{ij}^{(k)} \leq L$ for all i, j, k. Hence $(W^{(k)})$ is bounded and thus has at least one limit point W^* in compact $[0, L]^{n\times n}$. We now show that if $\|\mathbf{f} - \mathbf{d}^{(k)}\|_1 \to 0$, then any limit point W^* of the m-sequence satisfies $W^*\mathbf{1} = \mathbf{f}$. Let $(W^{(k_i)})$ denote any subsequence that converges to W^*. Then $\|\mathbf{f} - W^*\mathbf{1}\|_1 \leq \|\mathbf{f} - \mathbf{d}^{(k_i)}\| + \|(W^{(k_i)} - W^*)\mathbf{1}\|_1 \longrightarrow 0$. Now assume that $m = m_G$. Then $W^{(k)} = T^{(k)}WT^{(k)}$ for some $T^{(k)} \in \operatorname{diag}(\mathbb{R}^n_{>0})$, which gives that any limit point W^* is a biproportional scaling of W, that is $W^* = \lim_{i\to\infty} T^{(k_i)}WT^{(k_i)}$ for a subsequence $(W^{(k_i)})_{i\geq 0}$. Because of $W^*\mathbf{1} = \mathbf{f}$ we get that W^* is unique by the uniqueness of biproportional fits. Finally, any bounded sequence with a single limit point converges to it, hence $W^{(k)} \to W^*$. $\square$

H Proof of Lemma 4.7 (Strong convergence)

Lemma 4.7 (Strong convergence). *For any $\mathbf{x} := (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0}$, $\mathbf{a} := (a_1, \ldots, a_n) \in \mathbb{R}^n_{>0}$ with $\sum_i x_i \leq \sum_i a_i$ let $f(\mathbf{x}) := \sum_i a_i \log \frac{a_i}{x_i}$. Then*

$$f(\mathbf{x}) \geq 0 \text{ with equality iff } \mathbf{x} = \mathbf{a}. \tag{1}$$

Further, for any sequence $(\mathbf{x}^{(k)})_{k\geq 0}$ in $\mathbb{R}^n_{>0}$ with $\sum_i x_i^{(k)} \leq \sum_i a_i$ it holds that

$$\lim_{k\to\infty} f(\mathbf{x}^{(k)}) = 0 \quad \Leftrightarrow \quad \lim_{k\to\infty} \mathbf{x}^{(k)} = \mathbf{a}. \tag{2}$$

Proof. (1) follows from the Bregman divergence $RE(\mathbf{a}\|\mathbf{x}) = \sum_i a_i \log(a_i/x_i) - a_i + x_i \geq 0$, hence $\sum_i a_i \log(a_i/x_i) \geq \sum_i a_i - \sum_i x_i \geq 0$ with equality if and only if $\mathbf{a} = \mathbf{x}$. We now prove (2). "$\Leftarrow$" follows from continuity of f, so it remains to show "$\Rightarrow$". From $\sum_i x_i^{(k)} \leq \sum_i a_i =: a$ we get that $x_i^{(k)} \in (0, a]$ for all $i \in \{1, \ldots, n\}$ and $k \geq 0$. Compactness of $[0, a]^n$ implies that all limit points of $S := (\mathbf{x}^{(k)})$ lie within $[0, a]^n$, and that there exists at least one. Let $\mathbf{c}$ denote a limit point of S and $(\mathbf{y}^{(k)})$ a subsequence that converges to $\mathbf{c}$. Note that $\lim_{\epsilon\to 0} a_i \log(a_i/\epsilon) = \infty$. Hence we get by $f(\mathbf{y}^{(k)}) \to 0$ and by continuity of f that $\mathbf{c} \in (0, a]^n$ and $f(\mathbf{c}) = 0$. Now (1) implies that $\mathbf{c} = \mathbf{a}$ is the unique limit point of S, thus $\mathbf{x}^{(k)} \to \mathbf{a}$. $\square$

I Proof (sketch) of Proposition 5.2 (Strictly positive feasibility)

Consider the following specialization of Proposition 5.2 to connected graphs.

Proposition 5.2* (Strictly positive feasibility for connected graphs). *Let $\mathcal{G}(W) = (V, E, W)$ denote the graph corresponding to $W \in \mathcal{S}$, and assume that it is connected. For any $\mathbf{f} \in \mathbb{R}^n_{>0}$ there exists a strictly positive solution in $\mathcal{S}(\mathbf{f}, W)$ if and only if $\mathcal{G}(W)$ is a weak $\mathbf{f}$-expander that is strict for all $S \notin \{\emptyset, V_1, V_2, V\}$, with either $V_1 := V_2 := \emptyset$ if $G(W)$ is non-bipartite, or $V =: V_1 \,\dot\cup\, V_2$ the bipartition of $G(W)$.*

It is straightforward to derive Proposition 5.2 from Proposition 5.2*, that is to generalize from connected graphs to possibly unconnected graphs, by applying Proposition 5.2* to each connected component, and aggregating the exceptional sets that force equality. So it remains to proof Proposition 5.2*.

We need some additional notation and concepts for the proof:

For $i, j \in V$ with $ij \in E$ we also write $i \sim j$. For $i \in V$ and $X \subseteq V$ we write $i \sim X$ if there exists a vertex $x \in X$ with $i \sim x$. Otherwise, if i is not adjacent to any of the vertices in X, we write $i \not\sim X$. For $X, Y \subseteq V$ let $G[X, Y] := \{ij = ji \in E \mid (i, j) \in X \times Y\}$ denote the set of edges connecting some vertex in X to some vertex in Y. Thus, $B \subseteq V$ is an independent set if and only if $G[B, B] = \emptyset$. Observe that for all $X, Y \subseteq V$ it holds that

$$G[X \cup Y, Y] = \emptyset \quad \Leftrightarrow \quad G[X, Y] = G[Y, Y] = \emptyset.$$

For any set $B \subseteq V$, we call any disjoint $\tilde{B} \subseteq V \setminus B$ satisfying $G[\tilde{B}, B] = \emptyset$ a *non-adjacent opponent of* B. This also defines their *rest* $R(B, \tilde{B}) := V \setminus (B \,\dot{\cup}\, \tilde{B})$, which partitions V into $V = B \,\dot{\cup}\, \tilde{B} \,\dot{\cup}\, R(B, \tilde{B})$. The set $B^* := \{i \in V \setminus B \mid i \not\sim B\}$ is the (unique) *maximum non-adjacent opponent* of B. Thus, all vertices from the rest $R(B, B^*)$ are adjacent to B. Any other non-adjacent opponent $\tilde{B}$ of B satisfies that $\tilde{B} \subset B^*$.

Let $\delta B := \{i \in V \setminus B \mid \exists j \in B : i \sim j\}$ denote the *vertex boundary* of B. It holds that $\delta B = N(B) \setminus B$. Further $\delta B = N(B)$ if and only if B is an independent set. For any set B and its maximum non-adjacent opponent B^* their rest is $R(B, B^*) = \delta B$.

Proof (sketch) of Proposition 5.2★. We start with Theorem 7 of Behrend (2013), restricted to a connected graph with $n \geq 2$ vertices. This is equivalent to the tri-partition statement (i) below. We then transform (i) along a sequence of equivalent statements (i) ⇔ (ii) ⇔ (iii) ⇔ (iv) ⇔ (v) to the more intuitive statement (v) in terms of $\mathbf{f}$-expansion, which is equivalent to Proposition 5.2★.

Let $G = (V, E, W)$ for $W \in \Omega$ be connected, and $\mathbf{f} \in \mathbb{R}^n_{>0}$. Then there exists a strictly positive solution in $\mathcal{S}(\mathbf{f}, W)$ if and only if any of the following equivalent statements holds:

(i) $\sum_{i \in A} f_i \geq \sum_{j \in B} f_j$ for all $A, \tilde{A}, B$ with $V = A \,\dot{\cup}\, \tilde{A} \,\dot{\cup}\, B$ and $G[\tilde{A} \,\dot{\cup}\, B, B] = \emptyset$, where equality holds if and only if additionally $G[A, A \,\dot{\cup}\, \tilde{A}] = \emptyset$.

(ii) $\sum_{i \in A} f_i \geq \sum_{j \in B} f_j$ for all $A, \tilde{A}, B$ with $V = A \,\dot{\cup}\, \tilde{A} \,\dot{\cup}\, B$ and $G[\tilde{A} \,\dot{\cup}\, B, B] = \emptyset$, where equality holds if and only if additionally either $A = B = \emptyset$ or ($V = A \,\dot{\cup}\, B$ and $G[A, A] = G[B, B] = \emptyset$).

(iii) $\sum_{i \in A} f_i \geq \sum_{j \in B} f_j$ for all independent sets B, non-adjacent opponents $\tilde{B}$ and $A = R(B, \tilde{B})$, where equality holds if and only if additionally either $A = B = \emptyset$ or ($V = A \,\dot{\cup}\, B$ and $G[A, A] = G[B, B] = \emptyset$).

(iv) $\sum_{i \in A} f_i \geq \sum_{j \in B} f_j$ for all independent sets B, and $A = N(B)$, where equality holds if and only if additionally either $A = B = \emptyset$ or ($V = A \,\dot{\cup}\, B$ and $G[A, A] = G[B, B] = \emptyset$).

(v) $\sum_{i \in A} f_i \geq \sum_{j \in B} f_j$ for all subsets $B \subseteq V$, and $A = N(B)$, where equality holds if and only if additionally either $B = \emptyset$, or $B = V$, or ($V = A \,\dot{\cup}\, B$ and $G[A, A] = G[B, B] = \emptyset$).

Each intermediate step follows more or less straightforward by carefully applying all definitions. □