Convergence of BDRS as a Matrix Scaling Algorithm

Shiqian Ma
CMOR Department, Rice University. sqma@rice.edu

May 19, 2026

Abstract

This note proves the convergence of the Bregman Douglas-Rachford splitting method (BDRS) under certain verifiable condition, when it is treated as a matrix scaling algorithm. This is the first non-trivial convergence result of BDRS. The proof was generated by ChatGPT 5.5 and verified by the author.

1 The Bregman Douglas-Rachford Splitting Method

Recently, Ma et al. proposed the Bregman Douglas-Rachford Splitting method (BDRS) [8] for solving monotone operator inclusion problems:

$$\text{Find } x, \text{ s.t., } 0 \in A(x) + B(x), \tag{1}$$

where A and B are two maximal monotone operators. For a Legendre function h, the Bregman resolvent operator of a maximal monotone operator T is defined as [3]:

$$J_T^h := (\nabla h + T)^{-1} \circ \nabla h. \tag{2}$$

A typical iteration of BDRS for solving (1) is as follows:

$$x^k := J_{\gamma B}^h(z^k) \tag{3a}$$
$$y^k := J_{\gamma A}^h \circ \nabla h^*(2\nabla h(x^k) - \nabla h(z^k)) \tag{3b}$$
$$z^{k+1} := \nabla h^*(\nabla h(z^k) - \nabla h(x^k) + \nabla h(y^k)), \tag{3c}$$

where $\gamma > 0$ is a parameter, and h^* is the conjugate function of h.

A representative application of the BDRS algorithm is the discrete optimal transport problem:

$$\min_{X\in\mathbb{R}^{m\times n}} \langle C, X\rangle, \text{ s.t., } X\mathbf{1}_n = r, X^\top \mathbf{1}_m = c, X \geq 0, \tag{4}$$

where $X \in \mathbb{R}^{m\times n}$ is the transportation plan, $C \in \mathbb{R}^{m\times n}$ is the cost matrix, r and c are the row and column sums satisfying $r > 0$, $c > 0$, $\sum_i r_i = \sum_j c_j = 1$, $\mathbf{1}$ denotes the all-one vector. BDRS (3) can be applied to solve (4), because (4) can be written in the form of (1) by defining

$$A(X) = C + \partial\mathbb{I}(X\mathbf{1}_n = r) + \partial\mathbb{I}(X \geq 0), \quad B(X) = \partial\mathbb{I}(X^\top \mathbf{1}_m = c) + \partial\mathbb{I}(X \geq 0). \tag{5}$$

We now introduce some definition and notation. The Bregman distance corresponding to the Legendre function h, denoted $D_h(x, y)$, is defined as

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y\rangle. \tag{6}$$

Throughout this paper, we use $x \odot y$ to denote the componentwise multiplication of vectors (or matrices) x and y with same dimension. When it is clear from the context, we use $\frac{x}{y}$ to denote the componentwise division of vectors (or matrices) x and y with same dimension. That is,

$$(x \odot y)_i = x_i y_i, \quad \left(\frac{x}{y}\right)_i = \frac{x_i}{y_i}.$$

Moreover, x^2 denotes the componentwise square of vector (or matrix) x, i.e., $(x^2)_i = x_i^2$.

By denoting $\eta := 1/\gamma$, the BDRS (3) for solving (4) can be written as:

$$X^k := \underset{X}{\operatorname{argmin}} \ D_h(X, Z^k), \text{ s.t., } X^\top \mathbf{1}_m = c \tag{7a}$$

$$Y^k := \underset{Y}{\operatorname{argmin}} \ \langle C, Y\rangle + \eta D_h\left(Y, \frac{(X^k)^2}{Z^k}\right), \text{ s.t., } Y\mathbf{1}_n = r \tag{7b}$$

$$Z^{k+1} := \frac{Z^k \odot Y^k}{X^k}. \tag{7c}$$

The BDRS algorithm presented in (7) is actually closely related to the Sinkhorn's algorithm for matrix scaling and optimal transport [10, 11]. To see this, we first present the dual problem of (4):

$$\min_{\alpha,\beta} \ -r^\top \alpha - c^\top \beta, \text{ s.t., } \alpha_i + \beta_j \le C_{ij}, i \in [m], j \in [n]. \tag{8}$$

As established in [8], the BDRS for solving (4) is equivalent to the alternating direction exponential method of multipliers (ADEMM) for solving (8), which updates the iterates as follows:

$$\alpha^k := \underset{\alpha}{\operatorname{argmin}} \ -r^\top \alpha + \eta \sum_{ij} X_{ij}^k e^{\frac{1}{\eta}(\alpha_i + \beta_j^{k-1} - C_{ij})} \tag{9a}$$

$$\beta^k := \underset{\beta}{\operatorname{argmin}} \ -c^\top \beta + \eta \sum_{ij} X_{ij}^k e^{\frac{1}{\eta}(\alpha_i^k + \beta_j - C_{ij})} \tag{9b}$$

$$X_{ij}^{k+1} := X_{ij}^k e^{\frac{1}{\eta}(\alpha_i^k + \beta_j^k - C_{ij})}, \ i \in [m], j \in [n]. \tag{9c}$$

By denoting $u = e^{\alpha/\eta}$, $v = e^{\beta/\eta}$, and $K_{ij} = e^{-C_{ij}/\eta}$, (9) can be further rewritten as

$$u^k := \underset{u}{\operatorname{argmin}} \ \sum_{ij} u_i X_{ij}^k K_{ij} v_j^{k-1} - \sum_i r_i \log u_i \tag{10a}$$

$$v^k := \underset{v}{\operatorname{argmin}} \ \sum_{ij} u_i^k X_{ij}^k K_{ij} v_j - \sum_j c_j \log v_j \tag{10b}$$

$$X_{ij}^{k+1} := u_i^k X_{ij}^k K_{ij} v_j^k, \ i \in [m], j \in [n]. \tag{10c}$$

The two subproblems (10a) and (10b) admit closed-form solutions, and (10) can be equivalently written as:

$$u^k := \frac{r}{(X^k \odot K)v^{k-1}} \tag{11a}$$

$$v^k := \frac{c}{(X^k \odot K)^\top u^k} \tag{11b}$$

$$X^{k+1} := \operatorname{diag}(u^k)(X^k \odot K)\operatorname{diag}(v^k). \tag{11c}$$

However, the convergence of BDRS (3) is still missing in the literature. Even its special case (11), which solves the optimal transport problem (4), still lacks a convergence proof.

2 BDRS as a Matrix Scaling Algorithm

Matrix scaling problem has a long history and many important applications. We refer to the recent survey [5] for a thorough review on this topic. Matrix scaling aims to find a matrix with prescribed row and column sums that is on the diagonal orbit of a given positive matrix. More specifically, given a positive $m \times n$ matrix $M \in \mathbb{R}^{m\times n}_{++}$ and prescribed row and column sums $r > 0$, $c > 0$ satisfying $\sum_i r_i = \sum_j c_j = 1$, matrix scaling finds diagonal matrices $D_1 \in \mathbb{R}^{m\times m}$ and $D_2 \in \mathbb{R}^{n\times n}$ with positive diagonals such that $D_1 M D_2 \in \Omega := \{X \geq 0 \mid X\mathbf{1}_n = r, X^\top \mathbf{1}_m = c\}$. A widely used algorithm for matrix scaling is the Sinkhorn's algorithm [10, 11], which iterates as follows:

$$u^k := \frac{r}{Mv^{k-1}} \tag{12a}$$

$$v^k := \frac{c}{M^\top u^k} \tag{12b}$$

$$X^{k+1} := \mathrm{diag}(u^k) M \mathrm{diag}(v^k). \tag{12c}$$

It has been proved that $X^k \to X^* \in \Omega$, $\mathrm{diag}(u^k) \to D_1$, and $\mathrm{diag}(v^k) \to D_2$ (see, e.g., [10, 4]).

Back to the BDRS algorithm for solving (4), consider the special case where $C = 0$. In this case, K becomes an all-one matrix, and (11) reduces to

$$u^k := \frac{r}{X^k v^{k-1}} \tag{13a}$$

$$v^k := \frac{c}{(X^k)^\top u^k} \tag{13b}$$

$$X^{k+1} := \mathrm{diag}(u^k) X^k \mathrm{diag}(v^k). \tag{13c}$$

Suppose we choose initial point $X^0 = M$ in (13), then (13) can be viewed as a matrix scaling algorithm. If the sequence $\{X^k, u^k, v^k\}$ generated by (13) converges, then it should converge to a point in Ω that is on the diagonal orbit of M. So, the BDRS (13) becomes a matrix scaling algorithm. The BDRS (13) is very similar to the Sinkhorn's algorithm (12), with the only difference being the static matrix M in (12) replaced by a varying matrix X^k. This subtle difference brings significant difficulty for the convergence analysis. As BDRS still lacks a convergence guarantee, it is not clear whether and when (13) converges. In Section 4, we show that (13) indeed converges when $X^0 = M$ satisfies certain verifiable condition.

3 Preparations of the Proof

Our analysis utilizes properties of the Hilbert's projective metric. For positive vectors $x, y > 0$, the Hilbert's projective metric is defined as

$$d_H(x, y) = \log \frac{\max_i x_i/y_i}{\min_i x_i/y_i}. \tag{14}$$

We refer to [2, 6, 7] for applications, extensions, and historical remarks of the Hilbert's projective metric. This metric is invariant under positive scalar rescaling:

$$d_H(\xi x, \zeta y) = d_H(x, y), \quad \text{for any scalars } \xi, \zeta > 0.$$

It also satisfies the elementary inequalities

$$d_H(x \odot y, z \odot w) \leq d_H(x, z) + d_H(y, w), \text{ and } d_H(x, y) \leq d_H(x, z) + d_H(z, y).$$

The Sinkhorn projective map is defined as

$$S(z) := \frac{c}{M^\top(r/(Mz))}. \tag{15}$$

Note that the step (12c) is not needed in the algorithm, and (12a) and (12b) can be rewritten more compactly as one single update $v^k = S(v^{k-1})$.

The following proposition is in fact Sinkhorn's matrix scaling theorem [10, 4]. We include an elementary proof here for completeness.

Proposition 3.1. *Let* $M \in \mathbb{R}^{m\times n}_{++}$, $r \in \mathbb{R}^m_{++}$, $c \in \mathbb{R}^n_{++}$, $\sum_i r_i = \sum_j c_j = 1$. *There exist positive vectors* $p^* \in \mathbb{R}^m_{++}$, $q^* \in \mathbb{R}^n_{++}$, *unique up to the scaling transformation*

$$p^* \to \gamma p^*, \quad q^* \to \gamma^{-1} q^*,$$

such that

$$p^* = \frac{r}{Mq^*}, \quad q^* = \frac{c}{M^\top p^*}.$$

Equivalently,

$$q^* = S(q^*), \tag{16}$$

where S *is the Sinkhorn projective map defined in* (15).

Proof. Define a convex function

$$F(\alpha, \beta) := \sum_{ij} M_{ij} e^{\alpha_i + \beta_j} - \sum_i r_i \alpha_i - \sum_j c_j \beta_j.$$

Now consider the constrained minimization problem

$$\min_{\alpha,\beta} F(\alpha,\beta), \quad \text{s.t.,} \quad \sum_j \beta_j = 0.$$

We now prove that F is coercive on the affine subspace $\sum_j \beta_j = 0$. That is, we want to prove

$$\|(\alpha,\beta)\| \to \infty, \ \sum_j \beta_j = 0 \Longrightarrow F(\alpha,\beta) \to +\infty.$$

Define $R_{ij} := r_i c_j$, $f_{ij}(t) := M_{ij} e^t - R_{ij} t$ and $t_{ij} := \alpha_i + \beta_j$. Then F can be rewritten as

$$F(\alpha,\beta) = \sum_{ij} f_{ij}(t_{ij}).$$

It is easy to verify that each scalar function f_{ij} is coercive. To see this, note that $M_{ij} > 0$ and $R_{ij} > 0$. If $t \to +\infty$, then $M_{ij} e^t \to +\infty$ and $f_{ij}(t) \to +\infty$. If $t \to -\infty$, then $e^t \to 0$, and $-R_{ij} t \to +\infty$, and therefore $f_{ij}(t) \to +\infty$. Hence, $f_{ij}(t) \to +\infty$ as $|t| \to \infty$, which implies that

$$F(\alpha,\beta) \to +\infty, \qquad \text{whenever } \max_{ij} |\alpha_i + \beta_j| \to \infty.$$

Since $\sum_j \beta_j = 0$, for each fixed i, we have

$$\frac{1}{n}\sum_j t_{ij} = \frac{1}{n}\sum_j (\alpha_i + \beta_j) = \alpha_i + \frac{1}{n}\sum_j \beta_j = \alpha_i.$$

Therefore, if $\max_{ij} |\alpha_i + \beta_j|$ is upper bounded, then t_{ij}'s are uniformly bounded, and therefore α_i's are uniformly bounded. This then implies that $\beta_j = t_{ij} - \alpha_i$ is also uniformly bounded. As a result,

$$\|(\alpha, \beta)\| \to \infty, \quad \sum_j \beta_j = 0 \quad \Longrightarrow \quad \max_{ij} |\alpha_i + \beta_j| \to \infty.$$

This proves that F is coercive on $\sum_j \beta_j = 0$.

Now since F is continuous and coercive on the closed affine subspace $\sum_j \beta_j = 0$, it attains a minimizer (α^*, β^*) satisfying the first-order optimality conditions:

$$\sum_j M_{ij} e^{\alpha_i^* + \beta_j^*} = r_i, \quad \sum_i M_{ij} e^{\alpha_i^* + \beta_j^*} = c_j + \lambda, \quad \sum_j \beta_j^* = 0, \tag{17}$$

where λ is the Lagrange multiplier associated with the constraint $\sum_j \beta_j = 0$. Since $\sum_i r_i = \sum_j c_j = 1$, (17) implies $\lambda = 0$ and therefore,

$$\sum_j M_{ij} e^{\alpha_i^* + \beta_j^*} = r_i, \quad \sum_i M_{ij} e^{\alpha_i^* + \beta_j^*} = c_j. \tag{18}$$

Set $p_i^* = e^{\alpha_i^*}$, and $q_j^* = e^{\beta_j^*}$. Define $X^* = \text{diag}(p^*) M \text{diag}(q^*)$. Then X^* satisfies $X^* \mathbf{1} = r$ and $(X^*)^\top \mathbf{1} = c$. Therefore, $p_i^* (M q^*)_i = r_i$. So $p^* = \frac{r}{M q^*}$. Similarly, $q_j^* (M^\top p^*)_j = c_j$, so $q^* = \frac{c}{M^\top p^*}$. This proves the existence.

Now we prove the uniqueness of p^* and q^* up to scaling transformation. We claim that F is strictly convex, on the affine subspace $\sum_j \beta_j = 0$. This can be proved as follows. The second directional derivative of F at (α, β) in direction (h, k) is

$$D^2 F(\alpha, \beta)[(h, k), (h, k)] = \sum_{ij} M_{ij} e^{\alpha_i + \beta_j} (h_i + k_j)^2. \tag{19}$$

Because $M_{ij} > 0$, this quadratic form is zero only if

$$h_i + k_j = 0, \quad \forall i, j.$$

Therefore all h_i are equal to a constant ξ, and all k_j are equal to $-\xi$. But the constraint $\sum_j \beta_j = 0$ gives $\sum_j k_j = 0$, so $\xi = 0$. Hence $h = 0$ and $k = 0$. Therefore F is strictly convex on the constraint set $\sum_j \beta_j = 0$. Therefore, the minimizer of F over $\sum_j \beta_j = 0$ is unique. Suppose two pairs (α, β) and $(\tilde{\alpha}, \tilde{\beta})$ satisfy the scaling equations. Normalize both by imposing $\sum_j \beta_j = 0$ and $\sum_j \tilde{\beta}_j = 0$. By strict convexity of F on the gauge-fixed subspace, the normalized pairs coincide. Hence the original pairs differ only by a gauge transformation $(\alpha, \beta) \to (\alpha + t\mathbf{1}_m, \beta - t\mathbf{1}_n)$. Denote

$$p_i = e^{\alpha_i}, q_j = e^{\beta_j}, \tilde{p}_i = e^{\tilde{\alpha}_i}, \tilde{q}_j = e^{\tilde{\beta}_j}.$$

We have

$$\tilde{p} = e^t p, \quad \tilde{q} = e^{-t} q.$$

This proves the uniqueness up to the scaling transformation. □

The following lemma shows an important property of Algorithm (13).

Lemma 3.2. *$\{X^k\}$ generated by* (13) *satisfies*

$$(X^{k+1})^\top \mathbf{1}_m = c. \tag{20}$$

Proof. This desired result is a direct consequence of the updates (13b) and (13c):

$$(X^{k+1})^\top \mathbf{1}_m = \text{diag}(v^k)(X^k)^\top \text{diag}(u^k) \mathbf{1}_m = \text{diag}(v^k)(X^k)^\top u^k = v^k \odot ((X^k)^\top u^k) = c,$$

where the first equality is from (13c), and the last equality is from (13b). □

4 Convergence Proof of BDRS (13)

Now we are ready to prove the convergence of BDRS as a matrix scaling algorithm, i.e., algorithm given in (13).

Theorem 4.1. *Define the projective diameter of matrix M as*

$$\Delta(M) := \log \max_{i,i',j,j'} \frac{M_{ij}M_{i'j'}}{M_{ij'}M_{i'j}}.$$

Assume the given matrix $M > 0$ satisfies

$$\Delta(M) < 4\text{arctanh}(1/\sqrt{3}), \ \textit{i.e.,} \ \tanh^2\left(\frac{\Delta(M)}{4}\right) < \frac{1}{3}. \tag{21}$$

In Algorithm (13), choose initial points $X^0 = M > 0$, and $v^{-1} > 0$. Then X^k generated by (13) converges to a positive feasible coupling X^*, satisfying $X^*\mathbf{1}_n = r$ and $(X^*)^\top \mathbf{1}_m = c$.

Proof. Define the cumulative scaling vectors

$$p^k := u^{k-1} \odot u^{k-2} \odot \cdots \odot u^0, \quad q^k := v^{k-1} \odot v^{k-2} \odot \cdots \odot v^0, \quad \text{for } k \geq 1,$$

and $p^0 = \mathbf{1}_m$ and $q^0 = \mathbf{1}_n$. From (13c) we have

$$X^k = \text{diag}(p^k)M\text{diag}(q^k).$$

This immediately yields

$$X^k v^{k-1} = \text{diag}(p^k)M(q^k \odot v^{k-1}),$$

which, together with (13a), yields,

$$u^k = \frac{r}{p^k \odot M(q^k \odot v^{k-1})}.$$

Therefore,

$$p^{k+1} = p^k \odot u^k = \frac{r}{M(q^k \odot v^{k-1})}. \tag{22}$$

Moreover, since

$$(X^{k+1})^\top \mathbf{1}_m = \text{diag}(q^{k+1})M^\top p^{k+1},$$

using (20) we have

$$q^{k+1} = \frac{c}{M^\top p^{k+1}}. \tag{23}$$

Combining (22) and (23), we have

$$q^{k+1} = \frac{c}{M^\top (r/[M(q^k \odot v^{k-1})])} = S(q^k \odot v^{k-1}),$$

which, together with $v^{k-1} = q^k/q^{k-1}$ for $k \geq 1$, yields the second-order recurrence:

$$q^{k+1} = S\left(\frac{(q^k)^2}{q^{k-1}}\right), \quad k \geq 1.$$

Here S is the Sinkhorn projective map defined in (15). Because $M > 0$, its projective diameter is finite: $\Delta(M) < \infty$. Birkhoff's contraction theorem [1, 4] gives

$$d_H(Mx, My) \leq \tau(M) d_H(x, y),$$

where $\tau(M) = \tanh(\Delta(M)/4)$. The same bound holds for $M^\top$, since $\Delta(M^\top) = \Delta(M)$. The componentwise reciprocal map $x \to \frac{1}{x}$ is an isometry for Hilbert's metric:

$$d_H(1/x, 1/y) = d_H(x, y).$$

Multiplication by a fixed positive vector is also an isometry:

$$d_H(a \odot x, a \odot y) = d_H(x, y).$$

Therefore, the map S defined in (15) satisfies

$$d_H(S(x), S(y)) \leq \tau(M)^2 d_H(x, y). \tag{24}$$

Let $\kappa := \tau(M)^2$. From (21), we have $\kappa < 1/3$.

Pick q^* that satisfies (16). Define

$$e_k := d_H(q^k, q^*).$$

From (24) we get:

$$e_{k+1} = d_H\left(S\left(\frac{(q^k)^2}{q^{k-1}}\right), S(q^*)\right) \leq \kappa d_H\left(\frac{(q^k)^2}{q^{k-1}}, q^*\right). \tag{25}$$

Since

$$\frac{(q^k)^2}{q^{k-1}} = q^k \odot \frac{q^k}{q^{k-1}}, \text{ and } q^* = q^* \odot \mathbf{1}_n,$$

we get

$$d_H\left(\frac{(q^k)^2}{q^{k-1}}, q^*\right) \leq d_H(q^k, q^*) + d_H\left(\frac{q^k}{q^{k-1}}, \mathbf{1}_n\right) = d_H(q^k, q^*) + d_H(q^k, q^{k-1}) \leq 2d_H(q^k, q^*) + d_H(q^{k-1}, q^*).$$

Therefore,

$$e_{k+1} \leq \kappa(2e_k + e_{k-1}).$$

Now define the linear comparison sequence s_k by

$$s_0 = e_0, \quad s_1 = e_1, \quad s_{k+1} = \kappa(2s_k + s_{k-1}).$$

By induction, we have $e_k \leq s_k, \forall k$. The characteristic equation of the recurrence for s_k is (see, e.g., [9])

$$\lambda^2 - 2\kappa\lambda - \kappa = 0,$$

whose larger root is

$$\rho = \kappa + \sqrt{\kappa^2 + \kappa}.$$

The condition $\kappa < 1/3$ is equivalent to $\rho < 1$. Therefore, $s_k \to 0$, which implies $e_k = d_H(q^k, q^*) \to 0$. This means that q^k converges projectively to q^*.

Now we prove the projective convergence of p^k. Define

$$y^k := q^k \odot v^{k-1} = \frac{(q^k)^2}{q^{k-1}}, k \geq 1.$$

Since $q^k \to q^*$ projectively, we have $y^k \to q^*$ projectively from

$$d_H(y^k, q^*) \le 2d_H(q^k, q^*) + d_H(q^{k-1}, q^*) \to 0.$$

Denote $a_j^k := \frac{q_j^{k+1}}{y_j^k}$. Since $d_H(q^{k+1}, q^*) \to 0$, $d_H(y^k, q^*) \to 0$, we have $d_H(q^{k+1}, y^k) \to 0$ and therefore $\frac{\max_j a_j^k}{\min_j a_j^k} \to 1$. Denote $\lambda_k := \min_j a_j^k$. We now prove

$$\max_j \left| \frac{a_j^k}{\lambda_k} - 1 \right| \to 0. \tag{26}$$

In fact, we have

$$1 \le \frac{a_j^k}{\lambda_k} \le \frac{\max_j a_j^k}{\min_j a_j^k},$$

which implies

$$0 \le \frac{a_j^k}{\lambda_k} - 1 \le \frac{\max_j a_j^k}{\min_j a_j^k} - 1.$$

Taking maximum over j, we have

$$0 \le \max_j \left| \frac{a_j^k}{\lambda_k} - 1 \right| \le \max_j \left(\frac{a_j^k}{\lambda_k} - 1 \right) \le \frac{\max_j a_j^k}{\min_j a_j^k} - 1,$$

which proves (26). Therefore, we further have

$$\frac{q^{k+1}}{y^k} = \lambda_k \mathbf{1}_n + o(\lambda_k), \text{ and } \frac{Mq^{k+1}}{My^k} = \lambda_k \mathbf{1}_m + o(\lambda_k). \tag{27}$$

From (22), we have

$$p^{k+1} = \frac{r}{My^k}. \tag{28}$$

Therefore, from (27) we have

$$X^{k+1}\mathbf{1}_n = \operatorname{diag}(p^{k+1})Mq^{k+1} = r \odot \frac{Mq^{k+1}}{My^k} = r \odot \lambda_k \mathbf{1}_m + o(\lambda_k).$$

Together with (20), we have

$$\mathbf{1}_n^\top (X^{k+1})^\top \mathbf{1}_m = 1 = \lambda_k + o(\lambda_k) = \lambda_k(1 + o(1)).$$

Therefore, $\lambda_k = 1/(1+o(1)) \to 1$, $\frac{q^{k+1}}{y^k} \to \mathbf{1}_n$, $\frac{Mq^{k+1}}{My^k} \to \mathbf{1}_m$ and

$$X^{k+1}\mathbf{1}_n \to r \tag{29}$$

Now since $y^k \to q^*$ projectively, there exist scalars $\gamma_k > 0$ such that $\gamma_k y^k \to q^*$. Since $\frac{q^{k+1}}{y^k} \to \mathbf{1}_n$, we have

$$\gamma_k q^{k+1} \to q^*. \tag{30}$$

From (28) we have

$$\gamma_k^{-1} p^{k+1} = \frac{r}{M(\gamma_k y^k)} \to \frac{r}{Mq^*} = p^*. \tag{31}$$

From (30) and (31) we have

$$X^{k+1} = \operatorname{diag}(p^{k+1})M\operatorname{diag}(q^{k+1}) = \operatorname{diag}(\gamma_k^{-1}p^{k+1})M\operatorname{diag}(\gamma_k q^{k+1}) \to \operatorname{diag}(p^*)M\operatorname{diag}(q^*) = X^*,$$

which, together with (29) and (20), completes the proof. □

Remark 4.2. *The condition* (21) *is equivalent to the following verifiable condition:*

$$\max_{i,i',j,j'} \frac{M_{ij} M_{i'j'}}{M_{ij'} M_{i'j}} < e^{2.6339} \approx 13.93.$$

5 Concluding Remarks

The author started to use ChatGPT to solve this problem in Jan 2026. Back then, ChatGPT was not able to generate useful arguments. It hallucinated constantly and made false statements all the time. But in May 2026, the model generated a complete and correct proof with prompts provided by the author. Although some steps needed further clarification and proofs, all the steps generated by ChatGPT were essentially correct. The AI model evolves very impressively. The result presented in this paper is the first non-trivial convergence result for BDRS. Although we are not able to prove the convergence of BDRS in more general case at the moment, this result may shed light on future attempts to prove the convergence of BDRS.

References

[1] Garrett Birkhoff. Extensions of Jentzsch's theorem. *Trans. Ameri. Math. Soc.*, 85:219–227, 1957.

[2] P. J. Bushell. Hilbert's metric and positive contraction mappings in a Banach space. *Archive for Rational Mechanics and Analysis*, 52:330–338, 1973.

[3] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, 18(1):202–226, 1993.

[4] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and Its Applications*, 114/115:717–735, 1989.

[5] M. Idel. A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps. *arXiv preprint arXiv:1609.06349*, 2016.

[6] E. Kohlberg. The Perron-Frobenius theorem without additivity. *Journal of Mathematical Economics*, 10:299–303, 1982.

[7] E. Kohlberg and J. W. Pratt. The contraction mapping approach to the Perron-Frobenius theory: Why Hilbert's metric? *Mathematics of Operations Research*, 7(2):198–210, 1982.

[8] Shiqian Ma, Lin Xiao, and Renbo Zhao. Bregman Douglas-Rachford splitting method. *arxiv:2509.08739*, 2025.

[9] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):791–803, 1964.

[10] R. Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405, 1967.

[11] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.