Notes on Statistical and Machine Learning

Matrix Decompositions, Approximations, and Completion

Chapter: *32* **Prepared by:** *Chenxi Zhou*

This note is prepared based on *Chapter 7, Matrix Decompositions, Approximations, and Completion* in Hastie, Tibshirani, and Wainwright (2015).

I. Introduction

1. **Question of Main Interest:** Given a matrix $\mathbf{Z} \in \mathbb{R}^{m\times n}$, we find a matrix $\widehat{\mathbf{Z}}$ that approximates $\mathbf{Z}$ in a suitable sense. Examples include:
   (a) the approximation $\widehat{\mathbf{Z}}$ is much simpler (in certain sense) than $\mathbf{Z}$ so that we can gain a better understanding of $\mathbf{Z}$, and
   (b) we impute or fill in any missing entries in $\mathbf{Z}$, which is known as *matrix completion.*

2. **General Approach:** The general approach is to consider estimators based on an optimization problem of the form

$$\begin{aligned} \underset{\mathbf{M}\in\mathbb{R}^{m\times n}}{\text{minimize}} \ & \|\mathbf{Z} - \mathbf{M}\|_{\mathrm{F}}^2 \\ \text{subject to } \ & \Phi(\mathbf{M}) \le c, \end{aligned}$$

   where
   (a) $\|\cdot\|_{\mathrm{F}}^2$ is the squared Frobenius norm of a matrix, and
   (b) Φ is a constraint function that encourages the solution of the optimization problem to be sparse in some general sense.

3. **Summary of Various Methods:**

	Constraint	Method
(a)	$\|\mathbf{M}\|_1 \le c$	Sparse matrix approximation
(b)	$\text{rank}(\mathbf{M}) \le k$	Singular value decomposition
(c)	$\|\mathbf{M}\|_* \le c$	Convex matrix approximation
(d)	$\mathbf{M} = \mathbf{UDV}^\top, \Phi_1(\mathbf{u}_j) \le c_1, \Phi_2(\mathbf{v}_k) \le c_2$	Penalized singular value decompositon
(e)	$\mathbf{M} = \mathbf{LR}^\top, \Phi_1(\mathbf{L}) \le c_1, \Phi_2(\mathbf{R}) \le c_2$	Max-margin matrix factorization
(f)	$\mathbf{M} = \mathbf{L} + \mathbf{S}, \Phi_1(\mathbf{L}) \le c_1, \Phi_2(\mathbf{S}) \le c_2$	Additive matrix decomposition

(a) The constraint is $\|\mathbf{M}\|_1 \le c$, that is, we put an L^1-norm constraint on all entries in $\mathbf{M}$. This constraint leads to a soft-threshold version of the original matrix and the solution $\widehat{\mathbf{Z}}$ takes the form

$$\widehat{z}_{i,j} = \text{sign}(z_{i,j})(|z_{i,j}| - \gamma)_+, \qquad \text{for all } i = 1, 2, \cdots, m \text{ and } j = 1, 2, \cdots, n,$$

where the scalar $\gamma > 0$ is chosen so that $\sum_{i=1}^{m} \sum_{j=1}^{n} |\widehat{z}_{i,j}| = c$.

Remark. The resulting $\widehat{\mathbf{Z}}$ is useful in sparse covariance matrix estimation.

(b) The constraint is that the rank of $\mathbf{M}$ does *not* exceed a pre-specified value k, which is equivalent to the number of non-zero singular values in $\mathbf{M}$ not exceeding k.

The optimal solution can be found by computing the singular value decomposition (SVD) and truncating it to its top k components.

Remark. The formulation of the constraint $\text{rank}(\mathbf{M}) \le k$ leads to a *non-convex* optimization problem.

(c) The constraint $\|\mathbf{M}\|_* \le c$, where $\|\cdot\|_*$ is the *nuclear norm* and is equal to the sum of the singular values of a matrix, is a relaxation of $\text{rank}(\mathbf{M}) \le k$.

Remark. The nuclear norm is a convex matrix function, so the associated problem is convex and can be solved by computing the SVD, and soft-thresholding its singular values.

(d) The constraints $\Phi_1(\mathbf{u}_j) \le c_1$ and $\Phi_2(\mathbf{v}_k) \le c_2$ impose penalties on the left and right singular vectors. Examples include the usual L^1- or L^2-norms, with the former choice yielding sparsity in the elements of the singular vectors.

Remark. Sparse singular vectors are useful for problems where the interpretation of the singular vectors is important.

(e) The constraint

$$\mathbf{M} = \mathbf{L}\mathbf{R}^\top, \Phi_1(\mathbf{L}) \le c_1, \Phi_2(\mathbf{R}) \le c_2$$

imposes penalties directly on the components of the LR-matrix factorization.

(f) The constraint

$$\mathbf{M} = \mathbf{L} + \mathbf{S}, \Phi_1(\mathbf{L}) \le c_1, \Phi_2(\mathbf{S}) \le c_2$$

seeks an additive decomposition of the matrix, imposing penalties on both components in the sum.

II. Singular Value Decomposition

1. **Singular Value Decomposition:** Given a matrix $\mathbf{Z} \in \mathbb{R}^{m \times n}$ with $\text{rank}(\mathbf{Z}) = r \le \min\{m, n\}$, its singular value decomposition is given by

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^\top,$$

where

(a) $\mathbf{U} \in \mathbb{R}^{m\times r}$ is an orthogonal matrix satisfying $\mathbf{U}^\top\mathbf{U} = \mathbf{I}_r$ whose columns $\mathbf{u}_j \in \mathbb{R}^m$ are called the *left singular vectors*, for $j = 1, 2, \cdots, r$,

(b) $\mathbf{V} \in \mathbb{R}^{n\times r}$ is an orthogonal matrix satisfying $\mathbf{V}^\top\mathbf{V} = \mathbf{I}_r$ whose columns $\mathbf{v}_j \in \mathbb{R}^n$ are called the *right singular vectors*, for $j = 1, 2, \cdots, r$, and

(c) $\mathbf{D} \in \mathbb{R}^{r\times r}$ is diagonal, with diagonal elements $d_1 \geq d_2 \geq \cdots \geq d_r \geq 0$ known as the *singular values*.

Remark 1. If the diagonal entries $d_1, d_2, \cdots, d_r$ are unique, then so are $\mathbf{U}$ and $\mathbf{V}$, up to column-wise sign flips.

Remark 2. By convention, singular values are always non-negative, which should be distinguished from the eigenvalues that could be negative.

2. **Rank Constrained Optimization Problem:** Let $\mathbf{Z} \in \mathbb{R}^{m\times n}$ be given and $r_0 \leq \operatorname{rank}(\mathbf{Z})$ be given as well. We assume that $m \leq n$ and $\operatorname{rank}(\mathbf{Z}) = m$.

Consider the following optimization problem

$$\underset{\operatorname{rank}(\mathbf{M})=r_0}{\operatorname{minimize}} \|\mathbf{Z} - \mathbf{M}\|_{\mathrm{F}}^2. \tag{1}$$

We show that the solution to (1) is

$$\widehat{\mathbf{Z}}_{r_0} := \underset{\operatorname{rank}(\mathbf{M})=r_0}{\arg\min}\ \|\mathbf{Z} - \mathbf{M}\|_{\mathrm{F}}^2 = \mathbf{U}\mathbf{D}_{r_0}\mathbf{V}^\top$$

where $\mathbf{D}_{r_0} \in \mathbb{R}^{n\times n}$ is the same as the matrix $\mathbf{D}$ except all but the first r_0 diagonal elements are set to 0.

We first note that any matrix $\mathbf{M}$ of rank r_0 can be factored as $\mathbf{M} = \mathbf{Q}\mathbf{A}$, where $\mathbf{Q} \in \mathbb{R}^{m\times r_0}$ is an orthogonal matrix satisfying $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_{r_0}$ and $\mathbf{A} \in \mathbb{R}^{r_0\times n}$. Then, given $\mathbf{Q}$, the optimal value for $\mathbf{A}$ is $\mathbf{Q}^\top\mathbf{Z}$. To see this, notice that

$$\begin{aligned}
\|\mathbf{Z} - \mathbf{M}\|_{\mathrm{F}}^2 &= \|\mathbf{Z} - \mathbf{Q}\mathbf{A}\|_{\mathrm{F}}^2 \\
&= \operatorname{trace}\big(\mathbf{Z}^\top\mathbf{Z} - \mathbf{Z}^\top\mathbf{Q}\mathbf{A} - \mathbf{A}^\top\mathbf{Q}^\top\mathbf{Z} + \mathbf{A}^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{A}\big) \\
&= \operatorname{trace}\big(\mathbf{Z}^\top\mathbf{Z} - \mathbf{Z}^\top\mathbf{Q}\mathbf{A} - \mathbf{A}^\top\mathbf{Q}^\top\mathbf{Z} + \mathbf{A}^\top\mathbf{A}\big) \\
&= \operatorname{trace}\big((\mathbf{A} - \mathbf{Q}^\top\mathbf{Z})^\top(\mathbf{A} - \mathbf{Q}^\top\mathbf{Z}) - (\mathbf{Q}^\top\mathbf{Z})^\top(\mathbf{Q}^\top\mathbf{Z}) + \mathbf{Z}^\top\mathbf{Z}\big) \\
&\geq \operatorname{trace}(-\mathbf{Z}^\top\mathbf{Q}\mathbf{Q}^\top\mathbf{Z} + \mathbf{Z}^\top\mathbf{Z}).
\end{aligned}$$

Hence, the optimal value for $\mathbf{A}$ is $\mathbf{Q}^\top\mathbf{Z}$. With the optimal $\mathbf{A} = \mathbf{Q}^\top\mathbf{Z}$, the objective function $\|\mathbf{Z} - \mathbf{M}\|_{\mathrm{F}}^2$ can be written as

$$\|\mathbf{Z} - \mathbf{M}\|_{\mathrm{F}}^2 = \|\mathbf{Z} - \mathbf{Q}\mathbf{Q}^\top\mathbf{Z}\|_{\mathrm{F}}^2 = \operatorname{trace}(-\mathbf{Z}^\top\mathbf{Q}\mathbf{Q}^\top\mathbf{Z} + \mathbf{Z}^\top\mathbf{Z}).$$

Since the term $\mathbf{Z}\mathbf{Z}^\top$ does *not* depend on $\mathbf{Q}$, we can ignore it and have

$$\begin{aligned}
\arg\min_{\mathbf{Q}} \|\mathbf{Z} - \mathbf{Q}\mathbf{Q}^\top\mathbf{Z}\|_{\mathrm{F}}^2 &= \arg\min_{\mathbf{Q}} \operatorname{trace}(-\mathbf{Z}^\top\mathbf{Q}\mathbf{Q}^\top\mathbf{Z}) \\
&= \arg\max_{\mathbf{Q}} \operatorname{trace}(\mathbf{Z}^\top\mathbf{Q}\mathbf{Q}^\top\mathbf{Z}) \\
&= \arg\max_{\mathbf{Q}} \operatorname{trace}(\mathbf{Q}^\top\mathbf{Z}\mathbf{Z}^\top\mathbf{Q}),
\end{aligned}$$

subject to the constraint $\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_{r_0}$.

By the singular value decomposition $\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$, we have

$$\mathbf{Z}\mathbf{Z}^\top = \mathbf{U}\mathbf{D}^2\mathbf{U}^\top.$$

Then,

$$\mathbf{Q}^\top\mathbf{Z}\mathbf{Z}^\top\mathbf{Q} = (\mathbf{U}^\top\mathbf{Q})^\top\mathbf{D}^2(\mathbf{U}^\top\mathbf{Q}) = \widetilde{\mathbf{Q}}^\top\mathbf{D}^2\widetilde{\mathbf{Q}},$$

where $\widetilde{\mathbf{Q}} = \mathbf{U}^\top\mathbf{Q}$. Since $\mathbf{U} \in \mathbb{R}^{m\times m}$ is an orthogonal matrix, we have $\mathbf{U}\mathbf{U}^\top = \mathbf{I}_m$, and hence,

$$\widetilde{\mathbf{Q}}^\top\widetilde{\mathbf{Q}} = \mathbf{Q}^\top\mathbf{U}\mathbf{U}^\top\mathbf{Q} = \mathbf{Q}^\top\mathbf{Q}.$$

Hence, the problem of interest can be transformed as

$$\begin{aligned} &\underset{\mathbf{Q}\in\mathbb{R}^{m\times r_0}}{\text{maximize}} \ \ \text{trace}(\mathbf{Q}^\top\mathbf{D}^2\mathbf{Q}) \\ &\text{subject to} \ \ \mathbf{Q}^\top\mathbf{Q} = \mathbf{I}_{r_0}. \end{aligned} \tag{2}$$

Now, if we let $\mathbf{H} = \mathbf{Q}\mathbf{Q}^\top \in \mathbb{R}^{m\times m}$, it is plain to see $\mathbf{H} = \mathbf{H}^\top$ and

$$\mathbf{H}\mathbf{H} = \mathbf{Q}\mathbf{Q}^\top\mathbf{Q}\mathbf{Q}^\top = \mathbf{Q}\mathbf{Q}^\top = \mathbf{H}.$$

Hence, if we let $h_{i,i}$ denote the i-th diagonal element of $\mathbf{H}$, for $i = 1, 2, \cdots, m$, we have

$$h_{i,i} = \sum_{j=1}^m h_{i,j}h_{j,i} = \sum_{j=1}^m h_{i,j}^2 = h_{i,i}^2 + \sum_{j\neq i} h_{i,j}^2 \geq h_{i,i}^2.$$

Hence, we have $h_{i,i} \in [0, 1]$, for all $i = 1, 2, \cdots, m$. In addition, note that

$$\sum_{i=1}^m h_{i,i} = \text{trace}(\mathbf{H}) = \text{trace}(\mathbf{Q}\mathbf{Q}^\top) = \text{trace}(\mathbf{Q}^\top\mathbf{Q}) = \text{trace}(\mathbf{I}_{r_0}) = r_0,$$

and, by simple algebra,

$$\text{trace}(\mathbf{Q}^\top\mathbf{D}^2\mathbf{Q}) = \sum_{i=1}^m h_{i,i}d_i^2$$

Therefore, the problem (2) is equivalent to the following one

$$\underset{h_{i,i}\in[0,1],\sum_{i=1}^m h_{i,i}=r_0}{\text{maximize}} \sum_{i=1}^m h_{i,i}d_i^2. \tag{3}$$

Finally, suppose $d_1^2 \geq d_2^2 \geq \cdots \geq d_m^2 \geq 0$ be the squared singular values of $\mathbf{Z}$. The solution to (3) is obtained by setting $h_{1,1} = h_{2,2} = \cdots = h_{r_0,r_0} = 1$ and all remaining to be 0. An optimal choice of $\mathbf{Q}$ that satisfies such conditions is $\mathbf{U}_{r_0}$, the matrix formed by the first r_0 columns of $\mathbf{U}$.

Remark. If the singular values of $\mathbf{Z}$ satisfy

$$d_1^2 > d_2^2 > \cdots > d_m^2 \geq 0,$$

the solution to (3) is unique.

III. Missing Data and Matrix Completion

1. **Matrix Completion:** Let $\mathbf{Z} \in \mathbb{R}^{m\times n}$ be a matrix containing missing values. *Matrix completion* refers to the problem of filling in or imputing missing values.

2. **Problem Constraint:** The matrix completion problem is ill-specified unless additional constraints on the unknown matrix $\mathbf{Z}$ are imposed. We will specify the constraints related to the rank.

3. **Notation:** We let
$$\Omega \subset \{1, 2, \cdots, m\} \times \{1, 2, \cdots, n\}$$
denote the indices of the observed entries of the matrix $\mathbf{Z} \in \mathbb{R}^{m\times n}$.

4. **Naive Problem Formulation:** Given $\mathbf{Z} \in \mathbb{R}^{m\times n}$, we seek the lowest rank approximation to $\mathbf{Z}$ that interpolates the observed entries of $\mathbf{Z}$. The corresponding optimizaiton problem is
$$\begin{aligned} &\underset{\mathbf{M}\in\mathbb{R}^{m\times n}}{\text{minimize}} \ \ \text{rank}(\mathbf{M}) \\ &\text{subject to} \ \ m_{i,j} = z_{i,j} \text{ for all } (i,j) \in \Omega, \end{aligned} \tag{4}$$
where $m_{i,j}$ and $z_{i,j}$ denote the (i,j)-th entry of $\mathbf{M}$ and $\mathbf{Z}$, respectively, for all $i = 1, 2, \cdots, m$ and $j = 1, 2, \cdots, n$.

   Issues of Optimization Problem (4):

   (a) The problem (4) is computationally intractable (NP-hard), and cannot be solved in general even for moderately large matrices.

   (b) The constraint $z_{i,j} = m_{i,j}$ for all $(i,j) \in \Omega$ can be too restrictive and may lead to overfitting.

5. **Matrix Completion by Low-Rank Approximation:**

   (a) *Motivation:* Due to the two issues mentioned above about (4), it is generally better to allow $\mathbf{M}$ to make some errors on the observed data.

   (b) *Problem Formulation:* We consider the following optimization problem
$$\begin{aligned} &\text{minimize} \ \ \text{rank}(\mathbf{M}) \\ &\text{subject to} \ \ \sum_{(i,j)\in\Omega} (z_{i,j} - m_{i,j})^2 \le \delta, \end{aligned} \tag{5}$$
   or equivalently,
$$\underset{\text{rank}(\mathbf{M})\le r}{\text{minimize}} \sum_{(i,j)\in\Omega} (z_{i,j} - m_{i,j})^2. \tag{6}$$

   Remark. The family of solutions generated by varying δ in (5) is the same as that generated by varying r in problem (6).

(c) *Non-convexity of the Problem:* The problems (5) and (6) are non-convex. Exact solutions are in general *not* available.

(d) *Heuristic Algorithm:* Heuristic algorithms can be used to find local minima of (5) and (6). One example is the following:

i. Start with an initial guess for the missing values, and use them to complete $\mathbf{Z}$;

ii. Compute the rank-r SVD approximation of the filled-in matrix as in (1), and use it to provide new estimates for the missing values;

iii. Repeat the preceding step till convergence.

The missing value imputation for a missing entry $z_{i,j}$ is simply the (i,j)-th entry of the final rank-r approximation $\widehat{\mathbf{Z}}$.

6. **Matrix Completion Using Nuclear Norm — Version 1:** A convex relaxation of (4) is the following

$$\begin{aligned} &\underset{\mathbf{M}\in\mathbb{R}^{m\times n}}{\text{minimize}} \ \|\mathbf{M}\|_* \\ &\text{subject to} \ \ m_{i,j} = z_{i,j} \text{ for all } (i,j)\in\Omega, \end{aligned} \tag{7}$$

where $\|\cdot\|_*$ denotes the nuclear norm of $\mathbf{M}$, i.e., the sum of singular values of $\mathbf{M}$.

Since the nuclear norm is a convex relaxation of the rank of a matrix, and hence the problem (7) is convex.

7. **Matrix Completion Using Nuclear Norm — Version 2:** Since it is unrealistic to model the observed entries as being noiseless, we instead consider the following optimization problem

$$\underset{\mathbf{M}\in\mathbb{R}^{m\times n}}{\text{minimize}} \left\{ \frac{1}{2}\sum_{(i,j)\in\Omega} (z_{i,j} - m_{i,j})^2 + \lambda\|\mathbf{M}\|_* \right\}, \tag{8}$$

where $\lambda > 0$ is the penalty parameter. The problem (8) is called the *spectral regularization.*

Remark 1. This modification from (7) to (8) allows for solutions that do *not* fit the observed entries exactly, reducing potential overfitting in the case of noisy entries.

Remark 2. The value of $\lambda > 0$ can be chosen using the cross-validation.

8. **Algorithm of Solving (8):**

(a) *Main Idea:* The main idea of the algorithm to solve (8) is the following:

i. Start with an initial guess for the missing values, compute the (full rank) SVD, and then soft-threshold its singular values by an amount λ;

ii. Reconstruct the corresponding SVD approximation and obtain new estimates for the missing values;

iii. Repeat the preceding step until convergence.

(b) *Projection Operator:* Given an observed subset Ω of indices of matrix entries, we define the *projection operator* $\mathcal{P}_\Omega : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ as

$$[\mathcal{P}_\Omega(\mathbf{Z})]_{i,j} = \begin{cases} z_{i,j}, & \text{if } (i,j) \in \Omega, \\ 0, & \text{if } (i,j) \notin \Omega. \end{cases}$$

Then, we have

$$\sum_{(i,j)\in\Omega} (z_{i,j} - m_{i,j})^2 = \|\mathcal{P}_\Omega(\mathbf{Z}) - \mathcal{P}_\Omega(\mathbf{M})\|_{\mathrm{F}}^2$$

(c) *Soft-thresholded Version of a Matrix:* Let $\mathbf{W}$ be a matrix of rank r whose SVD is given by $\mathbf{W} = \mathbf{U}\mathbf{D}\mathbf{V}^\top$. Its *soft-thresholded* version is

$$\mathcal{S}_\lambda(\mathbf{W}) = \mathbf{U}\mathbf{D}_\lambda\mathbf{V}^\top,$$

where

$$\mathbf{D}_\lambda = \operatorname{diag}\big((d_1-\lambda)_+, (d_2-\lambda)_+, \cdots, (d_r-\lambda)_+\big).$$

(d) *Soft-impute Algorithm for Matrix Completion:* The following algorithm solves (8).

Algorithm 1 Soft-impute for Matrix Completion

1: Initialize $\mathbf{Z}^{\text{old}} = \mathbf{0}_{m\times n}$ and create a decreasing grid $\lambda_1 > \lambda_2 > \cdots > \lambda_K$;
2: For each $k = 1, 2, \cdots, K$, set $\lambda = \lambda_k$ and iterate until convergence:

- Compute $\widehat{\mathbf{Z}}_\lambda \leftarrow \mathcal{S}_\lambda(\mathcal{P}_\Omega(\mathbf{Z}) + \mathcal{P}_\Omega^\perp(\mathbf{Z}^{\text{old}}))$;
- Update $\mathbf{Z}^{\text{old}} \leftarrow \widehat{\mathbf{Z}}_\lambda$;

3: Output the sequence of solutions $\widehat{\mathbf{Z}}_{\lambda_1}, \widehat{\mathbf{Z}}_{\lambda_2}, \cdots, \widehat{\mathbf{Z}}_{\lambda_K}$.

Remark. Each iteration requires an SVD computation of a (potentially large) dense matrix, namely, $\mathcal{S}_\lambda(\mathcal{P}_\Omega(\mathbf{Z}) + \mathcal{P}_\Omega^\perp(\mathbf{Z}^{\text{old}}))$, even though $\mathcal{P}_\Omega(\mathbf{Z})$ is sparse. Note that we can write

$$\mathcal{P}_\Omega(\mathbf{Z}) + \mathcal{P}_\Omega^\perp(\mathbf{Z}^{\text{old}}) = \underbrace{\mathcal{P}_\Omega(\mathbf{Z}) - \mathcal{P}_\Omega(\mathbf{Z}^{\text{old}})}_{\text{sparse}} + \underbrace{\mathbf{Z}^{\text{old}}}_{\text{low rank}},$$

where

- the first component is sparse, with $|\Omega|$ non-missing entries, and
- the second component is a soft-thresholded SVD, so can be represented using the corresponding components.

IV. Maximum Margin Factorization and Related Methods

1. **Overview:** The *maximum margin matrix factorization* (MMMF) uses a factor model to approximate a given matrix.

2. **Problem Formulation:** Consider a matrix factorization of the form $\mathbf{M} = \mathbf{A}\mathbf{B}^\top$, where $\mathbf{A} \in \mathbb{R}^{m\times r}$ and $\mathbf{B} \in \mathbb{R}^{n\times r}$. One way to estimate such a factorization is to solve the following optimization problem

$$\underset{\mathbf{A}\in\mathbb{R}^{m\times r},\mathbf{B}\in\mathbb{R}^{n\times r}}{\text{minimize}} \left\{ \|\mathcal{P}_\Omega(\mathbf{Z}) - \mathcal{P}_\Omega(\mathbf{A}\mathbf{B}^\top)\|_\mathrm{F}^2 + \lambda(\|\mathbf{A}\|_\mathrm{F}^2 + \|\mathbf{B}\|_\mathrm{F}^2) \right\}, \tag{9}$$

   and the resulting factorization is called the *maximum margin matrix factorization*.

3. **An Equivalent Way of Expressing Nuclear Norm:** For any matrix $\mathbf{M} \in \mathbb{R}^{m\times n}$, the following identity holds

$$\|\mathbf{M}\|_* = \min_{\mathbf{A}\in\mathbb{R}^{m\times r},\mathbf{B}\in\mathbb{R}^{n\times r},\mathbf{M}=\mathbf{A}\mathbf{B}^\top} \left\{ \frac{1}{2}\big(\|\mathbf{A}\|_\mathrm{F}^2 + \|\mathbf{B}\|_\mathrm{F}^2\big) \right\}. \tag{10}$$

   Remark. The solution to (10) is *not unique*.

4. **Theorem — Connection between (8) and (9):** Let $\mathbf{Z}$ be an $m \times n$ matrix with observed entries indexed by Ω.

   (a) The solutions to the MMMF criterion (9) with $r = \min\{m, n\}$ and the nuclear norm regularized criterion (8) coincide for all $\lambda \geq 0$;

   (b) The solution space of the objective (8) is contained in that of (9). More precisely, for some fixed $\lambda^* > 0$, suppose that the objective (8) has an optimal solution with rank r^*. Then, for any optimal solution $(\widehat{\mathbf{A}}, \widehat{\mathbf{B}})$ to the problem (9) with $r \geq r^*$ and $\lambda = \lambda^*$, the matrix $\widehat{\mathbf{M}} = \widehat{\mathbf{A}}\widehat{\mathbf{B}}^\top$ is an optimal solution for the problem (8).

   Remark. The MMMF criterion (9) defines a two-dimensional family of models indexed by the pair (r, λ), while the soft-impute criterion (8) defines a one-dimensional family.

   According to the preceding theorem, the latter one-dimensional family is a special path in the two-dimensional grid of solutions $(\widehat{\mathbf{A}}_{(r,\lambda)}, \widehat{\mathbf{B}}_{(r,\lambda)})$.

5. **Comparison of (8) and (9):**

   (a) The formulation (8) is preferable, since it is convex and it does both rank reduction and regularization at the same time.

   (b) Using (9), we need to choose both the rank of the approximation and the regularization parameter λ.

6. **A Related Problem to (9):** A related problem to (9) in the literature is the following one

$$\underset{\mathbf{U},\mathbf{S},\mathbf{V}}{\text{minimize}} \left\{ \|\mathcal{P}_\Omega(\mathbf{Z}) - \mathcal{P}_\Omega(\mathbf{U}\mathbf{S}\mathbf{V}^\top)\|_\mathrm{F}^2 + \lambda \|\mathbf{S}\|_\mathrm{F}^2 \right\}, \tag{11}$$

where $\mathbf{U}$ and $\mathbf{V}$ satisfy $\mathbf{U}^\top\mathbf{U} = \mathbf{V}^\top\mathbf{V} = \mathbf{I}_r$ and $\mathbf{S} \in \mathbb{R}^{r\times r}$.

For a fixed rank r, the problem (11) can be solved by gradient descent.

Remark 1. The problem (11) is similar to the original MMMF problem (9), except that the matrices $\mathbf{U}$ and $\mathbf{V}$ are constrained to be orthonormal so that the "signal" and corresponding regularization are shifted to the (full) matrix $\mathbf{S}$.

Remark 2. Like MMMF, the problem (11) is non-convex so that gradient descent is *not* guaranteed to converge to the global optimum. In addition, it must be solved separately for different values of the rank r.

V. Penalized Matrix Decomposition

1. **Problem Formulation:** Given a matrix $\mathbf{Z} \in \mathbb{R}^{m\times n}$ that has *no* missing values, inspired by the maximum margin matrix factorization (9), we consider the following optimization problem

$$\underset{\mathbf{U}\in\mathbb{R}^{m\times r},\mathbf{V}\in\mathbb{R}^{n\times r},\mathbf{D}\in\mathbb{R}^{r\times r}}{\text{minimize}} \left\{ \|\mathbf{Z} - \mathbf{U}\mathbf{D}\mathbf{V}^\top\|_\mathrm{F}^2 + \lambda_1 \|\mathbf{U}\|_1 + \lambda_2 \|\mathbf{V}\|_2 \right\}, \tag{12}$$

where $\mathbf{D}$ is diagonal and non-negative and $\mathbf{U}^\top\mathbf{U} = \mathbf{V}^\top\mathbf{V} = \mathbf{I}_r$.

Remark. With the L^1-penalty on $\mathbf{U}$ and $\mathbf{V}$, we can obtain sparse versions of the singular vectors for easier interpretation.

2. **Problem (12) in $r = 1$ Case — Version 1:** Consider the one-dimensional case of (12) in the constrained form

$$\begin{aligned} &\underset{\mathbf{u}\in\mathbb{R}^m,\mathbf{v}\in\mathbb{R}^n,d\geq 0}{\text{minimize}} \left\{ \|\mathbf{Z} - d\mathbf{u}\mathbf{v}^\top\|_\mathrm{F}^2 \right\} \\ &\text{subject to } \ \|\mathbf{u}\|_1 \leq c_1, \|\mathbf{u}\|_2 = 1, \|\mathbf{v}\|_1 \leq c_2, \|\mathbf{v}\|_2 = 1. \end{aligned} \tag{13}$$

The issues with (13) are the following:

(a) it tends to produce solutions that are too sparse, and

(b) it is *not* convex due to the constraints $\|\mathbf{u}\|_2 = 1$ and $\|\mathbf{v}\|_2 = 1$.

To a rough idea of the first issue, consider the possibly simplest case where $m = 2$ and $n = 1$, and the resulting problem can be written as

$$\begin{aligned} &\underset{\mathbf{u}\in\mathbb{R}^2,v\in\mathbb{R},d\geq 0}{\text{minimize}} \left\{ \|\mathbf{Z} - dv\mathbf{u}\|_\mathrm{F}^2 \right\} \\ &\text{subject to } \ \|\mathbf{u}\|_1 \leq c_1, \|\mathbf{u}\|_2 = 1, |v| = 1, \end{aligned}$$

or, equivalently,

$$\begin{aligned}&\underset{u_1,u_2\in\mathbb{R}^m,v\in\mathbb{R},d\geq 0}{\text{minimize}} \left\{\left\|\begin{pmatrix}z_1\\z_2\end{pmatrix} - dv\begin{pmatrix}u_1\\u_2\end{pmatrix}\right\|_{\mathrm{F}}^2\right\}\\ &\text{subject to} \;\; |u_1|+|u_2|\leq c_1, u_1^2+u_2^2=1, |v|=1,\end{aligned}$$

that is,

$$\begin{aligned}&\underset{u_1,u_2\in\mathbb{R}^m,v\in\mathbb{R},d\geq 0}{\text{minimize}} \left\{(dvu_1-z_1)^2+(dvu_2-z_2)^2\right\}\\ &\text{subject to} \;\; |u_1|+|u_2|\leq c_1, u_1^2+u_2^2=1, |v|=1.\end{aligned}$$

With $c_1 \in [1, \sqrt{2}]$, it is easy to see that the feasible set over $(u_1, u_2, v)^\top \in \mathbb{R}^3$ is the union of 8 arcs that each intersect with $u_1 = 0$ or $u_2 = 0$ in $\mathbb{R}^3$. Typically, the optimal solution occurs when $u_1 = 0$ or $u_2 = 0$, resulting in a sparse solution.

3. **Equivalent Formulation of (13):** The objective function in (13) can be written equivalently as

$$\|\mathbf{Z} - d\mathbf{u}\mathbf{v}^\top\|_{\mathrm{F}}^2 = -2d\mathbf{u}^\top\mathbf{Z}\mathbf{v} + d^2\|\mathbf{u}\|_2^2\|\mathbf{v}\|_2^2 + \|\mathbf{Z}\|_{\mathrm{F}}^2.$$

Under the constraint $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$, we can simplify the preceding equation

$$\|\mathbf{Z} - d\mathbf{u}\mathbf{v}^\top\|_{\mathrm{F}}^2 = -2d\mathbf{u}^\top\mathbf{Z}\mathbf{v} + d^2 + \|\mathbf{Z}\|_{\mathrm{F}}^2.$$

For given $\mathbf{u}$ and $\mathbf{v}$, the value of d optimizing the preceding equation is $\mathbf{u}^\top\mathbf{Z}\mathbf{v}$. In addition, the last term $\|\mathbf{Z}\|_{\mathrm{F}}^2$ does *not* depend on $\mathbf{u}$ or $\mathbf{v}$. Hence, an equivalent formulation of (13) is

$$\begin{aligned}&\underset{\mathbf{u}\in\mathbb{R}^m,\mathbf{v}\in\mathbb{R}^n}{\text{maximize}} \; \mathbf{u}^\top\mathbf{Z}\mathbf{v}\\ &\text{subject to} \;\; \|\mathbf{u}\|_1\leq c_1, \|\mathbf{u}\|_2=1, \|\mathbf{v}\|_1\leq c_2, \|\mathbf{v}\|_2=1.\end{aligned} \tag{14}$$

We will only consider the problem (14) in the sequel.

Remark. Problem (14) still suffers the two issues we discussed earlier about (13).

4. **Problem (12) in $r = 1$ Case — Version 2:** In order to remedy the issues mentioned above about (13), we instead consider the following problem

$$\begin{aligned}&\underset{\mathbf{u}\in\mathbb{R}^m,\mathbf{v}\in\mathbb{R}^n}{\text{maximize}} \; \mathbf{u}^\top\mathbf{Z}\mathbf{v}\\ &\text{subject to} \;\; \|\mathbf{u}\|_1\leq c_1, \|\mathbf{v}\|_1\leq c_2, \|\mathbf{u}\|_2\leq 1, \|\mathbf{v}\|_2\leq 1.\end{aligned} \tag{15}$$

Remark 1. If we fix the component $\mathbf{v}$, the criterion (15) is linear in $\mathbf{u}$.

Remark 2. If we choose

$$1 \leq c_1 \leq \sqrt{m} \qquad \text{and} \qquad 1 \leq c_2 \leq \sqrt{n},$$

then the solution of (15) automatically satisfies $\|\mathbf{u}\|_2 = 1$ and $\|\mathbf{v}\|_2 = 1$. This follows from the Karush-Kuhn-Tucker conditions from convex optimization. Therefore, for c_1 and c_2 appropriately chosen, the solution to (15) solves (14).

Remark 3. The L^1 penalties above may be replaced by other kinds of penalties such as the fused lasso penalty

$$\Phi(\mathbf{u}) = \sum_{j=2}^{m} |u_j - u_{j-1}|,$$

where $\mathbf{u} = (u_1, u_2, \cdots, u_m)^\top \in \mathbb{R}^m$. This choice is useful in enforcing smoothness along the 1-dimensional ordering.

5. **Bi-convexity of Problem (15):** With $\mathbf{v}$ fixed, the problem (15) becomes

$$\begin{aligned} &\underset{\mathbf{u}\in\mathbb{R}^m}{\text{maximize}} \ \mathbf{u}^\top \mathbf{Z}\mathbf{v} \\ &\text{subject to } \|\mathbf{u}\|_1 \le c_1, \|\mathbf{u}\|_2 \le 1, \end{aligned} \tag{16}$$

which is convex. With $\mathbf{u}$ fixed, the resulting problem with respect to $\mathbf{v}$ is convex as well. This means that the problem (15) is *bi-convex*, and suggests an alternating algorithm for optimizing it.

6. **Characterizing the Solution to (16):** Using standard results from convex optimization, the solution to (16), denoted by $\mathbf{u}^*$, is given by

$$\mathbf{u}^* = \frac{\mathcal{S}_\lambda(\mathbf{Z}\mathbf{v})}{\|\mathcal{S}_\lambda(\mathbf{Z}\mathbf{v})\|_2},$$

with λ being the smallest positive value such that $\|\mathbf{u}^*\|_1 = c_1$. Here, $\mathcal{S}_\lambda$ is the soft-thresholding operator applied to each component of the vector $\mathbf{Z}\mathbf{v}$.

7. **Algorithm to Solve (15):** With the results above, we minimize (15) in an alternating fashion. The resulting algorithm is shown in Algorithm 2.

Algorithm 2 Alternating Soft-Thresholding for Rank-1 Penalized Matrix Decomposition

1: Set $\mathbf{v}$ to the top left singular vector from the SVD of $\mathbf{Z}$;
2: Perform the update

$$\mathbf{u} \quad \longleftarrow \quad \frac{\mathcal{S}_{\lambda_1}(\mathbf{Zv})}{\|\mathcal{S}_{\lambda_1}(\mathbf{Zv})\|_2},$$

with λ_1 being the smallest positive value such that $\|\mathbf{u}\|_1 \leq c_1$;
3: Perform the update

$$\mathbf{v} \quad \longleftarrow \quad \frac{\mathcal{S}_{\lambda_2}(\mathbf{Z}^\top\mathbf{u})}{\|\mathcal{S}_{\lambda_2}(\mathbf{Z}^\top\mathbf{u})\|_2},$$

with λ_2 being the smallest positive value such that $\|\mathbf{v}\|_1 \leq c_2$;
4: Iterate the preceding two steps until convergence;
5: **return** $\mathbf{u}$, $\mathbf{v}$ and $d = \mathbf{u}^\top\mathbf{Zv}$.

Remark. If $c_1 > \sqrt{m}$ and $c_2 > \sqrt{n}$, then the L^1 constraints have no effect.

8. **Multi-factor Penalized Matrix Decomposition:** Algorithm 2 leads the decomposition of $\mathbf{Z}$ with a single factor.

   To obtain a decomposition of K factors, we can apply Algorithm 2 K times, which leads to the following K-factor penalized matrix decomposition algorithm.

Algorithm 3 Multi-factor Penalized Matrix Decomposition

1: Set $\mathbf{R} \leftarrow \mathbf{Z}$;
2: For $k = 1, 2, \cdots, K$:

   (a) Find $\mathbf{u}_k$, $\mathbf{v}_k$, and d_k by applying the single-factor algorithm (Algorithm 2 to data $\mathbf{R}$;

   (b) Update $\mathbf{R} \leftarrow \mathbf{R} - d_k\mathbf{u}_k\mathbf{v}_k^\top$.

Remark 1. If we omit L^1-penalties on $\mathbf{u}_k$ and $\mathbf{v}_k$ (or equivalently, set $\lambda_1 = \lambda_2 = 0$), Algorithm 3 leads to the rank-K SVD of $\mathbf{Z}$. In particular, the successive solutions are orthogonal.

However, if we do impose L^1 penalties, the resulting solutions are *not* orthogonal.

Remark 2. Alternating minimization of biconvex functions, unlike the minimization of convex functions, is *not* guaranteed to find a global optimum, and is only guaranteed to move downhill to a local minimum.

Remark 3. Differences between matrix completion and penalized matrix decomposition are the following:

(a) For successful matrix completion, the singular vectors of $\mathbf{Z}$ need to be dense;

(b) In sparse matrix decomposition, we seek sparse singular vectors for interpretability.

VI. Additive Matrix Decomposition

1. **Overview:** In the problem of additive matrix decomposition, we seek to decompose a matrix into the sum of two or more matrices.

   Remark. The components in the additive decomposition should have *complementary* structures. For instance, if we decompose a matrix into a sum of two matrices, one component can have low rank and the other one is sparse.

2. **Applications:** Applications of additive matrix decompositions include factor analysis, and robust forms of PCA and matrix completion.

3. **Problem Formulation:** Given a matrix $\mathbf{Z} \in \mathbb{R}^{m\times n}$, we decompose it as

$$\mathbf{Z} = \mathbf{L} + \mathbf{S} + \mathbf{W},$$

   where $\mathbf{L} \in \mathbb{R}^{m\times n}$ is a low rank matrix, $\mathbf{S} \in \mathbb{R}^{m\times n}$ is a sparse matrix, and $\mathbf{W} \in \mathbb{R}^{m\times n}$ is a noise matrix. This leads to the following optimization problem

$$\underset{\mathbf{L}\in\mathbb{R}^{m\times n},\mathbf{S}\in\mathbb{R}^{m\times n}}{\text{minimize}} \left\{\frac{1}{2}\|\mathbf{Z} - (\mathbf{L} + \mathbf{S})\|_{\mathrm{F}}^2 + \lambda_1\Phi_1(\mathbf{L}) + \lambda_2\Phi_2(\mathbf{S})\right\}, \tag{17}$$

   with $\Phi_1(\mathbf{L}) = \|\mathbf{L}\|_*$ enforcing the low rank and $\Phi_2(\mathbf{S}) = \|\mathbf{S}\|_1$ enforcing the sparsity.

4. **Application 1 — Factor Analysis with Sparse Noise:**

   (a) *Setup:* We regard factor analysis as a generative model and let $Y_i \in \mathbb{R}^p$, for all $i = 1, 2, \cdots, n$, be generated as the following mechanism

$$Y_i = \boldsymbol{\mu} + \boldsymbol{\Gamma} U_i + W_i, \tag{18}$$

   where

   - $\boldsymbol{\mu} \in \mathbb{R}^p$ is the mean vector,
   - $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times r}$ is a (unknown) loading matrix,
   - $U_i \overset{\text{i.i.d}}{\sim} \text{Normal}_r(\mathbf{0}_r, \mathbf{I}_{r\times r})$,
   - $W_i \overset{\text{i.i.d}}{\sim} \text{Normal}_p(\mathbf{0}_p, \mathbf{S}^*)$, and
   - U_i and W_i are independent.

   (b) *Goal:* Given $Y_1, Y_2, \cdots, Y_n$, the goal is to estimate the column of the loading matrix $\boldsymbol{\Gamma}$, or, equivalently, the rank r matrix $\mathbf{L}^* = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top \in \mathbb{R}^{p\times p}$ that spans the column space of $\boldsymbol{\Gamma}$.

(c) *Variance of Y_i:* By the model (18), it is easy to see

$$\mathbf{\Sigma} := \mathrm{Var}[Y_i] = \mathbf{\Gamma}\mathbf{\Gamma}^\top + \mathbf{S}^*, \qquad \text{for all } i = 1, 2, \cdots, n.$$

(d) *Special Case:* If $\mathbf{S}^* = \sigma^2 \mathbf{I}_{p\times p}$, then the column span of $\mathbf{\Gamma}$ is equivalent to the span of the top r eigenvectors of $\mathbf{\Sigma}$, and we can recover it via standard principal component analysis.

(e) *When $\mathbf{S}^*$ Is Sparse:* Assume $\boldsymbol{\mu} = \mathbf{0}_p$. When $\mathbf{S}^*$ is a sparse matrix, with $Y_1, Y_2, \cdots, Y_n$ from the model (18), we can let

$$\mathbf{Z} = \frac{1}{n}\sum_{i=1}^{n} Y_i Y_i^\top \in \mathbb{R}^{p\times p}$$

be the sample covariance matrix and write $\mathbf{Z} = \mathbf{L}^* + \mathbf{S}^* + \mathbf{W}$, where $\mathbf{L}^* = \mathbf{\Gamma}\mathbf{\Gamma}^\top$ is of rank r. We can then estimate $\mathbf{L}^*$ and $\mathbf{S}^*$ by the problem (17).

5. **Application 2 — Robust PCA:**

(a) *Review of Standard PCA:* Let $\mathbf{Z} \in \mathbb{R}^{n\times p}$ be the data matrix, where the i-th row represents the i-th sample of a p-dimensional data vector. Standard PCA can be formulated as the problem of minimizing

$$\|\mathbf{Z} - \mathbf{L}\|_{\mathrm{F}}^2$$

subject to a rank constraint on $\mathbf{L}$.

(b) *Motivation of Robust PCA:* If some entries or rows of the data matrix $\mathbf{Z}$ is corrupted, standard PCA may be very sensitive to the perturbations of data.

(c) *Idea of Robust PCA:* Additive matrix decompositions provide one solution that introduces robustness to PCA. Let $\mathbf{L}$ be a low-rank matrix and $\mathbf{S}$ be a sparse matrix. We approximate $\mathbf{Z}$ with the sum $\mathbf{L} + \mathbf{S}$.

Remark. The specification of the sparse matrix $\mathbf{S}$ depends on the corruption type of $\mathbf{Z}$.

i. In the case of element-wise corruption, $\mathbf{S}$ would be modeled as being element-wise sparse, having relatively few nonzero entries;

ii. In the case of having entirely corrupted rows, $\mathbf{S}$ would be modeled as a row-sparse matrix.

(d) *Naive Optimization Problem:* Given some target rank r and sparsity k, robust PCA can be formulated as the following optimization problem

$$\begin{aligned} &\underset{\mathbf{L},\mathbf{S}}{\text{minimize}} \ \frac{1}{2}\|\mathbf{Z} - (\mathbf{L} + \mathbf{S})\|_{\mathrm{F}}^2, \\ &\text{subject to } \ \mathrm{rank}(\mathbf{L}) \le r, \mathrm{card}(\mathbf{S}) \le k, \end{aligned} \tag{19}$$

where "card" denotes a cardinality constraint, either

i. the total number of nonzero entries (in the case of element-wise corruption), or

ii. the total number of nonzero rows (in the case of row-wise corruption).

(e) *Convex Relaxation of* (19)*:* Note that the problem (19) is *doubly non-convex*, due to both the rank and cardinality constraints.

A natural convex relaxation is to replace the low-rank constraint by

$$\Phi_1(\mathbf{L}) = \|\mathbf{L}\|_*$$

and the sparsity constraint by

$$\Phi_2(\mathbf{S}) = \sum_{i=1}^{n} \sum_{j=1}^{p} |s_{i,j}|$$

for the element-wise sparsity or by

$$\Phi_2(\mathbf{S}) = \sum_{i=1}^{n} \|\mathbf{s}_i\|_2$$

for the row-wise sparsity, where $\mathbf{s}_i \in \mathbb{R}^p$ denotes the i-th row of $\mathbf{S}$.

6. **Application 3 — Robust Matrix Completion:** Similar to robust PCA presented above, we can also introduce a sparse component $\mathbf{S}$ to the optimization problem (8) which adds robustness to the matrix completion.

   Let $\mathbf{Z} \in \mathbb{R}^{m\times n}$ be the matrix containing missing values. We impose the following row-wise sparsity penalty to (8)

$$\Phi(\mathbf{S}) = \sum_{i=1}^{m} \|\mathbf{s}_i\|_2,$$

   where $\mathbf{s}_i \in \mathbb{R}^n$ denotes the i-th row of $\mathbf{S}$. Then, the optimization (8) can be modified as

$$\underset{\mathbf{L},\mathbf{S}\in\mathbb{R}^{m\times n}}{\text{minimize}} \left\{ \frac{1}{2} \sum_{(i,j)\in\Omega} \big(z_{i,j} - (l_{i,j} + s_{i,j})\big)^2 + \lambda_1 \|\mathbf{L}\|_* + \lambda_2 \Phi(\mathbf{S}) \right\},$$

   where $l_{i,j}$ and $s_{i,j}$ denote the (i,j)-th entry of the matrices $\mathbf{L}$ and $\mathbf{S}$, respectively.

References

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman & Hall/CRC. ISBN: 1498712169.