Estimation of Monge Matrix

Jiaxin Hu

June 22, 2021

1 Background

Definition 1 (Supermodular function)**.** A function $f : \mathbb{R}^k \mapsto \mathbb{R}$ is *supermodular* if and only if

$$f(x) + f(y) \leq f(x \uparrow y) + f(x \downarrow y),$$

for all $x, y \in \mathbb{R}^k$, where $x \uparrow y$ denotes the component-wise maximum and $x \downarrow y$ denotes the component-wise minimum.

If $-f$ is supermodular, then f is a *submodular* function.

Definition 2 (Monge Matrix)**.** A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a *Monge matrix* or *submodular matrx* if for all i, j, k, l such that $1 \leq i < k \leq m$ and $1 \leq j < l \leq n$, we have

$$\boldsymbol{A}[i, j] + \boldsymbol{A}[k, l] \leq \boldsymbol{A}[i, l] + \boldsymbol{A}[j, k],$$

i.e., $\boldsymbol{A}$ is a Monge matrix if and only if $\boldsymbol{A}[i, j]$ is a submodular function of discrete variables i, j. In addition, a matrix $\boldsymbol{A}$ is called an *anti-Monge matrix* or *supermodular matrix* if $-\boldsymbol{A}$ is a Monge matrix.

Definition 3 (Pre-Monge Matrix)**.** A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is called pre-Monge matrix if there exist two permutations $\pi_1 : [m] \mapsto [m]$ and $\pi_2 : [n] \mapsto [n]$ such that

$$\boldsymbol{A}(\pi_1(i), \pi_2(j)), \quad (i, j) \in [m] \times [n],$$

is a Monge matrix. We denote such matrix as $\boldsymbol{A}(\pi_1, \pi_2) = [\![\boldsymbol{A}(\pi_1(i), \pi_2(j))]\!]$.

Proposition 1 (Equivalent definition)**.** *A matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a Monge matrix if and only if for all $i \in [m], j \in [n]$ we have*

$$\boldsymbol{A}[i, j] + \boldsymbol{A}[i + 1, j + 1] \leq \boldsymbol{A}[i, j + 1] + \boldsymbol{A}[i + 1, j]. \tag{1}$$

Proof. ($\Rightarrow$) Simply follows by the definition.

($\Leftarrow$) Suppose for all $i \in [m], j \in [n]$ the inequality (1) holds. Then we have

$$\boldsymbol{A}[i, j] - \boldsymbol{A}[i, j + 1] \leq \boldsymbol{A}[i + 1, j] - \boldsymbol{A}[i + 1, j + 1], \tag{2}$$

for all $i \in [m], j \in [n]$, which still holds after replacing j by $j+1, j+1, ..., l-1$ with any $l > j$. Summing up the inequality (2) with $j, j+1, ..., l-1$, we obtain

$$\boldsymbol{A}[i,j] - \boldsymbol{A}[i,l] \leq \boldsymbol{A}[i+1,j] - \boldsymbol{A}[i+1,l],$$

which is equivalent to

$$\boldsymbol{A}[i+1,l] - \boldsymbol{A}[i,l] \leq \boldsymbol{A}[i+1,j] - \boldsymbol{A}[i,j]. \tag{3}$$

Note that the above inequality holds after replacing i by $i+1, i+2, ..., k-1$ with any $k > i$. Summing up the inequality (3) with $i+1, i+2, ..., k-1$ with any $k > i$, we obtain the desired inequality. □

Proposition 2. *(Linear combination) Consider a sequence of Monge matrices $\boldsymbol{A}_i$ for $i = 1, ..., n$ and a positive sequence $\{c_i\}_{i=1}^n$. Then the linear combination $\sum c_i \boldsymbol{A}_i$ is still a Monge matrix.*

Proof. Simply follows by the definition. □

Proposition 3. *Any subarray produced by selecting certain rows and columns of Monge matrix is still a Monge matrix.*

Proof. Simply follows by the definition. □

Proposition 4. *The minima of each row of a Monge matrix occur in a non-decreasing sequence of columns.*

Proof. Suppose the minima for row i occurs in j^*-th column. We only need to show that the minima for row $i+1$ occurs in j' where $j' \geq j^*$. We show it by contradiction.

Suppose $j' < j^*$. Then by the definition of Monge matrix, we have

$$\boldsymbol{A}[i,j^*] + \boldsymbol{A}[i+1,j'] \geq \boldsymbol{A}[i,j'] + \boldsymbol{A}[i+1,j^*],$$

which implies

$$0 > \boldsymbol{A}[i,j^*] - \boldsymbol{A}[i,j'] \geq \boldsymbol{A}[i+1,j^*] - \boldsymbol{A}[i+1,j'].$$

Then, $\boldsymbol{A}[i+1,j^*]$ is smaller than $\boldsymbol{A}[i+1,j']$ which contradicts to the assumption that j' is the row minima. □

Proposition 5. *Define the difference operator $D_1 \in \mathbb{R}^{(m-1)\times m}$ as following.*

$$D = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

Define the difference operator $D_2 \in \mathbb{R}^{(n-1)\times n}$ analogously. For a Monge matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, we have

$$D_1 \boldsymbol{A} D_2^T \leq 0,$$

where $\leq$ denotes the entrywise inequality.

Proof. By the definition, we have

$$D_1 \boldsymbol{A} D_2^T[i,j] = \boldsymbol{A}[i+1,j+1] - \boldsymbol{A}[i+1,j] - \boldsymbol{A}[i,j+1] + \boldsymbol{A}[i,j] \leq 0,$$

where the inequality follows by the definition of Monge matrix. □

2 Estimation of Monge Matrix

The paper (Hütter et al., 2020) tackles to two problems in statistics

1. recovering the Monge matrix from a noisy data matrix;
2. recovering the pre-Monge matrix from a noisy data matrix.

The second question is harder since the parameter space for pre-Monge matrix is larger than Monge matrix. On the other hand, the second question is more realistic, and the estimation of permutation brings extra insights to seriation of the features. In practice, one may implement two-stage procedures to estimate the permutation and magnitude of the signal separately. The purposed two-stage algorithms lead to suboptimal convergence rates, which implies a potential research direction.

Note that it is equivalent to study the learning problem to anti-Monge matrix. For simplicity, the main results are presented with anti-Monge matrix. In the following context, let $\mathcal{M}, \bar{\mathcal{M}}$ denote the space for anti-Monge and pre-anti-Monge matrix, respectively, where

$$\begin{aligned}
&\mathcal{M} = \{\Theta \in \mathbb{R}^{m\times n} : D_1 \Theta D_2^T \geq 0\},\\
&\mathcal{M}(\pi_1,\pi_2) = \{\Theta(\pi_1,\pi_2) \in \mathbb{R}^{m\times n} : \Theta(\pi_1,\pi_2) \in \mathcal{M}\},\\
&\bar{\mathcal{M}} = \bigcup_{\pi_1\in\mathcal{S}_1,\pi_2\in\mathcal{S}_2} \mathcal{M}(\pi_1,\pi_2),
\end{aligned}$$

where $\mathcal{S}_1, \mathcal{S}_2$ are the collections of all the possible permutations. Define the quantity for $\Theta \in \mathcal{M}$

$$V(\Theta) = \Theta[1,1] + \Theta[m,n] - \Theta[m,1] - \Theta[1,n] = \left\| D_1 \Theta D_2^T \right\|_1,$$

and define $\mathcal{M}_{V_0}$ as

$$\mathcal{M}_{V_0} = \{\Theta \in \mathcal{M} : V(\Theta) \leq V_0\}.$$

In addition, we assume $m \geq n$.

2.1 Anti-Monge matrix Estimation

Consider the problem

$$\boldsymbol{Y} = \Theta^* + \epsilon,$$

where $\Theta^* \in \mathcal{M}$ and $\epsilon \sim \mathrm{subG}_{m\times n}(\sigma^2)$, and the least square estimator

$$\hat{\Theta} \in \operatorname*{arg\,min}_{\Theta\in\mathcal{M}} \|\Theta - \boldsymbol{Y}\|_F^2.$$

Theorem 2.1 (Upper bound for anti-Monge matrix estimation)**.** *The least square estimator* $\hat{\Theta}$ *achieves the bound*

$$\frac{1}{mn}\left\|\hat{\Theta}-\Theta^*\right\|_F^2 \lesssim \left[\frac{\sigma^2}{n}+\left(\frac{\sigma^2 V(\Theta^*)}{mn}\right)^{2/3}(\log m)^{1/2}(\log n)^{2/3}\right]\wedge\sigma^2,$$

with probability at least $1-\exp(-m)$*. The same bound holds for expectation.*

Theorem 2.2 (Lower bound for anti-Monge matrix estimation)**.** *Suppose* $\epsilon\mathcal{N}_{m\times n}(0,\sigma^2)$ *and* $\Theta^*\in\mathcal{M}_{V_0}$*, then we have*

$$\inf_{\tilde{\Theta}}\sup_{\Theta^*\in\mathcal{M}_{V_0}}\mathbb{E}\left[\frac{1}{mn}\left\|\tilde{\Theta}-\Theta^*\right\|_F^2\right]\gtrsim\left[\frac{\sigma^2}{n}+\left(\frac{\sigma^2 V(\Theta^*)}{mn}\right)^{2/3}\right]\wedge\sigma^2.$$

Remark 1. The minimax lower bound implies the least square estimator is approximately optimal. To intuitively understand the upper and lower bound, I believe an important fact is that any anti-Monge matrix is the sum of a rank-two matrix and a bivariate isotonic matrix, which follows by the Lemma 7 in (Hütter et al., 2020). My conjecture is that the first term of order $\mathcal{O}(n^{-1})$ correspond to the low-rank matrix and the second term of $\mathcal{O}(n^{4/3})$ comes from the isotonic structure. If $V(\Theta)=0$, then the Monge matrix reduces to a rank two matrix and the rate $\mathcal{O}(n^{-1})$ is reasonable for low-rank matrix estimation.

2.2 Pre-anti-Monge matrix Estimation

Consider the problem

$$\boldsymbol{Y}=\Theta^*(\pi_1^*,\pi_2^*)+\epsilon,$$

where $\Theta^*\in\mathcal{M}$ and $\epsilon\sim\mathrm{subG}_{m\times n}(\sigma^2)$, the global least square estimator

$$\hat{\Theta}^g\in\underset{\Theta\in\bar{\mathcal{M}}_{V_0}}{\arg\min}\|\Theta-\boldsymbol{Y}\|_F^2.$$

Theorem 2.3 (Upper bound for pre-anti-Monge matrix estimation)**.** *Suppose* $\Theta^*\in\mathcal{M}_{V_0}$*. The global least square estimator* $\hat{\Theta}^g$ *achieves the rate*

$$\frac{1}{mn}\left\|\hat{\Theta}^g-\Theta^*(\pi_1^*,\pi_2^*)\right\|_F^2\lesssim\left[\frac{\sigma^2}{n}+\left(\frac{\sigma^2 V_0}{mn}\right)^{2/3}(\log m)^{1/2}(\log n)^{2/3}\right]\wedge\sigma^2,$$

with probability at least $1-m^{-m}$*. The same bound holds for expectation.*

Remark 2. This theorem implies the estimation of pre-anti-Monge matrix won't be affected too much by the extra permutation settings. Note that the error $\left\|\hat{\Theta}^g-\Theta^*(\pi_1^*,\pi_2^*)\right\|_F^2$ does not separate the permutation error and the estimation error since the permutation may not be identifiable if two rows or columns differ by a constant vector. Besides, since pre-anti-Monge matrix possesses a large parameter space, the lower bound for a smaller space for anti-Monge matrix also holds for pre-anti-Monge matrix.

2.3 Algorithms

Though the least square estimators possess good convergence property, they may not be computationally doable. Hence, author in (Hütter et al., 2020) purpose two efficient algorithms:

1. Variance sorting based algorithm: use the fact that the variance of the difference between the first row and the other rows,
$$\sum_{k=1}^{n}\left[\Theta^*_{ik}-\Theta^*_{1k}-\frac{1}{n}\sum_{l=1}^{n}(\Theta^*_{il}-\Theta^*_{1l})\right]^2,$$
is monotonically increasing in i. Similar fact holds for the column.
2. Singular value thresholding (SVT) based algorithm, which may produce estimator that is out of the space of pre-anti-Monge matrix.

Both estimators achieve the suboptimal rates.

3 Possible Extension to tensor

I have two possible extensions of Monge matrix estimation to the tensor case.

3.1 Joint (pre-)anti-Monge estimation

Similarly with joint covariance or precision matrix estimation, we may have several sample matrices and encouraging the joint structure among these matrices would benefit the estimation. For example, suppose we have $\boldsymbol{Y}_k \in \mathbb{R}^{n\times n}$ which is the brain network of individual k, for $k = 1, ..., K$. Consider the model

$$\boldsymbol{Y}_k = \Theta^* + \epsilon_k,$$

where $\Theta^* \in \mathcal{M}$ and $\epsilon_k \sim_{i.i.d.} \text{subG}(\sigma^2)$. The conjectured upper bound for least square estimator is

$$\frac{1}{n^2}\left\|\hat{\Theta}-\Theta^*\right\|_F^2 \lesssim \left[\frac{\sigma^2}{Kn}+\left(\frac{\sigma^2 V(\Theta^*)}{n^2K^{3/2}}\right)^{2/3}(\log m)^{1/2}(\log n)^{2/3}\right]\wedge\sigma^2.$$

Similarly idea and conjectured bound can be applied to $\Theta^* \in \bar{\mathcal{M}}$ with $\Theta^*(\pi_1^*, \pi_2^*) \in \mathcal{M}$.

More aggressively, we can assume different networks have different Monge matrix but share the same permutation, i.e., the order of the features. For example, assume $\boldsymbol{Y}_k \in \mathbb{R}^{n\times n}$ is some brain network in people k, for $k = 1, ..., K$. Consider the model,

$$\boldsymbol{Y}_k = \Theta^*_k(\pi_1^*, \pi_2^*) + \epsilon_k,$$

where $\Theta^*_k \in \mathcal{M}$ and $\epsilon_k \sim_{i.i.d.} \text{subG}(\sigma^2)$. Though the specific interpretation of permutation π_1^*, π_2^* may not be so clear at this point, it is natural to assume that such order for gene is invariant with different people while the signal of the network may vary in different people.

3.2 Monge Tensor estimation

In joint Monge matrix estimation, we do not assume the Monge property on individuals or the third mode of the tensor. So, adding the Monge property on each mode of the tensor may be a natural extension. I purpose a possible definition of the Monge tensor.

Definition 4 (Monge Tensor)**.** A tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times \cdots \times n_K}$ is a *Monge Tensor* if for all $\boldsymbol{i} = (i_1, \cdots, i_K)$ and $\boldsymbol{j} = (j_1, ..., j_K)$, we have

$$\mathcal{X}[\boldsymbol{i} \downarrow \boldsymbol{j}] + \mathcal{X}[\boldsymbol{i} \uparrow \boldsymbol{j}] \leq \mathcal{X}[\boldsymbol{i}] + \mathcal{X}[\boldsymbol{j}],$$

where $\boldsymbol{i} \downarrow \boldsymbol{j}$ denotes the component-wise minimum and $\boldsymbol{i} \uparrow \boldsymbol{j}$ denotes the component-wise maximum, i.e.,

$$\boldsymbol{i} \downarrow \boldsymbol{j} = (\min\{i_1, j_1\}, ..., \min\{i_K, j_K\}), \quad \boldsymbol{i} \uparrow \boldsymbol{j} = (\max\{i_1, j_1\}, ..., \max\{i_K, j_K\}).$$

Then, $\mathcal{X}[\boldsymbol{i}]$ is a submodular function of variables $i_1, ..., i_K$. Similarly, we may define the pre-Monge tensor $\mathcal{X}$ if there exist a sequence of permutations $\pi_1, ..., \pi_K$ such that

$$\mathcal{X}(\pi_1, ..., \pi_K)[\boldsymbol{i}] = \mathcal{X}[pi_1(i_1), ..., \pi_K(i_K)],$$

is a Monge tensor.

Trivially follows the matrix case, the Monge Tensor may be the sum of low-rank tensor and a multiway isotonic tensor. Such shape constraint is weaker than the purely rank constraint. However, the interpretation and application of Monge Tensor are not clear for me, neither its algebraic properties.

The correspondingly estimation problem is

$$\mathcal{Y} = \mathcal{X}^* + \epsilon,$$

where $\mathcal{X}^*$ is a Monge Tensor and $\epsilon \sim \text{subG}(\sigma^2)$.

References

Hütter, J.-C., Mao, C., Rigollet, P., Robeva, E., et al. (2020). Estimation of monge matrices. *Bernoulli*, 26(4):3051–3080.