

BDRS Can Fail for Matrix Scaling and Optimal Transport

Shiqian Ma

Data Science and AI Institute
Department of Applied Mathematics and Statistics
Department of Computer Science
Department of Electrical and Computer Engineering
Johns Hopkins University

September 28, 2026

Abstract

We show that the Bregman Douglas–Rachford splitting method (BDRS) can fail for matrix scaling and unregularized optimal transport. For matrix scaling, we construct a strictly positive 36×6 integer matrix, positive rational marginals of equal mass, and a positive auxiliary initialization for which the matrix iterates enter a nonconstant six-cycle after one iteration. The construction prescribes rotating scaling directions and uses cyclic symmetry to reduce the required transitions to a single interpolation condition, which is linear in the block row masses. Exact rational arithmetic certifies the example, and a direct induction proves the cycle for all iterations. A block lifting produces a 72×12 optimal transport counterexample with cost entries 2 and 3. For every fixed positive algorithm parameter, its iterates have six distinct subsequential limits and remain uniformly infeasible, although the transport problem has feasible solutions. These examples disprove the unconditional convergence of BDRS.

1 The Bregman Douglas-Rachford Splitting Method

Ma et al. proposed the Bregman Douglas–Rachford splitting method (BDRS) [5] for solving monotone operator inclusion problems:

$$\text{Find } x \text{ such that } 0 \in A(x) + B(x), \quad (1.1)$$

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 [1]:

$$J_T^h := (\nabla h + T)^{-1} \circ \nabla h. \quad (1.2)$$

A typical iteration of BDRS for solving (1.1) is as follows:

$$x^k := J_{\gamma B}^h(z^k) \quad (1.3a)$$

$$y^k := J_{\gamma A}^h \circ \nabla h^*(2\nabla h(x^k) - \nabla h(z^k)) \quad (1.3b)$$

$$z^{k+1} := \nabla h^*(\nabla h(z^k) - \nabla h(x^k) + \nabla h(y^k)), \quad (1.3c)$$

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:

$$\min_{X \in \mathbb{R}^{m \times n}} \langle C, X \rangle, \text{ s.t., } X \mathbf{1}_n = r, X^\top \mathbf{1}_m = c, X \geq 0, \quad (1.4)$$

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 $r \in \mathbb{R}_{++}^m$, $c \in \mathbb{R}_{++}^n$, $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j = 1$, $\mathbf{1}$ denotes the all-one vector. BDRS (1.3) can be applied to solve (1.4), because (1.4) can be written in the form of (1.1) by defining

$$A(X) = C + \partial \mathbb{I}(X \mathbf{1}_n = r) + \partial \mathbb{I}(X \geq 0), \quad B(X) = \partial \mathbb{I}(X^\top \mathbf{1}_m = c) + \partial \mathbb{I}(X \geq 0), \quad (1.5)$$

where $\mathbb{I}$ denotes the indicator function. We now introduce some definitions and notation. The Bregman distance corresponding to the Legendre function h , denoted $D_h(x, y)$, is defined as

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y \rangle. \quad (1.6)$$

Throughout this paper, we use $x \odot y$ to denote the componentwise multiplication of vectors (or matrices) x and y with the same dimensions. When it is clear from the context, we use $\frac{x}{y}$ to denote the componentwise division of vectors x and y with the same dimensions, and use x^2 to denote the componentwise square of a vector or matrix x . That is,

$$(x \odot y)_i = x_i y_i, \quad \left(\frac{x}{y}\right)_i = \frac{x_i}{y_i}, \quad (x^2)_i = x_i^2.$$

For the optimal transport application, we specialize to the negative entropy

$$h(X) = \sum_{i=1}^m \sum_{j=1}^n X_{ij} (\log X_{ij} - 1), \quad X \geq 0.$$

For $X > 0$, $\nabla h(X) = \log X$, and $\nabla h^*(Z) = e^Z$. Logarithms and exponentials of vectors and matrices are componentwise. With $\eta := 1/\gamma$, BDRS (1.3) for (1.4) takes the form

$$X^k := \operatorname{argmin}_X D_h(X, Z^k), \text{ s.t., } X^\top \mathbf{1}_m = c \quad (1.7a)$$

$$Y^k := \operatorname{argmin}_Y \langle C, Y \rangle + \eta D_h\left(Y, \frac{(X^k)^2}{Z^k}\right), \text{ s.t., } Y \mathbf{1}_n = r \quad (1.7b)$$

$$Z^{k+1} := \frac{Z^k \odot Y^k}{X^k}. \quad (1.7c)$$

The BDRS algorithm presented in (1.7) is closely related to Sinkhorn's algorithm for matrix scaling and optimal transport [6, 7]. To see this, we first write the dual of (1.4) in an equivalent minimization form:

$$\min_{\alpha, \beta} -r^\top \alpha - c^\top \beta, \text{ s.t., } \alpha_i + \beta_j \leq C_{ij}, i \in [m], j \in [n]. \quad (1.8)$$

As established in [5], the BDRS for solving (1.4) is equivalent to the alternating direction exponential method of multipliers (ADEMM) for solving (1.8), which updates the iterates as follows:

$$\alpha^k := \operatorname{argmin}_\alpha -r^\top \alpha + \eta \sum_{ij} X_{ij}^k e^{\frac{1}{\eta}(\alpha_i + \beta_j^{k-1} - C_{ij})} \quad (1.9a)$$

$$\beta^k := \operatorname{argmin}_\beta -c^\top \beta + \eta \sum_{ij} X_{ij}^k e^{\frac{1}{\eta}(\alpha_i^k + \beta_j - C_{ij})} \quad (1.9b)$$

$$X_{ij}^{k+1} := X_{ij}^k e^{\frac{1}{\eta}(\alpha_i^k + \beta_j^k - C_{ij})}, \quad i \in [m], j \in [n]. \quad (1.9c)$$

By denoting $u = e^{\alpha/\eta}$, $v = e^{\beta/\eta}$, and $H_{ij} = e^{-C_{ij}/\eta}$, (1.9) can be further rewritten as

$$u^k := \operatorname{argmin}_{u>0} \sum_{ij} u_i X_{ij}^k H_{ij} v_j^{k-1} - \sum_i r_i \log u_i \quad (1.10a)$$

$$v^k := \operatorname{argmin}_{v>0} \sum_{ij} u_i^k X_{ij}^k H_{ij} v_j - \sum_j c_j \log v_j \quad (1.10b)$$

$$X_{ij}^{k+1} := u_i^k X_{ij}^k H_{ij} v_j^k, \quad i \in [m], j \in [n]. \quad (1.10c)$$

The two subproblems (1.10a) and (1.10b) admit closed-form solutions, and (1.10) can be equivalently written as:

$$u^k := \frac{r}{(X^k \odot H) v^{k-1}} \quad (1.11a)$$

$$\text{(BDRS-OT)} \quad v^k := \frac{c}{(X^k \odot H)^\top u^k} \quad (1.11b)$$

$$X^{k+1} := \operatorname{diag}(u^k)(X^k \odot H) \operatorname{diag}(v^k). \quad (1.11c)$$

The projected and multiplicative formulations need not share the same initial matrix. For $k \geq 1$, the multiplicative iterates are column feasible. Setting

$$Z^k = X^k \operatorname{diag}\left(\frac{1}{v^{k-1}}\right), \quad Y^k = \operatorname{diag}(u^k)(X^k \odot H) \operatorname{diag}(v^{k-1})$$

recovers the three updates in (1.7). Thus the multiplicative trajectory from X^1 onward is an orbit of the projected BDRS formulation.

2 BDRS as a Matrix Scaling Algorithm

The matrix scaling problem has a long history and many important applications. We refer to the survey [3] 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 $r \in \mathbb{R}_{++}^m$, $c \in \mathbb{R}_{++}^n$ satisfying $\sum_{i=1}^m r_i = \sum_{j=1}^n c_j = 1$, 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 Sinkhorn's algorithm [6, 7], which iterates as follows:

$$u^k := \frac{r}{M v^{k-1}} \quad (2.1a)$$

$$v^k := \frac{c}{M^\top u^k} \quad (2.1b)$$

$$X^{k+1} := \operatorname{diag}(u^k) M \operatorname{diag}(v^k). \quad (2.1c)$$

It has been proved that $X^k \rightarrow X^* \in \Omega$, $\operatorname{diag}(u^k) \rightarrow D_1$, and $\operatorname{diag}(v^k) \rightarrow D_2$ (see, e.g., [6, 2]).

Back to the BDRS algorithm for solving (1.4), consider the special case where $C = 0$. In this case, H becomes an all-one matrix, and (1.11) reduces to

$$u^k := \frac{r}{X^k v^{k-1}} \quad (2.2a)$$

$$\text{(BDRS-MS)} \quad v^k := \frac{c}{(X^k)^\top u^k} \quad (2.2b)$$

$$X^{k+1} := \operatorname{diag}(u^k) X^k \operatorname{diag}(v^k). \quad (2.2c)$$

With $X^0 = M > 0$ and $v^{-1} > 0$, (2.2) is a matrix-scaling iteration: every iterate is in the positive diagonal orbit of M , and the goal is to obtain a matrix in that orbit with marginals r and c . (BDRS-MS) is very similar to Sinkhorn's algorithm (2.1), with the only difference being the fixed matrix M replaced by the current iterate X^k . The roles of the multipliers are different: in (BDRS-MS), u^k and v^k are incremental scalings of X^k , whereas in (2.1) they are the full scalings of the fixed matrix M . This difference makes the BDRS convergence analysis substantially more difficult. For the matrix-scaling specialization (BDRS-MS), Ma [4] proved its global convergence under the following verifiable sufficient condition:

$$\max_{\substack{i,i' \in [m] \\ j,j' \in [n]}} \frac{M_{ij}M_{i'j'}}{M_{ij'}M_{i'j}} < 7 + 4\sqrt{3} \approx 13.93.$$

In this paper, we construct counterexamples showing that (BDRS-MS) and (BDRS-OT) do not converge unconditionally. Each counterexample includes a specified strictly positive auxiliary initialization. We first derive a second-order recurrence for (BDRS-MS), i.e., (2.2). For a positive matrix K and positive marginals r, c , define the Sinkhorn map

$$S_K^{r,c}(z) = \frac{c}{K^\top(r/(Kz))}, \quad z > 0. \quad (2.3)$$

Define the cumulative scaling vectors

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

and $p^0 = \mathbf{1}_m$ and $q^0 = \mathbf{1}_n$. From (2.2c) 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 (2.2a), yields

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

Therefore,

$$p^{k+1} = p^k \odot u^k = \frac{r}{M(q^k \odot v^{k-1})}. \quad (2.4)$$

Moreover, since

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

and

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

we have

$$q^{k+1} = \frac{c}{M^\top p^{k+1}}. \quad (2.5)$$

Combining (2.4) and (2.5), we have

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

Define $q^{-1} := \mathbf{1}_n / v^{-1}$. Then $v^{k-1} = q^k / q^{k-1}$ for every $k \geq 0$, including the initialization. Consequently,

$$q^{k+1} = S_M^{r,c} \left(\frac{(q^k)^2}{q^{k-1}} \right), \quad k \geq 0. \quad (2.6)$$

The map $S_M^{r,c}$ is the ordinary Sinkhorn map; it is the extrapolated argument in (2.6) that makes the (BDRS-MS) recurrence second order.

3 The Construction Strategy of Counterexamples for (BDRS-MS)

In this section, we describe a strategy for constructing nonconstant cycles of (BDRS-MS). We seek positive data M, r, c and a positive auxiliary initialization v^{-1} such that, with $X^0 = M$, the matrix iterates $\{X^k\}$ of (BDRS-MS) enter a nonconstant cycle after the first iteration. Throughout this construction, $n \geq 2$.

Let $R = R_n \in \mathbb{R}^{n \times n}$ denote the cyclic right shift

$$(Rx)_j = x_{1+(j-2) \bmod n}, \quad j = 1, \dots, n. \quad (3.1)$$

For example,

$$R_3 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad R_3 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_3 \\ x_1 \\ x_2 \end{pmatrix}.$$

Also, $R^n = I$ and $R^{-1} = R^\top$. We want to find a positive matrix $K \in \mathbb{R}_{++}^{m \times n}$, positive vectors $r \in \mathbb{R}_{++}^m$, $c \in \mathbb{R}_{++}^n$ with $\sum_i r_i = \sum_j c_j = 1$ such that

$$S_K^{r,c}(Rz) = RS_K^{r,c}(z), \quad \forall z > 0, \quad (3.2)$$

and a positive vector $q \in \mathbb{R}_{++}^n$ with full cyclic period n , i.e.,

$$R^k q = q \iff k \equiv 0 \pmod{n},$$

such that

$$S_K^{r,c} \left(\frac{(R^{-1}q)^2}{R^{-2}q} \right) = \alpha q, \quad \text{for some } \alpha > 0, \quad (3.3)$$

The rotations describe cumulative column scalings relative to the base matrix K , not the scalings q^k relative to M in Section 2. Our eventual initialization of (BDRS-MS) is $X^0 = M = K \operatorname{diag}(R^{-1}q)$. Thus, setting

$$\tilde{v}^k := (R^{-1}q) \odot q^k,$$

we have

$$X^k = \operatorname{diag}(p^k) K \operatorname{diag}(\tilde{v}^k).$$

Substituting this representation into (BDRS-MS), or equivalently changing the base matrix in (2.6), gives

$$\tilde{v}^{k+1} = S_K^{r,c} \left(\frac{(\tilde{v}^k)^2}{\tilde{v}^{k-1}} \right).$$

With the auxiliary initialization $v^{-1} = \frac{R^{-1}q}{R^{-2}q}$, we have

$$\tilde{v}^{-1} = R^{-2}q, \quad \tilde{v}^0 = R^{-1}q.$$

We prescribe the successive projective directions

$$\dots, R^{-2}q, R^{-1}q, q, Rq, R^2q, \dots,$$

where a projective direction consists of all positive scalar multiples of a vector. Equation (3.3) gives $\tilde{v}^1 = \alpha q$. Since permutations commute with componentwise squares and division, we have

$$R\left(\frac{(R^{-1}q)^2}{R^{-2}q}\right) = \frac{q^2}{R^{-1}q}.$$

Using (3.2) and (3.3), we therefore obtain

$$S_K^{r,c}\left(\frac{q^2}{R^{-1}q}\right) = \alpha Rq.$$

The actual second update includes a scalar factor. The positive homogeneity $S_K^{r,c}(\lambda t) = \lambda S_K^{r,c}(t)$ for $\lambda > 0$, which follows directly from (2.3), gives

$$\tilde{v}^2 = S_K^{r,c}\left(\alpha^2 \frac{q^2}{R^{-1}q}\right) = \alpha^3 Rq.$$

Repeating the argument rotates the directions, while the scalar factors need not be periodic. The direct induction below keeps these factors and proves that the matrix iterates themselves form an exact cycle.

3.1 Constructing the cyclic symmetry

Let d denote a divisor of n . For a positive vector $s = (s_1, \dots, s_d)^\top$, define the $d \times n$ cyclic block

$$[C_d(s)]_{ij} = s_{1+((j-i) \bmod d)}, \quad 1 \leq i \leq d, \quad 1 \leq j \leq n. \quad (3.4)$$

The case $d = n$ is an ordinary circulant matrix.

Construct K . Choose N positive vectors

$$s^{(\ell)} \in \mathbb{R}_{++}^{d_\ell}, \quad \ell = 1, \dots, N, \quad (3.5)$$

where d_ℓ 's are all divisors of n . Define $K \in \mathbb{R}_{++}^{m \times n}$ as the following block matrix:

$$K = \begin{pmatrix} K^{(1)} \\ K^{(2)} \\ \vdots \\ K^{(N)} \end{pmatrix}, \quad m = \sum_{\ell=1}^N d_\ell, \quad \text{where } K^{(\ell)} = C_{d_\ell}(s^{(\ell)}). \quad (3.6)$$

Parameterize r . Introduce a positive weight vector

$$w \in \mathbb{R}_{++}^N, \quad \text{with} \quad \sum_{\ell=1}^N w_\ell = 1. \quad (3.7)$$

Define r as a block vector, and the i -th entry of its ℓ -th block $r^{(\ell)}$ is given by

$$r_{(\ell,i)} = r_i^{(\ell)} = \frac{w_\ell}{d_\ell}, \quad i = 1, \dots, d_\ell, \quad \ell = 1, \dots, N. \quad (3.8)$$

The weights w_ℓ 's are not fixed yet. Every choice satisfying (3.7) preserves the required row symmetry; we select weights satisfying the additional interpolation condition in the next subsection.

Construct c . c is simply chosen as

$$c = \frac{1}{n} \mathbf{1}_n. \quad (3.9)$$

Recall that R_d is the d -dimensional right shift. Define

$$\Pi = \text{diag}(R_{d_1}, \dots, R_{d_N}) \in \mathbb{R}^{m \times m}. \quad (3.10)$$

The following lemma shows that the constructions in (3.6)-(3.9) guarantee (3.2).

Lemma 3.1. *Constructions (3.6)-(3.9) imply*

$$KR = \Pi K, \quad K^\top \Pi = RK^\top, \quad \Pi r = r, \quad Rc = c, \quad (3.11)$$

which further leads to (3.2).

Proof. Fix $\ell \in \{1, \dots, N\}$, $i \in \{1, \dots, d_\ell\}$, and $j \in \{1, \dots, n\}$. In the calculation below, $j+1$ is interpreted cyclically modulo n and $i-1$ cyclically modulo d_ℓ . The condition $d_\ell \mid n$ ensures consistency when a column index wraps around. We compare the corresponding entries:

$$(KR)_{i,j}^{(\ell)} = K_{i,j+1}^{(\ell)} = s_{1+((j+1-i) \bmod d_\ell)}^{(\ell)} = K_{i-1,j}^{(\ell)} = (\Pi K)_{i,j}^{(\ell)}.$$

This proves the first identity in (3.11).

Transposing $KR = \Pi K$ gives $R^\top K^\top = K^\top \Pi^\top$. Since R and Π are permutation matrices, they are orthogonal. Multiplying on the left by R and on the right by Π yields $K^\top \Pi = RK^\top$, the second identity in (3.11).

The third and fourth identities in (3.11) follow immediately from the definitions of Π , R , r and c .

Now we show that (3.11) leads to (3.2). First, note that

$$\frac{r}{KRz} = \frac{r}{\Pi Kz} = \frac{\Pi r}{\Pi Kz} = \Pi \frac{r}{Kz},$$

where the first equality is from the first identity in (3.11), the second equality is from the third identity in (3.11), and the last equality is because Π is a permutation matrix. Thus

$$S_K^{r,c}(Rz) = \frac{c}{K^\top \Pi(r/(Kz))} = \frac{c}{RK^\top(r/(Kz))} = \frac{Rc}{RK^\top(r/(Kz))} = R \frac{c}{K^\top(r/(Kz))} = RS_K^{r,c}(z),$$

where the second equality uses the second identity in (3.11), the third equality uses the fourth identity in (3.11). This proves (3.2). $\square$

3.2 Enforcing the target transition by choosing the block weights

Fix the block matrices and a positive vector q with full cyclic period n . For ease of presentation, define

$$z = \frac{(R^{-1}q)^2}{R^{-2}q}. \quad (3.12)$$

The desired identity (3.3) is $S_K^{r,c}(z) = \alpha q$. Define $B \in \mathbb{R}_{++}^{n \times N}$ by

$$B_{j\ell} = \frac{1}{d_\ell} \sum_{i=1}^{d_\ell} \frac{K_{ij}^{(\ell)} z_j}{(K^{(\ell)} z)_i}, \quad j = 1, \dots, n, \quad \ell = 1, \dots, N. \quad (3.13)$$

Each column of B sums to one, because

$$\sum_{j=1}^n B_{j\ell} = \frac{1}{d_\ell} \sum_{i=1}^{d_\ell} \frac{\sum_{j=1}^n K_{ij}^{(\ell)} z_j}{(K^{(\ell)} z)_i} = 1.$$

For any weights satisfying (3.7), define r by (3.8). The j -th column sum of the row-normalized matrix

$$\text{diag}\left(\frac{r}{Kz}\right) K \text{diag}(z)$$

is

$$\begin{aligned} z_j \left[K^\top \left(\frac{r}{Kz} \right) \right]_j &= \sum_{\ell=1}^N \frac{w_\ell}{d_\ell} \sum_{i=1}^{d_\ell} \frac{K_{ij}^{(\ell)} z_j}{(K^{(\ell)} z)_i} \\ &= \sum_{\ell=1}^N B_{j\ell} w_\ell = (Bw)_j. \end{aligned}$$

Thus Bw is the actual intermediate column marginal and

$$\mathbf{1}_n^\top Bw = \mathbf{1}_N^\top w = 1, \quad S_K^{r,c}(z)_j = \frac{c_j z_j}{(Bw)_j} = \frac{z_j}{n(Bw)_j}.$$

Consequently, the target identity $S_K^{r,c}(z) = \alpha q$ requires

$$(Bw)_j = \frac{z_j}{n\alpha q_j}.$$

Summing these required equalities over j determines α . We therefore define the scalar α and the target vector p independently of w by

$$\alpha = \frac{1}{n} \sum_{j=1}^n \frac{z_j}{q_j}, \quad p_j = \frac{z_j/q_j}{\sum_{j'=1}^n z_{j'}/q_{j'}}, \quad j = 1, \dots, n. \quad (3.14)$$

Then $p > 0$, $\mathbf{1}_n^\top p = 1$, and $p_j = z_j/(n\alpha q_j)$. With these definitions,

$$S_K^{r,c}(z) = \alpha q \iff Bw = p.$$

3.3 The Complete Construction

The complete construction process is as follows.

- (i) Choose $n \geq 2$, a number of blocks $N \geq 1$, divisors $d_\ell \mid n$, and positive seeds $s^{(\ell)}$ as in (3.5); construct K by (3.6).
- (ii) Define c by (3.9).
- (iii) Choose $q > 0$ with full cyclic period n , and compute z , α , and p by (3.12) and (3.14).
- (iv) Compute B by (3.13) and find a strictly positive solution w of

$$Bw = p. \quad (3.15)$$

Because every column of B sums to one, this equality automatically gives $\mathbf{1}_N^\top w = \mathbf{1}_n^\top p = 1$.

(v) Define r by (3.8) using these weights.

If the equation in step (iv) has no positive solution, the chosen seeds or q must be changed.

Theorem 3.2. *Let $n \geq 2$, and suppose steps (i)–(v) above produce a solution $w > 0$ of (3.15). Initialize (BDRS-MS), i.e., (2.2), with $X^0 = M$ and*

$$M = K \operatorname{diag}(R^{-1}q), \quad v^{-1} = \frac{R^{-1}q}{R^{-2}q}. \quad (3.16)$$

Define

$$Y_k = \alpha \operatorname{diag}\left(\frac{r}{KR^k z}\right) K \operatorname{diag}(R^k q), \quad k \geq 0. \quad (3.17)$$

Then the sequence $\{X^k, v^k\}$ generated by (BDRS-MS) satisfies

$$X^{k+1} = Y_k, \quad v^k = \alpha^{k+1} \frac{R^k q}{R^{k-1}q}, \quad \forall k \geq 0. \quad (3.18)$$

Moreover, the matrices $Y_0, \dots, Y_{n-1}$ are pairwise distinct, and $Y_{k+n} = Y_k$. Hence the sequence $\{X^k\}$ does not converge.

Proof. Note that $Bw = p$ is equivalent to $S_K^{r,c}(z) = \alpha q$. Using (3.2) we have

$$\frac{c}{K^\top \left(\frac{r}{KR^k z}\right)} = S_K^{r,c}(R^k z) = R^k S_K^{r,c}(z) = \alpha R^k q, \quad \forall k \geq 0. \quad (3.19)$$

We prove (3.18) by induction. First, when $k = 0$, by (3.16) we have

$$X^0 v^{-1} = M v^{-1} = K \operatorname{diag}(R^{-1}q) \frac{R^{-1}q}{R^{-2}q} = K \frac{(R^{-1}q)^2}{R^{-2}q} = Kz.$$

Thus $u^0 = r/(Kz)$. Applying (3.19) with $k = 0$,

$$v^0 = \frac{c}{(X^0)^\top u^0} = \frac{c}{\operatorname{diag}(R^{-1}q) K^\top (r/(Kz))} = \frac{1}{R^{-1}q} \odot \frac{c}{K^\top (r/(Kz))} = \alpha \frac{q}{R^{-1}q}. \quad (3.20)$$

We then have

$$X^1 = \operatorname{diag}(u^0) X^0 \operatorname{diag}(v^0) = \operatorname{diag}\left(\frac{r}{Kz}\right) K \operatorname{diag}(R^{-1}q) \operatorname{diag}\left(\frac{\alpha q}{R^{-1}q}\right) = Y_0,$$

which together with (3.20) proves (3.18) for $k = 0$.

Fix $k \geq 1$ and assume that (3.18) holds for $k - 1$, i.e., $X^k = Y_{k-1}$, $v^{k-1} = \alpha^k R^{k-1}q / R^{k-2}q$, and we then prove that it holds for k . Since

$$\frac{(R^{k-1}q)^2}{R^{k-2}q} = \frac{(R^k R^{-1}q)^2}{R^k R^{-2}q} = R^k \frac{(R^{-1}q)^2}{R^{-2}q} = R^k z, \quad (3.21)$$

we have

$$\begin{aligned} X^k v^{k-1} &= \alpha \operatorname{diag}\left(\frac{r}{KR^{k-1}z}\right) K \operatorname{diag}(R^{k-1}q) \alpha^k R^{k-1}q / R^{k-2}q \\ &= \alpha^{k+1} \operatorname{diag}\left(\frac{r}{KR^{k-1}z}\right) K \frac{(R^{k-1}q)^2}{R^{k-2}q} \\ &= \alpha^{k+1} r \odot \frac{KR^k z}{KR^{k-1}z}, \end{aligned}$$

and

$$u^k = \frac{r}{X^k v^{k-1}} = \alpha^{-(k+1)} \frac{K R^{k-1} z}{K R^k z}.$$

It then follows that

$$\begin{aligned} (X^k)^\top u^k &= \alpha \operatorname{diag}(R^{k-1} q) K^\top \operatorname{diag}\left(\frac{r}{K R^{k-1} z}\right) \alpha^{-(k+1)} \frac{K R^{k-1} z}{K R^k z} \\ &= \alpha^{-k} \operatorname{diag}(R^{k-1} q) K^\top \left(\frac{r}{K R^k z}\right) \\ &= \alpha^{-(k+1)} c \odot \frac{R^{k-1} q}{R^k q}, \end{aligned}$$

where the last equality uses (3.19). We then have

$$v^k = \frac{c}{(X^k)^\top u^k} = \alpha^{k+1} \frac{R^k q}{R^{k-1} q}. \quad (3.22)$$

It then follows that

$$\begin{aligned} X^{k+1} &= \operatorname{diag}(u^k) X^k \operatorname{diag}(v^k) \\ &= \alpha^{-(k+1)} \operatorname{diag}\left(\frac{K R^{k-1} z}{K R^k z}\right) \alpha \operatorname{diag}\left(\frac{r}{K R^{k-1} z}\right) K \operatorname{diag}(R^{k-1} q) \alpha^{k+1} \operatorname{diag}\left(\frac{R^k q}{R^{k-1} q}\right) \\ &= \alpha \operatorname{diag}\left(\frac{r}{K R^k z}\right) K \operatorname{diag}(R^k q) \\ &= Y_k, \end{aligned}$$

which together with (3.22) completes the induction and thus proves (3.18) for all $k \geq 0$.

Since $R^n = I$, formula (3.17) gives $Y_{k+n} = Y_k$.

We now prove that $Y_0, \dots, Y_{n-1}$ are pairwise distinct. Suppose that $Y_k = Y_\ell$ for some $0 \leq k, \ell \leq n-1$. The (i, j) -th entry gives

$$\alpha \frac{r_i K_{ij} (R^k q)_j}{(K R^k z)_i} = \alpha \frac{r_i K_{ij} (R^\ell q)_j}{(K R^\ell z)_i}, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n.$$

Since $\alpha > 0$, $r_i > 0$, and $K_{ij} > 0$, we can cancel these common factors and obtain

$$\frac{(R^k q)_j}{(K R^k z)_i} = \frac{(R^\ell q)_j}{(K R^\ell z)_i}, \quad 1 \leq i \leq m, \quad 1 \leq j \leq n.$$

Fix any row index i . The scalar

$$\gamma := \frac{(K R^k z)_i}{(K R^\ell z)_i} > 0$$

is independent of j , and

$$(R^k q)_j = \gamma (R^\ell q)_j, \quad j = 1, \dots, n.$$

Hence

$$R^k q = \gamma R^\ell q.$$

Because $R^k q$ and $R^\ell q$ are merely permutations of the entries of q , they have the same sum. Therefore

$$\mathbf{1}^\top q = \mathbf{1}^\top R^k q = \gamma \mathbf{1}^\top R^\ell q = \gamma \mathbf{1}^\top q.$$

Since $q > 0$, we have $\mathbf{1}^\top q > 0$, and thus

$$\gamma = 1.$$

Consequently,

$$R^k q = R^\ell q,$$

so that

$$R^{k-\ell} q = q.$$

By the full-period assumption on q ,

$$k - \ell \equiv 0 \pmod{n}.$$

Since $k, \ell \in \{0, 1, \dots, n-1\}$, this implies $k = \ell$. Therefore, $Y_0, \dots, Y_{n-1}$ are pairwise distinct. $\square$

3.4 A feasible scaling exists

We also verify explicitly that the constructed matrix-scaling problem is feasible. Let $\sigma_\ell = \sum_{j=1}^{d_\ell} s_j^{(\ell)}$. Every row of $K^{(\ell)}$ sums to $(n/d_\ell)\sigma_\ell$, and every column sums to σ_ℓ . Consequently, $X^\star$ defined as

$$X^\star = \begin{pmatrix} (X^\star)^{(1)} \\ (X^\star)^{(2)} \\ \vdots \\ (X^\star)^{(N)} \end{pmatrix}, \quad \text{with} \quad (X^\star)_{ij}^{(\ell)} = \frac{w_\ell}{n\sigma_\ell} K_{ij}^{(\ell)}, \ell = 1, \dots, N$$

has row sums $r_{(\ell,i)} = w_\ell/d_\ell$ and column sums $\sum_{\ell=1}^N w_\ell/n = 1/n$, hence marginals r, c . It belongs to the diagonal orbit of M , since

$$X^\star = \text{diag}(\tilde{d}) M \text{diag}\left(\frac{1}{R^{-1}q}\right),$$

where $\tilde{d}$ is a block vector with

$$\tilde{d}_i^{(\ell)} = \frac{w_\ell}{n\sigma_\ell}, i = 1, \dots, d_\ell, \ell = 1, \dots, N.$$

Therefore, a feasible scaling indeed exists.

4 A Concrete 36×6 Counterexample for (BDRS-MS)

In this section, we provide a concrete counterexample that satisfies all the requirements discussed in Section 3.

4.1 The counterexample

The counterexample has the following dimensions: $m = 36$, $n = 6$, $N = 6$, $d_1 = d_2 = \dots = d_N = 6$. The seeds in (3.5) are specified by $s_j^{(\ell)} = 2^{E_{\ell j}}$ for $\ell, j = 1, \dots, 6$, where

$$E = \begin{pmatrix} 20 & 0 & 18 & 10 & 16 & 18 \\ 0 & 21 & 24 & 26 & 29 & 16 \\ 0 & 28 & 3 & 28 & 11 & 28 \\ 0 & 0 & 1 & 21 & 34 & 10 \\ 3 & 10 & 0 & 20 & 26 & 32 \\ 1 & 4 & 0 & 1 & 2 & 3 \end{pmatrix}. \quad (4.1)$$

The matrix K in (3.6) is a 36×6 matrix with 6 blocks. The (i, j) -th entry of its ℓ -th block is:

$$K_{ij}^{(\ell)} = 2^{E_{\ell, 1+((j-i) \bmod 6)}}, \quad 1 \leq \ell, i, j \leq 6.$$

More explicitly,

$$K = \begin{pmatrix} K^{(1)} \\ K^{(2)} \\ K^{(3)} \\ K^{(4)} \\ K^{(5)} \\ K^{(6)} \end{pmatrix} \in \mathbb{R}_{++}^{36 \times 6},$$

where

$$K^{(1)} = \begin{pmatrix} 2^{20} & 1 & 2^{18} & 2^{10} & 2^{16} & 2^{18} \\ 2^{18} & 2^{20} & 1 & 2^{18} & 2^{10} & 2^{16} \\ 2^{16} & 2^{18} & 2^{20} & 1 & 2^{18} & 2^{10} \\ 2^{10} & 2^{16} & 2^{18} & 2^{20} & 1 & 2^{18} \\ 2^{18} & 2^{10} & 2^{16} & 2^{18} & 2^{20} & 1 \\ 1 & 2^{18} & 2^{10} & 2^{16} & 2^{18} & 2^{20} \end{pmatrix},$$

$$K^{(2)} = \begin{pmatrix} 1 & 2^{21} & 2^{24} & 2^{26} & 2^{29} & 2^{16} \\ 2^{16} & 1 & 2^{21} & 2^{24} & 2^{26} & 2^{29} \\ 2^{29} & 2^{16} & 1 & 2^{21} & 2^{24} & 2^{26} \\ 2^{26} & 2^{29} & 2^{16} & 1 & 2^{21} & 2^{24} \\ 2^{24} & 2^{26} & 2^{29} & 2^{16} & 1 & 2^{21} \\ 2^{21} & 2^{24} & 2^{26} & 2^{29} & 2^{16} & 1 \end{pmatrix},$$

$$K^{(3)} = \begin{pmatrix} 1 & 2^{28} & 2^3 & 2^{28} & 2^{11} & 2^{28} \\ 2^{28} & 1 & 2^{28} & 2^3 & 2^{28} & 2^{11} \\ 2^{11} & 2^{28} & 1 & 2^{28} & 2^3 & 2^{28} \\ 2^{28} & 2^{11} & 2^{28} & 1 & 2^{28} & 2^3 \\ 2^3 & 2^{28} & 2^{11} & 2^{28} & 1 & 2^{28} \\ 2^{28} & 2^3 & 2^{28} & 2^{11} & 2^{28} & 1 \end{pmatrix},$$

$$K^{(4)} = \begin{pmatrix} 1 & 1 & 2 & 2^{21} & 2^{34} & 2^{10} \\ 2^{10} & 1 & 1 & 2 & 2^{21} & 2^{34} \\ 2^{34} & 2^{10} & 1 & 1 & 2 & 2^{21} \\ 2^{21} & 2^{34} & 2^{10} & 1 & 1 & 2 \\ 2 & 2^{21} & 2^{34} & 2^{10} & 1 & 1 \\ 1 & 2 & 2^{21} & 2^{34} & 2^{10} & 1 \end{pmatrix},$$

$$K^{(5)} = \begin{pmatrix} 2^3 & 2^{10} & 1 & 2^{20} & 2^{26} & 2^{32} \\ 2^{32} & 2^3 & 2^{10} & 1 & 2^{20} & 2^{26} \\ 2^{26} & 2^{32} & 2^3 & 2^{10} & 1 & 2^{20} \\ 2^{20} & 2^{26} & 2^{32} & 2^3 & 2^{10} & 1 \\ 1 & 2^{20} & 2^{26} & 2^{32} & 2^3 & 2^{10} \\ 2^{10} & 1 & 2^{20} & 2^{26} & 2^{32} & 2^3 \end{pmatrix},$$

$$K^{(6)} = \begin{pmatrix} 2 & 2^4 & 1 & 2 & 2^2 & 2^3 \\ 2^3 & 2 & 2^4 & 1 & 2 & 2^2 \\ 2^2 & 2^3 & 2 & 2^4 & 1 & 2 \\ 2 & 2^2 & 2^3 & 2 & 2^4 & 1 \\ 1 & 2 & 2^2 & 2^3 & 2 & 2^4 \\ 2^4 & 1 & 2 & 2^2 & 2^3 & 2 \end{pmatrix}.$$

The vector $q \in \mathbb{R}_{++}^6$ is chosen as

$$q = \begin{pmatrix} 2^8 \\ 2^{13} \\ 2^{15} \\ 2^{11} \\ 2^0 \\ 2^2 \end{pmatrix},$$

and then $z \in \mathbb{R}_{++}^6$ can be computed by

$$z = \frac{(R^{-1}q)^2}{R^{-2}q} = \begin{pmatrix} 2^{11} \\ 2^{19} \\ 2^{22} \\ 2^{-2} \\ 2^{-4} \\ 2^3 \end{pmatrix}.$$

From (3.14), we get

$$p = \frac{1}{1655297} \begin{pmatrix} 2^{16} \\ 2^{19} \\ 2^{20} \\ 1 \\ 2^9 \\ 2^{14} \end{pmatrix} = \frac{1}{1655297} \begin{pmatrix} 65536 \\ 524288 \\ 1048576 \\ 1 \\ 512 \\ 16384 \end{pmatrix},$$

and

$$\alpha = \frac{1655297}{6 \cdot 2^{13}} = \frac{1655297}{49152}.$$

We then use (3.13) to compute $B \in \mathbb{R}_{++}^{6 \times 6}$. With K and z given above, B can be computed exactly. For readability, we provide a numerical approximation of B below, but note that all calculations later (e.g., w) are done exactly.

$$B \approx \begin{pmatrix} 8.17 \times 10^{-4} & 1.6163 \times 10^{-1} & 2.44 \times 10^{-4} & 1.6667 \times 10^{-1} & 1.6659 \times 10^{-1} & 8.52 \times 10^{-4} \\ 3.3853 \times 10^{-1} & 1.8165 \times 10^{-1} & 4.9998 \times 10^{-1} & 1.6675 \times 10^{-1} & 1.6987 \times 10^{-1} & 1.9853 \times 10^{-1} \\ 6.6064 \times 10^{-1} & 6.5656 \times 10^{-1} & 4.9977 \times 10^{-1} & 3.3455 \times 10^{-1} & 4.9944 \times 10^{-1} & 8.0061 \times 10^{-1} \\ 1.17 \times 10^{-7} & 2.73 \times 10^{-7} & 2.38 \times 10^{-7} & 1.62 \times 10^{-4} & 2.52 \times 10^{-6} & 9.98 \times 10^{-8} \\ 6.01 \times 10^{-8} & 3.12 \times 10^{-7} & 7.45 \times 10^{-9} & 1.6521 \times 10^{-1} & 3.02 \times 10^{-5} & 1.96 \times 10^{-8} \\ 1.06 \times 10^{-5} & 1.60 \times 10^{-4} & 7.63 \times 10^{-6} & 1.6666 \times 10^{-1} & 1.6407 \times 10^{-1} & 2.67 \times 10^{-6} \end{pmatrix}.$$

Finally, the weight vector $w \in \mathbb{R}^N$, with $N = 6$, is computed by

$$w = B^{-1}p. \tag{4.2}$$

Exact rational Gaussian elimination and integer cross-multiplication certify that

$$\frac{6}{10^{11}} < \det B < \frac{7}{10^{11}}, \quad \frac{\beta_\ell}{10^{12}} < w_\ell < \frac{\beta_\ell + 1}{10^{12}}, \tag{4.3}$$

where the integer lower-bound numerators β_ℓ and the corresponding w_ℓ are given below. The decimals are explanatory, not the definition of w_ℓ .

ℓ	β_ℓ	Approximation of w_ℓ
1	546062288830	0.546062288830524
2	179725833975	0.179725833975301
3	154299387166	0.154299387166565
4	1860993790	0.001860993790767
5	58217179004	0.058217179004256
6	59834317232	0.059834317232588

Therefore, $w > 0$. Moreover, $Bw = p$ and the column sums of B imply $\mathbf{1}^\top w = 1$.

With these preparations, we choose the following data for the matrix scaling problem:

$$\begin{aligned} r_{(\ell,i)} &= w_\ell/6, \ell = 1, \dots, 6, i = 1, \dots, 6, \\ c &= \mathbf{1}_6/6, \\ M &= K \operatorname{diag}(8192, 32768, 2048, 1, 4, 256), \end{aligned}$$

and the initial point in the (BDRS-MS) algorithm is chosen as:

$$X^0 = M, \quad v^{-1} = (1/4, 16, 2048, 1/4, 1/64, 1/32)^\top.$$

Theorem 3.2 proves that there is a six-cycle in (BDRS-MS) for solving this matrix scaling problem.

4.2 The iterates are not feasible

The nonconstant cycle already proves that the matrix iterates do not converge. We now establish the additional fact that the cycle remains uniformly separated from the prescribed row marginal. Evaluating Y_0 from (3.17) with exact rational arithmetic and comparing fractions by integer cross-multiplication gives

$$\frac{206}{1000} < \|Y_0 \mathbf{1} - r\|_\infty < \frac{208}{1000}. \quad (4.4)$$

Thus Y_0 is not row feasible. The same exact-arithmetic verifier certifies (4.4). The following lemma transports this residual to every phase of the cycle.

Lemma 4.1. *For any $k \geq 0$,*

$$Y_k = \Pi^k Y_0 R^{-k}.$$

Proof. Recall that

$$Y_k = \alpha \operatorname{diag}\left(\frac{r}{KR^k z}\right) K \operatorname{diag}(R^k q).$$

From $KR = \Pi K$, we obtain

$$KR^k = \Pi^k K,$$

and hence

$$KR^k z = \Pi^k K z.$$

Since $\Pi r = r$, we also have $\Pi^k r = r$. Because Π^k is a permutation matrix and componentwise division commutes with permutations,

$$\frac{r}{KR^k z} = \frac{r}{\Pi^k K z} = \Pi^k \left(\frac{r}{K z} \right).$$

Therefore,

$$\operatorname{diag}\left(\frac{r}{KR^k z}\right) = \operatorname{diag}\left(\Pi^k \frac{r}{K z}\right).$$

For any permutation matrix P and vector x ,

$$\operatorname{diag}(Px) = P \operatorname{diag}(x) P^{-1}.$$

Applying this identity with $P = \Pi^k$ gives

$$\operatorname{diag}\left(\frac{r}{KR^k z}\right) = \Pi^k \operatorname{diag}\left(\frac{r}{K z}\right) \Pi^{-k}.$$

Hence

$$Y_k = \alpha \Pi^k \operatorname{diag}\left(\frac{r}{Kz}\right) \Pi^{-k} K \operatorname{diag}(R^k q).$$

Again using $KR^k = \Pi^k K$, we have

$$\Pi^{-k} K = KR^{-k}.$$

Thus

$$Y_k = \alpha \Pi^k \operatorname{diag}\left(\frac{r}{Kz}\right) KR^{-k} \operatorname{diag}(R^k q).$$

Since R^k is also a permutation matrix,

$$\operatorname{diag}(R^k q) = R^k \operatorname{diag}(q) R^{-k},$$

which is equivalent to

$$R^{-k} \operatorname{diag}(R^k q) = \operatorname{diag}(q) R^{-k}.$$

Therefore,

$$\begin{aligned} Y_k &= \alpha \Pi^k \operatorname{diag}\left(\frac{r}{Kz}\right) K \operatorname{diag}(q) R^{-k} \\ &= \Pi^k \left[\alpha \operatorname{diag}\left(\frac{r}{Kz}\right) K \operatorname{diag}(q) \right] R^{-k} \\ &= \Pi^k Y_0 R^{-k}, \end{aligned}$$

which completes the proof. $\square$

By Lemma 4.1, $R^{-k} \mathbf{1}_n = \mathbf{1}_n$, and $\Pi^k r = r$, we have

$$Y_k \mathbf{1}_n - r = \Pi^k Y_0 R^{-k} \mathbf{1}_n - r = \Pi^k (Y_0 \mathbf{1}_n - r).$$

Permutation matrices preserve the infinity norm. Hence

$$\|Y_k \mathbf{1}_n - r\|_\infty = \|Y_0 \mathbf{1}_n - r\|_\infty, \quad k \geq 0.$$

In particular, (4.4) holds at every phase, and $X^k = Y_{k-1}$ is infeasible for every $k \geq 1$. The initial matrix is also infeasible: every entry of M is a positive integer, so $(M \mathbf{1}_6)_i \geq 6$, whereas $r_i < 1$. Thus

$$\|X^0 \mathbf{1}_6 - r\|_\infty = \|M \mathbf{1}_6 - r\|_\infty > 5.$$

This last estimate concerns the initial matrix, not the cycle states.

5 Counterexamples for (BDRS-OT)

In this section, we show that a similar counterexample can be constructed for (BDRS-OT), i.e., (1.11). That is, BDRS can fail for a feasible unregularized optimal transport problem.

Throughout this section, M, r, c and v^{-1} are the data of the 36×6 example in Section 4, so $m = 36$ and $n = 6$. Let $\{X^k, u^k, v^k\}$ be the sequence actually generated by (BDRS-MS) with $X^0 = M$ and this auxiliary initialization. By Theorem 3.2,

$$X^k = Y_{k-1} \quad (k \geq 1).$$

Let $J = \mathbf{1}_m \mathbf{1}_n^\top$ be the $m \times n$ all-one matrix. Define

$$\hat{C} = \begin{pmatrix} 2J & 3J \\ 3J & 2J \end{pmatrix}, \quad \hat{r} = \frac{1}{2} \begin{pmatrix} r \\ r \end{pmatrix}, \quad \hat{c} = \frac{1}{2} \begin{pmatrix} c \\ c \end{pmatrix}, \quad (5.1)$$

and the initialization

$$\hat{X}^0 = \frac{1}{4} \begin{pmatrix} M & M \\ M & M \end{pmatrix}, \quad \hat{v}^{-1} = \begin{pmatrix} v^{-1} \\ v^{-1} \end{pmatrix}. \quad (5.2)$$

The lifted cost and initial matrix have size 72×12 and are strictly positive. The lifted OT feasible set is nonempty: the positive matrix $\hat{r}\hat{c}^\top$ has marginals $\hat{r}, \hat{c}$. As a nonempty compact transportation polytope with a finite linear objective, it also has an optimal solution. Furthermore, the cost cannot be removed by adding row and column potentials. Choosing one row from each row group and one column from each column group gives the mixed difference

$$2 + 2 - 3 - 3 = -2 \neq 0,$$

whereas every matrix of the form $a\mathbf{1}^\top + \mathbf{1}b^\top$ has all mixed differences equal to zero.

Theorem 5.1 (Nonconvergence of (BDRS-OT)). *Apply (BDRS-OT) to the optimal transport problem with data defined in (5.1) and initialization given in (5.2). For any fixed $\eta > 0$, set $\rho = e^{-1/\eta} \in (0, 1)$ and denote*

$$\hat{H} = e^{-\hat{C}/\eta} = \rho^2 \begin{pmatrix} J & \rho J \\ \rho J & J \end{pmatrix}, \quad (5.3)$$

where the exponential is componentwise. The sequence $\{\hat{X}^k, \hat{u}^k, \hat{v}^k\}$ generated by (BDRS-OT)

$$\hat{u}^k := \frac{\hat{r}}{(\hat{X}^k \odot \hat{H})\hat{v}^{k-1}} \quad (5.4a)$$

$$\hat{v}^k := \frac{\hat{c}}{(\hat{X}^k \odot \hat{H})^\top \hat{u}^k} \quad (5.4b)$$

$$\hat{X}^{k+1} := \text{diag}(\hat{u}^k)(\hat{X}^k \odot \hat{H})\text{diag}(\hat{v}^k), \quad (5.4c)$$

satisfies

$$\hat{X}^k = \frac{1}{2(1+\rho^k)} \begin{pmatrix} X^k & \rho^k X^k \\ \rho^k X^k & X^k \end{pmatrix}, \quad \hat{v}^{k-1} = \begin{pmatrix} v^{k-1} \\ v^{k-1} \end{pmatrix}, \quad k \geq 0. \quad (5.5)$$

The sequence $\{\hat{X}^k\}$ does not converge. More precisely,

$$\|\hat{X}^k \mathbf{1} - \hat{r}\|_\infty = \frac{1}{2} \|X^k \mathbf{1} - r\|_\infty > 0.103, \quad \forall k \geq 0. \quad (5.6)$$

The set of subsequential limits consists exactly of the six distinct matrices

$$Z_\ell = \frac{1}{2} \begin{pmatrix} Y_{\ell-1} & 0 \\ 0 & Y_{\ell-1} \end{pmatrix}, \quad \ell = 1, \dots, 6. \quad (5.7)$$

All six limits are infeasible.

Proof. We prove both identities in (5.5) simultaneously by induction on the state $(\hat{X}^k, \hat{v}^{k-1})$. At $k = 0$, the matrix identity follows from $X^0 = M$, $\hat{X}^0 = \frac{1}{4} \begin{pmatrix} M & M \\ M & M \end{pmatrix}$, and $\rho^0 = 1$. The auxiliary identity follows directly from $\hat{v}^{-1} = \begin{pmatrix} v^{-1} \\ v^{-1} \end{pmatrix}$ in (5.2). Assume now that both identities hold at some $k \geq 0$. Then

$$\hat{X}^k \odot \hat{H} = \frac{\rho^2}{2(1+\rho^k)} \begin{pmatrix} X^k & \rho^{k+1} X^k \\ \rho^{k+1} X^k & X^k \end{pmatrix}.$$

Multiplication by the duplicated vector $\hat{v}^{k-1}$ gives

$$(\hat{X}^k \odot \hat{H})\hat{v}^{k-1} = \frac{\rho^2(1+\rho^{k+1})}{2(1+\rho^k)} \begin{pmatrix} X^k v^{k-1} \\ X^k v^{k-1} \end{pmatrix}.$$

Dividing $\hat{r}$ defined in (5.1) by this vector and using (5.4a), we get

$$\hat{u}^k = \frac{1 + \rho^k}{\rho^2(1 + \rho^{k+1})} \begin{pmatrix} u^k \\ u^k \end{pmatrix}.$$

Therefore,

$$(\hat{X}^k \odot \hat{H})^\top \hat{u}^k = \frac{1}{2} \begin{pmatrix} (X^k)^\top u^k \\ (X^k)^\top u^k \end{pmatrix}.$$

Dividing $\hat{c}$ defined in (5.1) by this vector and using (5.4b), we get

$$\hat{v}^k = \begin{pmatrix} v^k \\ v^k \end{pmatrix}. \quad (5.8)$$

Finally, from (5.4c) we have

$$\begin{aligned} \hat{X}^{k+1} &= \text{diag}(\hat{u}^k)(\hat{X}^k \odot \hat{H}) \text{diag}(\hat{v}^k) \\ &= \frac{1}{2(1 + \rho^{k+1})} \begin{pmatrix} \text{diag}(u^k) & 0 \\ 0 & \text{diag}(u^k) \end{pmatrix} \begin{pmatrix} X^k & \rho^{k+1} X^k \\ \rho^{k+1} X^k & X^k \end{pmatrix} \begin{pmatrix} \text{diag}(v^k) & 0 \\ 0 & \text{diag}(v^k) \end{pmatrix} \\ &= \frac{1}{2(1 + \rho^{k+1})} \begin{pmatrix} X^{k+1} & \rho^{k+1} X^{k+1} \\ \rho^{k+1} X^{k+1} & X^{k+1} \end{pmatrix}, \end{aligned} \quad (5.9)$$

where the last equality uses the (BDRS-MS) update $X^{k+1} = \text{diag}(u^k)X^k \text{diag}(v^k)$. Combining (5.8) and (5.9) completes the induction for (5.5).

According to Theorem 3.2, $X^k = Y_{k-1}$ for $k \geq 1$. Thus

$$\hat{X}^k = \frac{1}{2(1 + \rho^k)} \begin{pmatrix} Y_{k-1} & \rho^k Y_{k-1} \\ \rho^k Y_{k-1} & Y_{k-1} \end{pmatrix}, \quad k \geq 1.$$

Since $\rho^k \rightarrow 0$ and $Y_{k+6} = Y_k$, we have

$$\lim_{t \rightarrow \infty} \hat{X}^{6t+\ell} = \frac{1}{2} \begin{pmatrix} Y_{\ell-1} & 0 \\ 0 & Y_{\ell-1} \end{pmatrix} = Z_\ell, \quad \text{for } \ell = 1, \dots, 6.$$

The six matrices $Z_1, \dots, Z_6$ are pairwise distinct because $Y_0, \dots, Y_5$ are pairwise distinct. Thus the full sequence cannot converge. Conversely, any convergent subsequence has a further subsequence whose iteration indices have a fixed residue class modulo 6. Its limit must be the corresponding Z_ℓ . Therefore these six matrices are exactly the subsequential limits.

Moreover, (5.5) gives, for every $k \geq 0$,

$$\hat{X}^k \mathbf{1}_{2n} = \frac{1}{2} \begin{pmatrix} X^k \mathbf{1}_n \\ X^k \mathbf{1}_n \end{pmatrix}.$$

Consequently,

$$\|\hat{X}^k \mathbf{1}_{2n} - \hat{r}\|_\infty = \frac{1}{2} \|X^k \mathbf{1}_n - r\|_\infty.$$

For $k \geq 1$, $X^k = Y_{k-1}$, so (4.4) and Lemma 4.1 imply

$$\|\hat{X}^k \mathbf{1}_{2n} - \hat{r}\|_\infty = \frac{1}{2} \|Y_{k-1} \mathbf{1}_n - r\|_\infty > 0.103.$$

For $k = 0$, every entry of the concrete M is a positive integer and $r_i < 1$. Hence

$$\|\hat{X}^0 \mathbf{1}_{12} - \hat{r}\|_\infty = \frac{1}{2} \|M \mathbf{1}_6 - r\|_\infty > \frac{5}{2} > 0.103.$$

This separately verifies the initial case and proves (5.6) for every $k \geq 0$. For each $\ell = 1, \dots, 6$, direct evaluation at the limit gives

$$\|Z_\ell \mathbf{1}_{2n} - \hat{r}\|_\infty = \frac{1}{2} \|Y_{\ell-1} \mathbf{1}_n - r\|_\infty > 0.103.$$

Thus every subsequential limit is infeasible. □

6 Conclusions

We constructed strictly positive data for which (BDRS-MS) has a nonconstant six-cycle after its first update. The construction separates cyclic symmetry from the interpolation requirement: symmetry is enforced by the block structure, while the remaining condition is the positive linear system $Bw = p$. Exact rational arithmetic certifies a 36×6 counterexample, and the induction proves its behavior for every iteration. A block lifting transfers the failure to (BDRS-OT) with a fixed, nonadditive cost, for every fixed $\eta > 0$. The lifted sequence has six distinct boundary limit points, all infeasible. The examples disprove unconditional convergence of BDRS. There still exists a gap between the sufficient condition provided in [4] and the counterexamples given in this paper. It is an important question to study the minimal requirements for guaranteeing the global convergence of BDRS.

AI Usage Disclosure

The counterexamples were provided by ChatGPT-6. The paper was written by the author.

References

- [1] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming. *Mathematics of Operations Research*, 18(1):202–226, 1993.
- [2] Joel Franklin and Jens Lorenz. On the scaling of multidimensional matrices. *Linear Algebra and Its Applications*, 114/115:717–735, 1989.
- [3] M. Idel. A review of matrix scaling and Sinkhorn’s normal form for matrices and positive maps. *arXiv preprint arXiv:1609.06349*, 2016.
- [4] Shiqian Ma. Convergence of BDRS as a matrix scaling algorithm. *Optimization Online*, 2026.
- [5] Shiqian Ma, Lin Xiao, and Renbo Zhao. Bregman Douglas-Rachford splitting method. *arxiv:2509.08739*, 2025.
- [6] R. Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405, 1967.
- [7] R. Sinkhorn and P. Knopp. Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*, 21(2):343–348, 1967.