A Heuristic for Statistical Seriation (Supplementary Material)

Komal Dhull[1] **Jingyan Wang**[1] **Nihar B. Shah**[1] **Yuanzhi Li**[1] **R. Ravi**[2]

[1]School of Computer Science, Carnegie Mellon University
[2]Tepper School of Business, Carnegie Mellon University

In this supplementary material, we present the proofs of all theoretical results.

A PRELIMINARY RESULTS

In this section, we present preliminary results that are used in the proofs. For the regularizer R, it can be verified that we have the symmetry

$$R_{ii'jj'} = R_{ii'j'j} = R_{i'ijj'} = R_{i'ij'j}.$$

We say that an entry (i, j) does not contribute to the regularizer if $R_{ii'jj'} = 0$ for all $i' \in [n]$ and $j' \in [d]$. We say that a row/column does not contribute to the regularizer if none of the entries in the row/column contributes to the regularizer. We say that (i, i', j, j') is a "conflicting quadruple" if we have $(A_{ij} - A_{ij'})(A_{i'j} - A_{i'j'}) < 0$. By the definition (7) of the regularizer, an entry (i, j) does not contribute to the regularizer if and only if the quadruple (i, i', j, j') is not a conflicting quadruple for each $i' \in [n]$ and $j' \in [d]$.

A.1 DERIVATIVE OF THE OBJECTIVE

We compute the derivative of the regularizer term $R_{ii'jj'}$ as

$$\frac{\partial R_{ii'jj'}}{\partial A_{ij}} = \begin{cases} 0 & \text{if } (A_{ij} - A_{ij'})(A_{i'j} - A_{i'j'}) \geq 0 \\ 2(A_{ij} - A_{ij'})(A_{i'j} - A_{i'j'})^2 & \text{otherwise.} \end{cases} \tag{13}$$

Hence, we have

$$\operatorname{sign}\left(\frac{\partial R_{ii'jj'}}{\partial A_{ij}}\right) = \operatorname{sign}(A_{ij} - A_{ij'}), \quad \text{if } (A_{ij} - A_{ij'})(A_{i'j} - A_{i'j'}) < 0. \tag{14}$$

It can be verified that we have the symmetry

$$\frac{\partial R_{ii'jj'}}{\partial A_{ij}} = \frac{\partial R_{ii'j'j}}{\partial A_{ij}} = \frac{\partial R_{i'ijj'}}{\partial A_{ij}} = \frac{\partial R_{i'ij'j}}{\partial A_{ij}}. \tag{15}$$

Combining (15) with the expression (6) of R, we have

$$\frac{\partial R}{\partial A_{ij}} = 4 \sum_{i' \in [n], j' \in [d]} \frac{\partial R_{ii'jj'}}{\partial A_{ij}}. \tag{16}$$

The derivative of the objective L is computed as

$$\nabla L(A) = 2(A - Y)_\Omega + \lambda \nabla R(A). \tag{17}$$

We have the partial derviative

$$\frac{\partial L}{\partial A_{ij}} = 2(A_{ij} - Y_{ij}) \cdot \mathbb{1}\{(i,j) \in \Omega\} + \lambda \frac{\partial R}{\partial A_{ij}} \tag{18}$$

$$\stackrel{\text{(i)}}{=} 2(A_{ij} - Y_{ij}) \cdot \mathbb{1}\{(i,j) \in \Omega\} + 4\lambda \sum_{i' \in [n], j' \in [d]} \frac{\partial R_{ii'jj'}}{\partial A_{ij}}, \tag{19}$$

where (i) is true by plugging in (16).

A.2 ADDITIONAL PRELIMINARY RESULTS

For notational simplicity, we denote the projection step (10b) as $\mathcal{P}_{[0,1]}A := \min\{1, \max\{0, A\}\}$ for any $A \in \mathbb{R}^{d \times n}$. The following lemma states that the objective L does not increase after a projection step.

Lemma 5. *Consider any $Y \in [0,1]^{n \times d}$. Then for any $A \in \mathbb{R}^{n \times d}$, we have $L(\mathcal{P}_{[0,1]}A) \leq L(A)$.*

Proof of Lemma 5 We consider the two terms in the objective (8). For the first term $\|A - Y\|_\Omega^2$, it is straightforward to verify that

$$\|Y - \mathcal{P}_{[0,1]}(A)\|_\Omega \leq \|Y - A\|_\Omega, \qquad \forall Y \in [0,1]^{n \times d}. \tag{20}$$

For the second term, we consider $R_{ii'jj'}$ for each quadruple (i, i', j, j'). Note that for any scalar values $a, b \in \mathbb{R}$, the term $(\mathcal{P}_{[0,1]}(a) - \mathcal{P}_{[0,1]}(b))$ either has the same sign as $(a - b)$ or has a value of 0. Now we discuss the following two cases depending on the sign of each quadruple (i, i', j, j').

Case 1: $(A_{ij} - A_{ij'})(A_{i'j} - A_{i'j'}) \geq 0$.

In this case, we have $(\mathcal{P}_{[0,1]}A_{ij'} - \mathcal{P}_{[0,1]}A_{i'j})(\mathcal{P}_{[0,1]}A_{i'j} - \mathcal{P}_{[0,1]}A_{i'j'}) \geq 0$. Hence, by the definition of the function $R_{ii'jj'}$, we have

$$0 = R_{ii'jj'}(A) = R_{ii'jj'}(\mathcal{P}_{[0,1]}A) \tag{21}$$

Case 2: $(A_{ij} - A_{ij'})(A_{i'j} - A_{i'j'}) < 0$.

In this case, we have $(\mathcal{P}_{[0,1]}A_{ij'} - \mathcal{P}_{[0,1]}A_{i'j})(\mathcal{P}_{[0,1]}A_{i'j} - \mathcal{P}_{[0,1]}A_{i'j'}) \leq 0$. Moreover, due to the projection we have

$$\left|\mathcal{P}_{[0,1]}A_{ij} - \mathcal{P}_{[0,1]}A_{ij'}\right| \leq |A_{ij} - A_{ij'}|$$
$$\left|\mathcal{P}_{[0,1]}A_{i'j} - \mathcal{P}_{[0,1]}A_{i'j'}\right| \leq |A_{i'j} - A_{i'j'}|.$$

By the definition of the function $R_{ii'jj'}$, it can be verified that

$$R_{ii'jj'}(A) \geq R_{ii'jj'}(\mathcal{P}_{[0,1]}A). \tag{22}$$

Combining (21) and (22) from the two cases, we have

$$R(A) \geq R(\mathcal{P}_{[0,1]}A), \quad \forall (i, i', j, j'), \forall A \in [0,1]^{n \times d}. \tag{23}$$

Finally, combining the two terms (20) and (23) of the objective L, we have

$$L(A) \geq L(\mathcal{P}_{[0,1]}A),$$

completing the proof.

Now we analyze the local optima of the objective. Standard results suggest that any local optimum in the interior of the domain satisfies the first-order optimality condition, namely having a gradient of 0. The following lemma suggests that any local optimum on the boundary of the domain also satisfies the first-order optimality condition. We define $\nabla L(A)$ as the gradient on $\mathbb{R}$, without restricting to the domain $[0,1]^{n \times d}$.

Lemma 6. *For any local optimum A of the objective* (8) *defined on the domain $[0,1]^{d\times n}$, we have $\nabla L(A) = 0$.*

Proof of Lemma 6 If any local optimum A is in the interior, then standard first-order optimality condition [Beck, 2014, Theorem 2.6] yields $\nabla L(A) = 0$. It remains to consider the case where A is on the boundary of the domain.

Assume for contradiction that there exists a local optimum A on the boundary with $\nabla L(A) \neq 0$. Without loss of generality we assume $\frac{\partial L(A)}{\partial A_{11}} \neq 0$. By definition of the local optimum, there exists some $\delta > 0$, such that $L(A') \geq L(A)$ for all $A' \in [0,1]^{n\times d}$ with $\|A' - A\|_F < \delta$. On the other hand, let E_{11} denote the matrix whose $(1,1)$-entry is 1 and all other entries are 0. By definition of the partial derivative, there exists some $\delta' \in (0, \delta)$ such that $L(A + \delta' E_{11}) < L(A)$. Now consider the point $\mathcal{P}_{[0,1]}(A + \delta' E_{11})$. By Lemma 5, we have

$$L(\mathcal{P}_{[0,1]}(A + \delta' E_{11})) \leq L(A + \delta' E_{11}) < L(A). \tag{24}$$

Since $[0,1]^{n\times d}$ is a convex set and $A \in [0,1]^{n\times d}$, by Lemma 5 we have

$$\left\|\mathcal{P}_{[0,1]}(A + \delta' E_{11}) - A\right\|_F \leq \|A + \delta' E_{11} - A\|_F = \delta' < \delta. \tag{25}$$

Combining (24) and (25), the point $\mathcal{P}_{[0,1]}(A + \delta' E_{11})$ yields a contradiction to the local optimality of A.

B PROOF OF THEOREM 1

The proof consists of two steps. First, we show that our objective L has a Lipschitz gradient. Second, we incorporate the projected step straightforwardly into standard analysis of gradient descent for functions with Lipschitz gradient.

Step 1: Bound the magnitude of the gradient $\|\nabla L\|_F$ and the Lipschitz constant

As a general definition, consider any $d \geq 1$. A function $f : \mathbb{R}^d \to \mathbb{R}$ is said to have a Lipschitz gradient with constant K on domain $D \subseteq \mathbb{R}^d$ if

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq K\|x - y\|_2, \text{ for all } x, y \in D.$$

For projected gradient descent, the gradient step (10a) may give solutions outside the domain $[0,1]^{n\times d}$, so we bound the gradient on an enlarged domain, namely $[-1,2]^{n\times d}$. For any $A \in [-1,2]^{n\times d}$, its partial derivative is given by (19) as:

$$\frac{\partial L}{\partial A_{ij}} = 2(A_{ij} - Y_{ij}) \cdot \mathbb{1}\{(i,j) \in \Omega\} + 4\lambda \sum_{i'\in[n], j'\in[d]} \frac{\partial R_{ii'jj'}}{\partial A_{ij}}. \tag{26}$$

Consider the term $\frac{\partial R_{ii'jj'}}{\partial A_{ij}}$ in (26). For each $i' \in [n]$ and $j' \in [d]$, we have

$$\left|\frac{\partial R_{ii'jj'}}{\partial A_{ij}}\right| \leq 2|A_{ij} - A_{ij'}| \cdot (A_{i'j} - A_{i'j'})^2 \leq 54. \tag{27}$$

Combining (27) and (26), we have

$$\left|\frac{\partial L}{\partial A_{ij}}\right| \leq 6 + 216\lambda nd, \tag{28}$$

and hence

$$\|\nabla L(A)\|_F \leq \sqrt{nd}(6 + 216\lambda nd). \tag{29}$$

Now we bound the Lipschitz constant of the objective L. Let $A, B \in [-1,2]^{n\times d}$ be any two matrices. Using (29), we have:

$$\begin{aligned}\|\nabla L(A) - \nabla L(B)\|_F^2 &\leq 4(nd)(6 + 216\lambda nd)^2 \\ &\overset{\text{(i)}}{\leq} 4(2 + 72\lambda nd)^2\|A - B\|_F^2,\end{aligned}$$

where (i) holds because $A, B \in [-1,2]^{n\times d}$ Hence, L has a Lipschitz gradient with $K = K(n,d,\lambda) := 4 + 144\lambda nd$ on $[-1,2]^{n\times d}$.

Step 2: Incorporate the projection step into standard analysis of gradient descent

The following standard result states that a gradient descent step with a sufficiently small stepsize decreases the objective.

Lemma 7 (Sufficient Decrease Lemma; Lemma 4.23 and Lemma 4.24 of Beck [2014])**.** *Suppose $f : \mathbb{R}^d \to \mathbb{R}$ has Lipschitz gradient with constant K. Then for any $x \in \mathbb{R}^d$ and $\gamma > 0$, we have*

$$f(x) - f(x - \gamma \nabla f(x)) \geq \left(1 - \frac{K\gamma}{2}\right) \|\nabla f(x)\|_2^2. \tag{30}$$

Now denote $\{A_t^{\text{grad}}\}_{t\geq 0}$ as the sequence after the gradient step (10a) in each iteration, and denote $\{A_t\}_{t\geq 0}$ as the sequence after the projection step (10b) in each iteration. We set the stepsize γ such that $\gamma \in (0, \frac{1}{4K})$. Due to the projection we have $A_t \in [0,1]^{n\times d}$ for all $t \geq 0$. Then for the gradient step, using (28) it can be verified that

$$A_t^{\text{grad}} = A_{t-1} - \gamma \nabla L(A_{t-1}) \in [-1, 2]^{n\times d}.$$

By Lemma 7 we have

$$L(A_{t-1}) - L(A_t^{\text{grad}}) \geq \left(1 - \frac{K\gamma}{2}\right) \|\nabla L(A_{t-1})\|_2^2 \geq 0. \tag{31}$$

For the projection step, by Lemma 5 we have

$$L(A_t^{\text{grad}}) - L(A_t) \geq 0. \tag{32}$$

Combining (31) and (32), we have

$$L(A_{t-1}) - L(A_t) \geq \left(1 - \frac{K\gamma}{2}\right) \|\nabla L(A_{t-1})\|_F^2 \geq 0. \tag{33}$$

Hence, the sequence $\{L(A_t)\}_{t\geq 0}$ is non-increasing. Furthermore, it is straightforward to verify that L is bounded below by 0. Since the sequence $\{L(A_t)\}_{t\geq 0}$ is non-increasing and bounded below by 0, we have

$$\lim_{t\to\infty} L(A_{t-1}) - L(A_t) = 0. \tag{34}$$

Plugging (34) into (33), we have $\lim_{t\to\infty} \|\nabla A_t\|_F = 0$, completing the proof.

C PROOF OF THEOREM 2

Since $Y \in \mathcal{M}$, we have A^* is a global minimum if and only if

$$\begin{aligned} A^*_\Omega &= Y_\Omega \\ \text{and } A^* &\in \mathcal{M}. \end{aligned}$$

By Lemma 6 any local optima (on the boundary) is a stationary point, so we only consider stationary points for the proof. To show that any stationary point is the global optimum, we separately discuss the three cases: $d = 2$, $d = 3$ and $n = 2$. In each case, we show that any stationary point A satisfies $A \in \mathcal{M}$. Since we have $\nabla L(A) = 0$ for any $A \in \mathcal{M}$, setting the derivative (17) to 0 gives $A_\Omega = Y_\Omega$.

C.1 $d = 2$

Consider any stationary point A. With $d = 2$, the matrix A has two columns. Assume for contradiction that $A \notin \mathcal{M}$. Denote the sets

$$I_+ := \{i \in [n] : A_{i1} - A_{i2} > 0\} \tag{35a}$$
$$I_- := \{i \in [n] : A_{i1} - A_{i2} < 0\}. \tag{35b}$$

By the assumption that $A \notin \mathcal{M}$, we have $I_+ \neq \emptyset$ and $I_- \neq \emptyset$.

For each $i \in I_+$, we have

$$\operatorname{sign}\left(\frac{\partial R}{\partial A_{i1}}\right) = \operatorname{sign}\left(\sum_{i' \in I_-} \frac{\partial R_{i,i',1,2}}{\partial A_{i1}}\right) \overset{\text{(i)}}{=} \operatorname{sign}\left(A_{i1} - A_{i2}\right) \overset{\text{(ii)}}{=} 1 \tag{36a}$$

$$\operatorname{sign}\left(\frac{\partial R}{\partial A_{i2}}\right) = \operatorname{sign}\left(\lambda \sum_{i' \in I_-} \frac{\partial R_{ii',2,1}}{\partial A_{i2}}\right) \overset{\text{(i)}}{=} \operatorname{sign}\left(A_{i2} - A_{i1}\right) \overset{\text{(ii)}}{=} -1, \tag{36b}$$

where the steps (i) are true due to (14), and the steps (ii) are true due to the definition (35a) of I_+. Likewise for each $i \in I_-$, we have

$$\operatorname{sign}\left(\frac{\partial R}{\partial A_{i1}}\right) = \operatorname{sign}(A_{i1} - A_{i2}) = -1 \tag{36c}$$

$$\operatorname{sign}\left(\frac{\partial R}{\partial A_{i2}}\right) = \operatorname{sign}(A_{i2} - A_{i1}) = 1. \tag{36d}$$

Case 1: If any entry (i, j) in the rows $I_+ \cup I_-$ is not observed (i.e., not in Ω), then by the gradient expression (18) we have

$$\frac{\partial L}{\partial A_{ij}} = \frac{\partial R}{\partial A_{ij}} \neq 0,$$

where the inequality holds due to (36). Contradiction to the assumption that A is a stationary point with $\nabla L(A) = 0$.

Case 2: All the entries in the rows $I_+ \cup I_-$ are observed.

Now consider any $i \in I_+$. Setting the gradient expression (18) to 0, we have

$$A_{i1} - Y_{i1} + \frac{\partial R}{\partial A_{i1}} = 0$$

$$Y_{i1} = A_{i1} + \frac{\partial R}{\partial A_{i1}}. \tag{37a}$$

Likewise, we have

$$Y_{i2} = A_{i2} + \frac{\partial R}{\partial A_{i2}}. \tag{37b}$$

Subtracting (37b) from (37a), we have

$$Y_{i1} - Y_{i2} = A_{i1} - A_{i2} + \left(\frac{\partial R}{\partial A_{i1}} - \frac{\partial R}{\partial A_{i2}}\right) > 0, \tag{38a}$$

where the last inequality holds because $(A_{i1} - A_{i2}) > 0$ by the definition (35a) of I_+, and because $\frac{\partial R}{\partial A_{i1}} - \frac{\partial R}{\partial A_{i2}} > 0$ due to (36a) and (36b). Likewise, for each $i \in I_-$,

$$Y_{i1} - Y_{i2} = A_{i1} - A_{i2} + \left(\frac{\partial R}{\partial A_{i1}} - \frac{\partial R}{\partial A_{i2}}\right) < 0. \tag{38b}$$

Combining (38) contradicts the assumption that $Y \in \mathcal{M}$.

C.2 $n = 2$

With $n = 2$, the matrix A has two rows. We prove by induction on the number of columns d. For $d = 1$, we trivially have $A \in \mathcal{M}$. For $d = 2$, the proof in Section C.1 yields the claimed result. Now suppose the claim holds for all $2 \times d$ matrices. We now consider any $2 \times (d + 1)$ matrix.

Let $A \in \mathbb{R}^{2\times(d+1)}$ be a stationary point given the observations $Y \in \mathbb{R}^{2\times(d+1)}$. Without loss of generality, we re-index the columns such that $A_{11} \leq A_{12} \leq \ldots \leq A_{1,d+1}$. Now consider the maximum entry in the second row of A.

Case 1: The entry $A_{2,d+1}$ is the maximum in the second row of A.

In this case, column $(d+1)$ contains the maximum for both rows. That is, we have $A_{i,d+1} \geq A_{ij}$ for each $i \in \{1,2\}$ and each $j \in [d]$. It can be verified that this column $(d+1)$ of the matrix, namely the column $\begin{bmatrix} A_{1,d+1} \\ A_{2,d+1} \end{bmatrix}$ does not contribute to the regularizer R. Hence, the gradient of the submatrix $\{A_{ij}\}_{i\in\{1,2\},j\in[d]}$ remains the same if the last column is removed. That is, for each $i \in \{1,2\}$ and $j \in [d]$, we have

$$\frac{\partial L(\{A_{ij}\}_{i\in\{1,2\},j\in[d]})}{\partial A_{ij}} = \frac{\partial L(A)}{\partial A_{ij}}.$$

Applying the induction hypothesis on the submatrix $\{A_{ij}\}_{i\in\{1,2\},j\in[d]}$, we have $\{A_{ij}\}_{i\in\{1,2\},j\in[d]} \in \mathcal{M}$. Since the last column $\begin{bmatrix} A_{1,d+1} \\ A_{2,d+1} \end{bmatrix}$ has the maximum entries in both rows, we have $A \in \mathcal{M}$.

Case 2: The entry $A_{2,d+1}$ is not a maximum in the second row.

Assume that a maximum in the second row is A_{2j^*} for some $1 \leq j^* < d$. Then we have $A_{2j^*} > A_{2,d+1}$.

Now consider the entry A_{1j^*}. By assumption we have $A_{1j^*} \leq A_{1,d+1}$. If $A_{1j^*} = A_{1,d+1}$, then the two entries in column j^* are both the maximum in their respective rows. Applying a similar inductive argument as in Case 1 to the submatrix $\{A_{ij}\}_{i\in\{1,2\},j\in[d+1]\setminus\{j^*\}}$ yields $A \in \mathcal{M}$. It remains to consider the case of $A_{1j^*} < A_{1,d+1}$.

We first analyze row 2. Using (14) combined with the fact that A_{2j^*} is the maximum entry in row 2, we have $\frac{\partial R}{\partial A_{2j^*}} \geq 0$. Moreover, since $A_{1j^*} < A_{1,d+1}$ and $A_{2j^*} > A_{2,d+1}$, the quadruple $(1, 2, j, d+1)$ is a conflicting quadruple, and hence we have the strict inequality

$$\frac{\partial R}{\partial A_{2j^*}} > 0. \tag{39}$$

On the other hand, we have $\frac{\partial R}{\partial A_{2,d+1}} \leq 0$, because for any conflicting quadruple $(2, d+1, 1, j)$ for some $j \in [d]$ that contributes to the derivative $\frac{\partial R}{\partial A_{2j}}$, we have

$$\operatorname{sign}\left(\frac{\partial R_{2,1,d+1,j}}{\partial A_{2j}}\right) \overset{\text{(i)}}{=} \operatorname{sign}(A_{2,d+1} - A_{2,j}) \overset{\text{(ii)}}{=} -\operatorname{sign}(A_{1,d+1} - A_{1,j}) \overset{\text{(iii)}}{=} -1,$$

where step (i) is true due to (14); step (ii) is true because $(2, 1, d+1, j)$ is assumed to be a conflicting quadruple and hence $(A_{2,d+1} - A_{2,j})(A_{1,d+1} - A_{1,j}) < 0$; step (ii) is true because by assumption $A_{1,d+1}$ is the maximum entry in the first row. Furthermore, the quadruple $(1, 2, j^*, d+1)$ is a conflicting quadruple, so we have strict inequality

$$\frac{\partial R}{\partial A_{2,d+1}} < 0. \tag{40}$$

Now consider whether the entries A_{2,j^*} and $A_{2,d+1}$ are observed. If either A_{2,j^*} or $A_{2,d+1}$ is not observed, then combining the gradient expression (18) with the strict inequalities (39) and (40), we have $\frac{\partial L}{\partial A_{2,d+1}} \neq 0$ or $\frac{\partial L}{\partial A_{2,j^*}} \neq 0$, contradicting the assumption that A is a stationary point. Hence, both A_{2,j^*} and $A_{2,d+1}$ are observed. Setting the gradient expression (18) to 0 respectively for the two entries $(2, j^*)$ and $(2, d+1)$, we have

$$Y_{2j^*} - Y_{2,d+1} = (A_{2j^*} - A_{2,d+1}) + \frac{\partial R}{\partial A_{2j^*}} - \frac{\partial R}{\partial A_{2,d+1}} > 0, \tag{41a}$$

where the inequality holds because $(A_{2j^*} - A_{2,d+1}) > 0$ as A_{2j^*} is the maximum entry in the second row, and because of (39) and (40).

Now we analyze row 1. Using a similar argument as in row 2, we have $\frac{\partial R}{\partial A_{1,d+1}} > 0$ because $A_{1,d+1}$ is the maximum entry in row 1, and strict inequality holds due to the existence of the conflicting quadruple $(1, 2, j^*, d+1)$. Moreover, we have $\frac{\partial R}{\partial A_{1j^*}} < 0$, because A_{2j^*} is the maximum entry in row 2 and the same conflicting quadruple $(1, 2, j^*, d+1)$. Similar to the analysis of row 2, we derive that both entries $(1, j^*)$ and $(1, d+1)$ are observed. Therefore,

$$Y_{1j^*} - Y_{1,d+1} = (A_{1j^*} - A_{1,d+1}) + \frac{\partial R}{\partial A_{1j^*}} - \frac{\partial R}{\partial A_{1,d+1}} < 0. \tag{41b}$$

Combining (41) contradicts the assumption that $Y \in \mathcal{M}$. Therefore, the entry $A_{2,d+1}$ is the maximum in row 2. From Case 1 we have $A \in \mathcal{M}$ for any $2 \times (d+1)$ matrices, completing the inductive step.

C.3 $d = 3$

With $d = 3$, the matrix has 3 columns. We consider the maximum entry in each row of the matrix. If a row has multiple maxima, one is chosen arbitrarily unless otherwise specified.

Case 1: The maxima in all the n rows of the matrix lie in the same column.

Without loss of generality, assume that the column containing all the maxima is column 3. It can be verified that all entries in column 3 do not contribute to the regularizer R. Applying the proof of the $d = 2$ case in Appendix C.1 to the submatrix $\{A_{ij}\}_{i\in[n],j\in\{1,2\}}$ yields $\{A_{ij}\}_{i\in[n],j\in\{1,2\}} \in \mathcal{M}$. Since column 3 contains the maximum of each row, we have $A \in \mathcal{M}$.

Case 2: The maxima of the n rows lie in two different columns.

If the 3 entries within each row are identical, then we have $A \in \mathcal{M}$, so it remains to consider the case where there exists a row whose values are not all identical. Without loss of generality, we assume that the entries are not all identical in row 1. We re-index the columns such that the first row is non-decreasing. Hence, we have $A_{11} < A_{13}$. We also re-index the rows, so that rows whose maxima are in the same column are grouped together. Then the matrix A is in one of the two following forms:

$$\left[\begin{array}{ccc} \text{min} & * & \text{max} \\ \vdots & \vdots & \vdots \\ * & * & \text{max} \\ \hline * & \text{max} & * \\ \vdots & \vdots & \vdots \\ * & \text{max} & * \end{array}\right] \tag{42a}$$

or

$$\left[\begin{array}{ccc} \text{min} & * & \text{max} \\ \vdots & \vdots & \vdots \\ * & * & \text{max} \\ \hline \text{max} & * & * \\ \vdots & \vdots & \vdots \\ \text{max} & * & * \end{array}\right], \tag{42b}$$

where we use "min" and "max" to indicate that the matrix entry is respectively a minimum or a maximum of its row (allowing ties). We use $*$ to indicate a general matrix entry, and use the horizontal line to indicate that the matrix structure decomposes into two blocks of rows. We denote the upper block and the lower block of the matrix as A_{U} and A_{L}, respectively, so that the matrix is also written as $\left[\begin{array}{c} A_{\mathrm{U}} \\ \hline A_{\mathrm{L}} \end{array}\right]$. We denote the row indices of the upper block and the lower block as $I_{\mathrm{U}}, I_{\mathrm{L}} \subseteq [n]$, respectively. By the assumption of the case, we have $I_{\mathrm{U}}, I_{\mathrm{L}} \neq \emptyset$.

Case 2.1: We consider the matrix form (42a).

We assume that in the lower block A_{L}, the entries in column 2 are strictly greater than the entries in column 3 within each row. That is, we assume $A_{i2} > A_{i3}$ for each $i \in I_{\mathrm{L}}$. This assumption is without loss of generality, because otherwise we have $A_{i2} = A_{i3}$, so that one can move row i to the upper block of the matrix.

Case 2.1.1: There exists a strict min-entry in column 2 in some row of the upper block. That is, there exists $i^* \in I_{\mathrm{U}}$ such that $A_{i^*1} > A_{i^*2}$. Since column 3 contains the maximum for all rows in the upper block, we have the strict inequality $A_{i^*2} < A_{i^*3}$.

Using (14), it can be verified that

$$\frac{\partial R}{\partial A_{i^*2}} < 0 \tag{43a}$$

$$\frac{\partial R}{\partial A_{i^*3}} > 0, \tag{43b}$$

where strict inequalities hold because the quadruple $(i^*, i', 2, 3)$ is a conflicting quadruple for each $i' \in I_{\mathrm{L}}$. Setting the gradient (18) for the stationary point A and combining with (43), we have the entries $(i^*, 1)$ and $(i^*, 2)$ must both be observed. Subtracting the gradient expression (18) on the entries $(i^*, 1)$ and $(i^*, 2)$, we have

$$Y_{i^*,2} - Y_{i^*,3} = (A_{i^*,2} - A_{i^*,3}) + \left(\frac{\partial R}{\partial A_{i^*,2}} - \frac{\partial R}{\partial A_{i^*,3}} \right) < 0,$$

where the last inequality holds by combining the fact of $A_{i^*2} < A_{i^*3}$ with inequalities (43). Hence, we have

$$Y_{i^*2} < Y_{i^*3}. \tag{44}$$

Now consider the case where there exists a min-entry in column 3 in the lower block, and denote this row as $i_{\mathrm{L}} \in I_{\mathrm{L}}$. Since we assume $A_{i2} > A_{i3}$ for each $i \in I_{\mathrm{L}}$ for Case 2.1, we have $(i_{\mathrm{L}}, 3)$ is a strict min-entry. Note that $(i^*, i_{\mathrm{L}}, 2, 3)$ is a conflicting quadruple. Using an argument similar to the derivation of (44), we have

$$Y_{i_{\mathrm{L}},2} > Y_{i_{\mathrm{L}},3}. \tag{45}$$

Combining (44) and (45) contradicts the assumption that $Y \in \mathcal{M}$. Hence, there does not exist any min-entry in column 3 in the lower block. Hence, the min-entry must lie in column 1 in the lower block, and all such min-entries are strict. Now the matrix A can be written in the form

$$\left[\begin{array}{ccc} \text{min} & * & \text{max} \\ \vdots & \vdots & \vdots \\ * & * & \text{max} \\ \hline \text{min} & \text{max} & * \\ \vdots & \vdots & \vdots \\ \text{min} & \text{max} & * \end{array}\right].$$

Now consider any row $i_{\mathrm{L}} \in I_{\mathrm{L}}$. We have $\frac{\partial R}{\partial A_{i_{\mathrm{L}},2}} > 0$ because column 2 contains a max-entry, and strict inequality holds due to the conflicting quadruple $(i^*, i_{\mathrm{L}}, 2, 3)$. On the other hand, we have $\frac{\partial R}{\partial A_{i_{\mathrm{L}},3}} \leq 0$, because no quadruple within the lower block contributes to the regularizer, and in the upper block column 3 contains the max-entry. Moreover, we have the strict inequality $\frac{\partial R}{\partial A_{i_{\mathrm{L}},3}} < 0$ due to the conflicting quadruple $(i^*, i_{\mathrm{L}}, 2, 3)$ again. Setting the gradient expression (18) for the stationary point A, we have that both entries $(i_{\mathrm{L}}, 2)$ and $(i_{\mathrm{L}}, 3)$ are observed. Subtracting the two gradient expression, we have

$$Y_{i_{\mathrm{L}},2} > Y_{i_{\mathrm{L}},3}. \tag{45'}$$

Combining (44) and (45') yields a contradiction to the assumption of $Y \in \mathcal{M}$, completing the proof of Case 2.1.1.

Case 2.1.2: There does not exist a min-entry in column 2 in the upper block.

In this case, the matrix is in the form

$$\left[\begin{array}{ccc} \text{min} & * & \text{max} \\ \vdots & \vdots & \vdots \\ \text{min} & * & \text{max} \\ \hline * & \text{max} & * \\ \vdots & \vdots & \vdots \\ * & \text{max} & * \end{array}\right].$$

We consider column 2 in the upper block. If $A_{i2} = A_{i3}$ for all $i \in I_{\mathrm{U}}$, then column 2 of the entire matrix only contains max-entries, and we apply the proof of Case 1 to column 2. It remains to consider the case where there exists some $i \in I_{\mathrm{U}}$ such that $A_{i_{\mathrm{U}},2} < A_{i_{\mathrm{U}},3}$. We have $\frac{\partial R}{\partial A_{i_{\mathrm{U}},3}} > 0$, where strict inequality holds due to the conflicting quadruple $(I_{\mathrm{U}}, I_{\mathrm{L}}, 2, 3)$ for any $i_{\mathrm{L}} \in I_{\mathrm{L}}$. Moreover, we have $\frac{\partial R}{\partial A_{i_{\mathrm{U}},2}} \leq 0$, because no quadruple within the upper block contributes to the regularizer, and in the lower block column 2 contains the max-entries. We have the strict inequality $\frac{\partial R}{\partial A_{i_{\mathrm{U}},2}} < 0$ due to the conflicting quadruple $(I_{\mathrm{U}}, I_{\mathrm{L}}, 2, 3)$ for any $i_{\mathrm{L}} \in I_{\mathrm{L}}$. Using the gradient expression (18), both entries $(i_{\mathrm{U}}, 2)$ and $(i_{\mathrm{U}}, 3)$ are observed, and we have

$$Y_{i_{\mathrm{U}},2} < Y_{i_{\mathrm{U}},3}. \tag{46a}$$

Now consider column 3 in the lower block. If for any row $i_\mathrm{L} \in I_\mathrm{L}$, column 3 contains the min-entry. Then due to the quadruple $(i_\mathrm{U}, i_\mathrm{L}, 2, 3)$ we have

$$Y_{i_\mathrm{L},2} < Y_{i_\mathrm{L},3}. \tag{46b}$$

Combining (46) yields a contradiction to the assumption that $Y \in \mathcal{M}$. Hence, column 3 does not contain any min-entry in the lower block. That is, the matrix can be written in the form

$$\left[\begin{array}{ccc} \text{min} & * & \text{max} \\ \vdots & \vdots & \vdots \\ \text{min} & * & \text{max} \\ \hline \text{min} & \text{max} & * \\ \vdots & \vdots & \vdots \\ \text{min} & \text{max} & * \end{array}\right].$$

Note that column 1 of the entire matrix only contains min-entries. Applying Case 1 to the minima (instead of the maxima) completes the proof of Case 2.1.2.

Case 2.2: We consider the form (42b).

Without loss of generality, we assume strict inequality $A_{i_\mathrm{L},1} > A_{i_\mathrm{L},3}$ for all $i_\mathrm{L} \in I_\mathrm{L}$. Otherwise, we have $A_{i_\mathrm{L},1} = A_{i_\mathrm{L},3}$ and one can move row i_L to the upper block. Assume that column 3 in the lower block contains a min-entry for some row $i_\mathrm{L} \in I_\mathrm{L}$. Combining row i_L with row 1 gives a conflicting quadruple $(1, i_\mathrm{L}, 1, 3)$. Using an argument similar to Case 2.1, we have

$$\begin{aligned} &Y_{11} < Y_{13} \\ &Y_{i_\mathrm{L},1} > Y_{i_\mathrm{L},3}, \end{aligned}$$

contradicting to the assumption $Y \in \mathcal{M}$. Hence, column 3 in the lower block does not contain any min-entry. Therefore, the matrix can be written as

$$\left[\begin{array}{ccc} \text{min} & * & \text{max} \\ \vdots & \vdots & \vdots \\ * & * & \text{max} \\ \hline \text{max} & \text{min} & * \\ \vdots & \vdots & \vdots \\ \text{max} & \text{min} & * \end{array}\right].$$

For any i_L, the quadruple $(1, i_\mathrm{L}, 1, 3)$ is again a conflicting quadruple. We have

$$\begin{aligned} &Y_{11} < Y_{13} \\ &Y_{i_\mathrm{L},1} > Y_{i_\mathrm{L},3}, \end{aligned}$$

contradicting to the assumption $Y \in \mathcal{M}$, completing the proof of Case 2.2.

Case 3: The maxima of the n rows span all the 3 columns. That is, the matrix can be written in the form:

$$\left[\begin{array}{ccc} \text{min} & * & \text{max} \\ \vdots & \vdots & \vdots \\ * & * & \text{max} \\ \hline * & \text{max} & * \\ \vdots & \vdots & \vdots \\ * & \text{max} & * \\ \hline \text{max} & * & * \\ \vdots & \vdots & \vdots \\ \text{max} & * & * \end{array}\right].$$

Denote the three blocks in the matrix as $A_{\text{U}}, A_{\text{M}}$ and A_{L} respectively, so that the matrix is also written as $\begin{bmatrix} A_{\text{U}} \\ \hline A_{\text{M}} \\ \hline A_{\text{L}} \end{bmatrix}$. Denote the corresponding sets of row indices as $I_{\text{U}}, I_{\text{M}}$ and I_{L}, respectively. Without loss of generality, we assume

$$A_{i2} > A_{i3} \qquad \forall i \in I_{\text{M}}$$
$$A_{i1} > \{A_{i2}, A_{i3}\} \qquad \forall i \in I_{\text{L}}.$$

Otherwise, we may move the rows in the middle block to the upper block, and move the rows in the lower block to the upper or middle blocks.

Now consider the lower block. Assume that there exists some min-entry in column 3 of the lower block. That is, assume that there exists some $i_{\text{L}} \in I_{\text{L}}$, such that $A_{i_{\text{L}},3}$ is a min-entry. Then the quadruple $(1, i_{\text{L}}, 1, 3)$ is a conflicting quadruple. Hence, we have

$$Y_{11} < Y_{13}$$
$$Y_{i_{\text{L}},1} > Y_{i_{\text{L}},3},$$

contradicting with the assumption that $Y \in \mathcal{M}$. Hence, there does not exist any min-entry in column 3 of the lower block. Then the matrix can be written in the form:

$$\begin{bmatrix} \text{min} & * & \text{max} \\ \vdots & \vdots & \vdots \\ * & * & \text{max} \\ \hline * & \text{max} & * \\ \vdots & \vdots & \vdots \\ * & \text{max} & * \\ \hline \text{max} & \text{min} & * \\ \vdots & \vdots & \vdots \\ \text{max} & \text{min} & * \end{bmatrix}.$$

Now consider row 1. The quadruple $(1, i_{\text{L}}, 1, 3)$ is a conflicting quadruple for each row $i_{\text{L}} \in I_{\text{L}}$ in the lower block. Hence, we have

$$Y_{11} < Y_{13}. \tag{47}$$

Assume without loss of generality that there exists some $i_{\text{M}} \in I_{\text{M}}$, such that $A_{i_{\text{M}},1} < A_{i_{\text{M}},2}$. Otherwise, the first column in the middle block contains all max-entries, and the matrix reduces to Case 2.2. Now consider any row $i_{\text{L}} \in I_{\text{L}}$. The quadruple $(i_{\text{M}}, i_{\text{L}}, 1, 2)$ is a conflicting quadruple. Hence, we have

$$Y_{i_{\text{L}},1} > Y_{i_{\text{L}},2}. \tag{48}$$

Combining (47) and (48) along with the assumption that $Y \in \mathcal{M}$, we have

$$Y_{i2} \leq Y_{i1} \leq Y_{i3}, \qquad \forall i \in [n]. \tag{49}$$

Now consider row i_{M} again in the middle block. Assume $A_{i_{\text{M}},1}$ is the min-entry in row i_{M}. The quadruple $(i_{\text{M}}, i_{\text{L}}, 1, 2)$ is a conflicting quadruple for any $i_{\text{L}} \in I_{\text{L}}$. Hence, we have

$$Y_{i_{\text{M}},2} > Y_{i_{\text{M}},1},$$

contradicting (49). Hence, it must be the case that $A_{i_{\text{M}},3}$ is the min-entry. Then we have $A_{I_{\text{M}},2} > A_{I_{\text{M}},1} \geq A_{i_{\text{M}},3}$. Now again consider any row $i_{\text{L}} \in I_{\text{L}}$. Recall that we have established that $A_{i_{\text{L}},3}$ cannot be a min-entry, so we have $A_{i_{\text{L}},3} > A_{i_{\text{L}},2}$. Then the quadruple $(i_{\text{M}}, i_{\text{L}}, 2, 3)$ is a conflicting quadruple. Hence, we have

$$Y_{i_{\text{M}},2} > Y_{i_{\text{M}},3},$$

again contradicting (49), completing the proof of Case 3.

Finally, combining the three cases yields the claimed result.

D PROOF OF PROPOSITION 3

Without loss of generality, we consider any $j, j' \in [d]$ such that

$$A_{1,j} < A_{1,j'} \tag{50a}$$
$$A_{2,j} > A_{2,j'}, \tag{50b}$$

and prove that

$$Y_{1,j} < Y_{1,j'}, \quad \text{if } (1,j),(1,j') \in \Omega.$$

First, consider the quadruple $(1, 2, j, j')$. By (50), it is a conflicting quadruple. By (14), we have

$$\frac{\partial R_{1,2,j,j'}}{\partial A_{1,j}} < 0 \tag{51a}$$
$$\frac{\partial R_{1,2,j,j'}}{\partial A_{1,j'}} > 0. \tag{51b}$$

Now consider quadruples involving any other column $k \in [d] \setminus \{j, j'\}$. We consider all possible orderings of the entries in column k relative to the columns j and j' as follows (we bold the entries in column k for better readability).

Case 1:

$$\begin{matrix} \mathbf{A_{1,k}} \le A_{1,j} < A_{1,j'} \\ \mathbf{A_{2,k}} \le A_{2,j'} < A_{2,j} \end{matrix}, \quad \text{or} \quad \begin{matrix} A_{1,j} < A_{1,j'} \le \mathbf{A_{1,k}} \\ A_{2,j'} < A_{2,j} \le \mathbf{A_{2,k}} \end{matrix}$$

It can be verified that column k does not form conflicting quadruples with columns j or j'. Hence, column k does not contribute to the gradient of the regularizer with respect to A_{1j} or $A_{1j'}$:

$$\frac{\partial R_{1,2,j,k}}{\partial A_{1j}} = \frac{\partial R_{1,2,j',k}}{\partial A_{1j'}} = 0.$$

Case 2:

$$\begin{matrix} A_{1,j} < \mathbf{A_{1,k}} \le A_{1,j'} \\ \mathbf{A_{2,k}} \le A_{2,j'} < A_{2,j} \end{matrix}, \quad \text{or} \quad \begin{matrix} A_{1,j} < A_{1,j'} \le \mathbf{A_{1,k}} \\ A_{2,j'} \le \mathbf{A_{2,k}} < A_{2,j} \end{matrix}$$

It can be verified that column k contributes a negative gradient to the regularizer with respect to A_{1j}, and no gradient to the regularizer with respect to $A_{1j'}$:

$$\frac{\partial R_{12jk}}{\partial A_{1j}} < 0 = \frac{\partial R_{12j'k}}{\partial A_{1j'}}.$$

Case 3:

$$\begin{matrix} A_{1,j} \le \mathbf{A_{1,k}} < A_{1,j'} \\ A_{2,j'} < A_{2,j} \le \mathbf{A_{2,k}} \end{matrix}, \quad \text{or} \quad \begin{matrix} \mathbf{A_{1,k}} \le A_{1,j} < A_{1,j'} \\ A_{2,j'} < \mathbf{A_{2,k}} \le A_{2,j} \end{matrix}$$

It can be verified that column k contributes no gradient to the regularizer with respect to $A_{1,j}$, and a positive gradient to the regularizer with respect to $A_{1j'}$:

$$\frac{\partial R_{1j2k}}{\partial A_{1j}} = 0 < \frac{\partial R_{1j'2k}}{\partial A_{1j'}}.$$

Case 4:

$$\begin{matrix} A_{1,j} < \mathbf{A_{1,k}} < A_{1,j'} \\ A_{2,j'} < \mathbf{A_{2,k}} < A_{2,j}, \end{matrix}$$

It can be verified that column k contributes a negative gradient to the regularizer with respect to A_{1j}, and a positive gradient to the regularizer with respect to $A_{1j'}$:

$$\frac{\partial R_{1j2k}}{\partial A_{1j}} < 0 < \frac{\partial R_{1j'2k}}{\partial A_{1j'}}.$$

Case 5:

$$\begin{aligned} \mathbf{A_{1,k}} &< A_{1,j} < A_{1,j'} \\ A_{2,j'} &< A_{2,j} < \mathbf{A_{2,k}}, \end{aligned}$$

It can be verified that column k contributes positive gradients to the regularizer with respect to both A_{1j} and $A_{1j'}$. By (13), we have

$$\begin{aligned} \frac{\partial R_{1j2k}}{\partial A_{1j}} &= 2(A_{1j} - A_{1k})(A_{2j} - A_{2k})^2 \\ \frac{\partial R_{1j'2k}}{\partial A_{1j'}} &= 2(A_{1j'} - A_{1k})(A_{2j'} - A_{2k})^2, \end{aligned}$$

and hence

$$0 < \frac{\partial R_{1j2k}}{\partial A_{1j}} < \frac{\partial R_{1j'2k}}{\partial A_{1j'}}.$$

Finally, combining all the 5 cases, it can be verified that they cover all possible orderings of the entries in column k relative to columns j and j'. Moreover, we have

$$\frac{\partial R_{1j2k}}{\partial A_{1j}} < \frac{\partial R_{1j'2k}}{\partial A_{1j'}} \qquad \forall k \in [d] \setminus \{j, j'\}. \tag{52}$$

Plugging (51) and (52) to (16), we have

$$\begin{aligned} \frac{\partial R}{\partial A_{1j}} &= 4\lambda \left(\frac{\partial R_{1,2,j,j'}}{\partial A_{1j}} + \sum_{k \in [d] \setminus \{j,j'\}} \frac{\partial R_{1j2k}}{\partial A_{1j}} \right) \\ &< 4\lambda \left(\frac{\partial R_{1,2,j',j'}}{\partial A_{1j'}} + \sum_{k \in [d] \setminus \{j,j'\}} \frac{\partial R_{1j'2k}}{\partial A_{1j'}} \right) = \frac{\partial R}{\partial A_{1j'}} \end{aligned} \tag{53}$$

Since we assume $(1, j), (1, j') \in \Omega$, using the gradient expression (18), we have

$$Y_{1j'} - Y_{1j} = (A_{1j} - A_{1j'}) + \left(\frac{\partial R}{\partial A_{1j}} - \frac{\partial R}{\partial A_{1j'}} \right) < 0,$$

where the inequality holds due to (50a) and (53), completing the proof.

E PROOF OF THEOREM 4

To present the main ideas of the proof, we first prove the following lemma under a simplified setting of Theorem 4, where the partition includes two subsets, $[d] = S \cup \overline{S}$ under full observations $\Omega = [n] \times [d]$. Then we present how to generalize Lemma 8 to any partition and partial observations.

Lemma 8. *Consider any matrix $Y \in \mathbb{R}^{n \times d}$, and full observations $\Omega = [n] \times [d]$. Consider $n = 2$. Assume there exists a partition of columns $[d] = S \cup \overline{S}$, such that any column in $\overline{S}$ dominates any column in S. That is, for any $j \in S$ and $j' \in \overline{S}$, we have*

$$Y_{i,j} < Y_{i,j'} \qquad \forall i \in \{1, 2\}. \tag{54}$$

Then we have the same relation for any stationary point A. That is,

$$A_{i,j} < A_{i,j'} \qquad \forall i \in \{1, 2\}, \forall j \in S \text{ and } j' \in \overline{S}. \tag{55}$$

E.1 PROOF OF LEMMA 8

We decompose the proof into the following steps.

Step 1: Show that conflicting quadruples cannot lie across $(S, \overline{S})$

Assume for contradiction that there exists a conflicting quadruple across $(S, \overline{S})$. That is, assume that there exists $j \in S$ and $j' \in \overline{S}$ such that $(A_{1,j} - A_{1,j'})(A_{2j} - A_{2j'}) < 0$. Applying Proposition 3, we have $(Y_{1,j} - Y_{1,j'})(Y_{2j} - Y_{2j'}) < 0$, contradicting the dominance assumption (54). Hence, all conflicting quadruples must lie within S, or within $\overline{S}$. Formally, for any $j, j' \in [d]$ such that $(1, 2, j, j')$ is a conflicting quadruple, we have either $j, j' \in S$ or $j, j' \in \overline{S}$.

Step 2: Partition columns into blocks We partition the columns into blocks $[d] = B_1 \cup B_2 \cup \ldots \cup B_K$ for some $K \geq 2$, such that the following conditions are satisfied:

(a) For $k \in [K]$, the block B_k includes columns only from S, or only from $\overline{S}$. That is, for each $k \in [K]$ we have $B_k \subseteq S$ or $B_k \subseteq \overline{S}$.
(b) For each $k \in [K-1]$, the blocks B_k and B_{k+1} are in different sets of the partition $(S, \overline{S})$. That is, for each $k \in [K-1]$, we have either $B_k \subseteq S$ and $B_{k+1} \subseteq \overline{S}$, or $B_k \subseteq \overline{S}$ and $B_{k+1} \subseteq S$.
(c) For each $k \in [K-1]$, the columns in B_{k+1} dominates the columns in B_k. That is,

$$A_{ij} \leq A_{ij'} \qquad \forall i \in \{1, 2\}, \forall k \in [K], \forall j \in B_k, \text{ and } \forall j' \in B_{k+1}.$$

Due to Step 1, all conflicting quadruples lie within S or $\overline{S}$, so it can be verified that a partition of blocks with $K \geq 2$ satisfying (a)-(c) exists.

Step 3: Show that A **satisfies the claimed dominance relation** (55).

We define

$$k_{\text{H}} := \max\{k \in [K] : B_k \subseteq S\} \tag{56a}$$
$$k_{\text{L}} := \min\{k \in [K] : B_k \subseteq \overline{S}\}, \tag{56b}$$

where ties are broken arbitrary. That is, $B_{k_{\text{L}}}$ is the block that is ordered the lowest among all blocks consisting of columns in $\overline{S}$, and $B_{k_{\text{H}}}$ is the block that is ordered the highest among all blocks consisting of columns in S. Furthermore, we define

$$j_{\text{H}} := \operatorname*{argmax}_{j \in B_{k_{\text{H}}}} A_{1j} \tag{57a}$$
$$j_{\text{L}} := \operatorname*{argmin}_{j \in B_{k_{\text{L}}}} A_{1,j}, \tag{57b}$$

where ties are broken arbitrarily. That is, $(1, j_{\text{H}})$ is the the maximum entry of A in row 1 among columns $B_{k_{\text{H}}}$, and $(1, j_{\text{L}})$ is the minimum entry of A in row 1 among columns $B_{k_{\text{L}}}$.

Case 1: $A_{1,j_{\text{H}}} < A_{1,j_{\text{L}}}$

By condition (c) of the construction, we have $k_{\text{L}} > k_{\text{H}}$. Hence, for all $j \in S$ and $j' \in \overline{S}$, we have

$$A_{1j} \overset{\text{(i)}}{\leq} A_{1j_{\text{H}}} < A_{1j_{\text{L}}} \overset{\text{(ii)}}{\leq} A_{1j'},$$

where steps (i) and (ii) are true due to the definitions (56) and (57) along with the fact that $k_{\text{L}} > k_{\text{H}}$. This completes Case 1.

Case 2: $A_{1,j_{\text{H}}} \geq A_{1,j_{\text{L}}}$

If any conflicting quadruple includes the entry $A_{1,j_{\text{H}}}$, then from Step 1 we have that all such conflicting quadruples are within S. By the definition (56a) of k_{H} and the definition (57a) of j_{H}, the entry $A_{1,j_{\text{H}}}$ is the maximum entry among all entries in row 1 among column S. Hence, we have

$$\frac{\partial R}{\partial A_{1,j_{\text{H}}}} \geq 0 \tag{58a}$$

and likewise

$$\frac{\partial R}{\partial A_{1,j_{\text{L}}}} \leq 0. \tag{58b}$$

Using the gradient expression (18), we have

$$Y_{1,j_{\mathrm{L}}} - Y_{1,j_{\mathrm{H}}} = (A_{1,j_{\mathrm{L}}} - A_{1,j_{\mathrm{H}}}) + \lambda \left(\frac{\partial R}{\partial A_{1,j_{\mathrm{L}}}} - \frac{\partial R}{\partial A_{1,j_{\mathrm{H}}}} \right) \leq 0,$$

where the last inequality is true due to (58) along with the assumption of the case. This contradicts the assumption (54) that the columns $\overline{S}$ dominates the columns S, completing Case 2.

Combining the two cases completes the proof.

E.2 PROOF OF THEOREM 4

Now we extend Lemma 8 to partial observations, stated as follows.

Lemma 9. *Consider any matrix* $Y \in [0,1]^{n \times d}$*, and partial observations* $\Omega \subseteq [n] \times [d]$*. Consider* $n = 2$*. Assume there exists a partition of columns* $[d] = S \cup \overline{S}$*, such that any column in* $\overline{S}$ *dominates any column in* S*. That is, we have*

$$Y_{i,j} < Y_{i,j'} \qquad \forall i \in \{1,2\}, \forall j \in S \text{ and } \forall j' \in \overline{S}. \tag{59}$$

Moreover, we assume that for each $j \in S, j' \in \overline{S}$*, we have*

$$\exists i \in \{1,2\} \text{ such that } (i,j),(i,j') \in \Omega. \tag{60}$$

Then for any stationary point A*, we have*

$$A_{i,j} < A_{i,j'} \qquad \forall i \in \{1,2\}, \forall j \in S \text{ and } j' \in \overline{S}. \tag{61}$$

We first use Lemma 9 to prove Theorem 4, and then prove Lemma 9. To prove Theorem 4, applying Lemma 9 with

$$S = \cup_{r=1}^{k} S_r$$
$$\overline{S} = \cup_{r=k+1}^{m} S_r$$

with every $k \in [m-1]$ gives

$$A_{ij} < A_{ij'} \qquad \forall i \in \{1,2\}, \forall j \in S_k, \text{ and } \forall j' \in S_{k+1},$$

completes the proof of Theorem 4. It now remains to prove Lemma 9.

Proof of Lemma 9 We extend the three steps in the proof of Lemma 8 to partial observations as follows.

Step 1: Show that conflicting quadruples cannot lie across $(S, \overline{S})$

Assume for contradiction that $(1,2,j,j')$ is a conflicting quadruple with $j \in S$ and $j' \in \overline{S}$. Assume without loss of generality that

$$A_{1j} < A_{1j'} \tag{62a}$$
$$A_{2j} > A_{2j'}. \tag{62b}$$

If all the 4 entries in this quadruple are observed, then applying Proposition 3 yields a contradiction. By assumption (60), one pair in the quadruple is observed. If the pair $A_{2j} > A_{2j'}$ is observed, then applying Proposition 3 gives $Y_{2j} > Y_{2j'}$, yielding a contradiction to (59). Hence, it remains to consider the case that the pair $A_{1j} < A_{1j'}$ is observed.

We first show that one entry in the pair $A_{2j} > A_{2j'}$ must be observed. Using the same argument as in the proof of Proposition 3, we have

$$\frac{\partial R}{\partial A_{2j}} > \frac{\partial R}{\partial A_{2j'}}. \tag{63}$$

If both entries in this pair are unobserved, then combining (63) with the gradient expression (18), we have

$$\frac{\partial L}{\partial A_{2j}} = \lambda \frac{\partial R}{\partial A_{2j}} > \lambda \frac{\partial R}{\partial A_{2j'}} = \frac{\partial L}{\partial A_{2j'}},$$

contradicting the assumption that A is a stationary point with a gradient of 0, and hence $\frac{\partial L}{\partial A_{2j}} = \frac{\partial L}{\partial A_{2j'}} = 0$. Hence, one entry in the pair $A_{2j} > A_{2j'}$ is observed. We now separately discuss the two cases depending on which entry in this pair is observed.

Case 1: A_{2j} is observed and $A_{2j'}$ is unobserved.

Since $A_{2j'}$ is unobserved, we have

$$\frac{\partial L}{\partial A_{2j'}} = \frac{\partial R}{\partial A_{2j'}} = 0. \tag{64}$$

Since $(1, 2, j, j')$ is a conflicting quadruple, we have

$$\frac{\partial R_{1,2,jj'}}{\partial A_{2j'}} < 0. \tag{65}$$

Combining (64) and (65), there must exist some $k \in [d]$ such that $\frac{\partial R_{1,2,j'k}}{\partial A_{2j'}} > 0$. That is, we have

$$A_{1j'} < A_{1k} \tag{66a}$$
$$A_{2j'} > A_{2k}. \tag{66b}$$

If $k \in S$, then $(1, 2, j', k)$ is a quadruple across the partition $(S, \overline{S})$. Recall by the assumption of the case that $A_{2j'}$ is unobserved, by condition (60), the pair $A_{1j'} < A_{1k}$ must be observed. Applying Proposition 3 yields $Y_{1j'} < Y_{1k}$, contradicting the dominance assumption (59).

It now remains to consider $k \in \overline{S}$. Recall by the assumption of the case that $A_{2j'}$ is unobserved. If A_{2k} is also unobserved, then the applying the arguments in Proposition 3 to the conflicting quadruple $(1, 2, j, k)$, we have

$$\frac{\partial R}{\partial A_{2j'}} < \frac{\partial R}{\partial A_{2k}},$$

and hence

$$\frac{\partial L}{\partial A_{2j'}} = \frac{\partial R}{\partial A_{2j'}} < \frac{\partial R}{\partial A_{2k}} = \frac{\partial L}{\partial A_{2k}},$$

contradicting the assumption that A is a stationary point with a gradient of 0. Hence, A_{2k} is observed. Combining (62) and (66), we have

$$A_{1j} < A_{1k}$$
$$A_{2j} > A_{2k}.$$

That is, $(1, 2, j, k)$ is a conflicting quadruple. Note that all the 4 entries in this conflicting quadruple are observed. Note that by the assumption that $j \in S$ and $k \in \overline{S}$, this conflicting quadruple is across the partition $(S, \overline{S})$. Applying Proposition 3 yields a contradiction with the dominance relation (59) of the partition $(S, \overline{S})$.

Case 2: A_{2j} is unobserved and $A_{2j'}$ is observed. A similar argument as in Case 1 applies.

Combining the two cases completes Step 1.

Step 2: Parition columns into blocks

We use the same construction of the blocks described in Step 2 of the proof of Lemma 8, and obtain the blocks $[d] = B_1 \cup \ldots \cup B_K$.

Step 3: Show that A satisfies the claimed dominance relation (61)

We follow Step 3 of the proof of Lemma 8, and use the same definition of $k_{\mathrm{H}}, k_{\mathrm{L}}$ from (56), and the definition of $(j_{\mathrm{H}}, j_{\mathrm{L}})$ from (57). Again assume for contradiction that the dominance relation (61) does not hold on A. We separately discuss the following cases depending on whether the entries $(1, j_{\mathrm{H}})$ and $(1, j_{\mathrm{L}})$ are observed.

Case 1: Both $(1, j_\mathrm{H})$ and $(1, j_\mathrm{L})$ are observed. Then Step 3 of Lemma 8 can be applied directly.

Case 2: Both $(1, j_\mathrm{H})$ and $(1, j_\mathrm{L})$ are unobserved. Due to the definitions (56a) and (57a), the entry $(1, j_\mathrm{H})$ is the maximum entry of A in row 1 among columns S. If the entry $(1, j_\mathrm{H})$ is involved in any conflicting quadruple, then due to Step 1, all such conflicting quadruples must lie within S. Hence, all conflicting quadruples contribute a positive gradient to $\frac{\partial R}{\partial A_{1,j_\mathrm{H}}}$. Since $(1, j_\mathrm{H})$ is unobserved, setting the gradient expression (18) to 0 for the stationary point A, we have

$$\frac{\partial L}{\partial A_{1,j_\mathrm{H}}} = \frac{\partial R}{\partial A_{1,j_\mathrm{H}}} = 0.$$

Hence, the entry $(1, j_\mathrm{H})$ cannot be in any conflicting quadruples. Therefore, $(2, j_\mathrm{H})$ is the maximum entry in row 2 among columns S. Likewise $(1, j_\mathrm{L})$ cannot be in any conflicting quadruples, and $(2, j_\mathrm{L})$ is the minimum entry in row 2 among columns $\overline{S}$. By the assumption (60), both $(2, j_\mathrm{L})$ and $(2, j_\mathrm{L})$ are observed. Applying the arguments in Case 1 to the pair of $(2, j_\mathrm{H})$ and $(2, j_\mathrm{L})$ completes Case 2.

Case 3: $(1, j_\mathrm{L})$ is observed and $(1, j_\mathrm{H})$ is unobserved.

Denote $(2, j'_\mathrm{L})$ as the minimum entry in row 2 among columns $\overline{S}$. If $(2, j'_\mathrm{L})$ is unobserved, then as in Case 2, the entry $(2, j'_\mathrm{L})$ cannot be in any conflicting quadruples, and hence $j'_\mathrm{L} = j_\mathrm{L}$. We have $(1, j_\mathrm{H})$ and $(2, j_\mathrm{L})$ both unobserved, contradicting (60). Hence, $(2, j'_\mathrm{L})$ must be observed, and likewise $(2, j'_\mathrm{H})$ must be observed, where $(2, j'_\mathrm{H})$ is the maximum entry in row 2 among columns S. Applying Case 1 to the pair of $(2, j'_\mathrm{L})$ and $(2, j'_\mathrm{H})$ completes the proof.

Case 4: $(1, j_\mathrm{L})$ is unobserved and $(1, j_\mathrm{H})$ is observed. By symmetry, a similar argument as in Case 3 applies.

Finally, combining the 4 cases completes the proof.

References

Amir Beck. *Introduction to Nonlinear Optimization.* Society for Industrial and Applied Mathematics, 2014.