

1 Original Problem

The original problem is:

$$\begin{aligned} \min_{S,Y} \quad & \|X - ASY\|_F^2 + \alpha \|Y\|_F^2 \\ \text{s.t.} \quad & S^T S = I, S \geq 0 \end{aligned} \quad (1)$$

where $\mathbf{X} \in \mathbf{R}^{Dim \times n}$, $\mathbf{A} \in \mathbf{R}^{Dim \times m}$, $\mathbf{Y} \in \mathbf{R}^{k \times n}$. And \mathbf{S} is a select matrix, $\mathbf{S} \in \mathbf{R}^{m \times k}$, $m \geq k$.

2 Transformed Problem

The original problem is strongly NP-hard and also difficult to solve approximately. A popular method for this original problem is the Augmented Lagrange Multiplier Method (ALM). The original problem is transformed below:

$$\begin{aligned} \min_{Y,S,Q,J,Z_1,Z_2} \quad & \|X - ASY\|_F^2 + \alpha \|Y\|_F^2 + \langle Z_1, S - Q \rangle \\ & + \frac{1}{2}\mu \|S - Q\|_F^2 + \langle Z_2, S - J \rangle + \frac{1}{2}\mu \|S - J\|_F^2 \\ \text{s.t.} \quad & S = Q, S = J, Q^T Q = I, J \geq 0 \end{aligned} \quad (2)$$

where $\mathbf{S}, \mathbf{Q}, \mathbf{J} \in \mathbf{R}^{m \times k}$.

3 Solution

The ALM can be decomposed into 4 subproblems. All this 4 subproblems have close-formed solutions. So we can do the optimization by 4 steps.

Steps 1: Fix S, Q, J to optimize Y and ignore the constant items, We can get:

$$\min_Y \|X - ASY\|_F^2 + \alpha \|Y\|_F^2 \quad (3)$$

We can get the first-order differential of this formulation, and let it to be zero:

$$S^T A^T (ASY - X) + \alpha Y = 0$$

$$(S^T A^T ASY + \alpha I) = S^T A^T X$$

$$Y = (S^T A^T ASY + \alpha I)^{-1} S^T A^T X$$

Now we can get Y .

Steps 2: Fix Y, Q, J to optimize S and ignore the constant items, We can get:

$$\begin{aligned} \min_S \quad & \|X - ASY\|_F^2 + \langle Z_1, S - Q \rangle + \frac{1}{2}\mu \|S - Q\|_F^2 \\ & + \langle Z_2, S - J \rangle + \frac{1}{2}\mu \|S - J\|_F^2 \end{aligned} \quad (4)$$

We can get the first-order differential of this formulation, and let it to be zero:

$$2A^T(ASY - X)Y^T + Z_1 + \mu(S - Q) + Z_2 + \mu(S - J) = 0$$

$$2A^TASY Y^T + 2\mu S = 2A^TXY^T + \mu(Q + J)$$

Use the property of *KroneckerProduct*.

$$\text{Vec}(C \cdot S \cdot B) = (B^T \otimes C) \cdot \text{Vec}(S)$$

We can get:

$$2(Y Y^T \otimes A^T A) \cdot \text{Vec}(S) + 2\mu \text{Vec}(S) = \text{Vec}(2A^TXY^T + \mu(Q + J))$$

$$2(Y Y^T \otimes A^T A + \mu I) \cdot \text{Vec}(S) = \text{Vec}(2A^TXY^T + \mu(Q + J))$$

$$\text{Vec}(S) = 2(Y Y^T \otimes A^T A + \mu I)^{-1} \cdot \text{Vec}(2A^TXY^T + \mu(Q + J))$$

Now We need to convert $\text{Vector}(S)$ back to a matrix which is $m \times k$.

Steps 3: Fix S, Y, J to optimize Q and ignore the constant items, We can get:

$$\begin{aligned} \min_Q \quad & \langle Z_1, S - Q \rangle + \frac{1}{2}\mu \|S - Q\|_F^2 \\ \text{s.t.} \quad & Q^T Q = I \end{aligned} \quad (5)$$

It equals to:

$$\begin{aligned} \min_Q \quad & \langle Z_1, -Q \rangle + \mu \langle S, Q \rangle \\ \text{s.t.} \quad & Q^T Q = I \end{aligned} \quad (6)$$

\Longleftrightarrow

$$\begin{aligned} \max_Q \quad & \langle Q, \mu S + Z_1 \rangle \\ \text{s.t.} \quad & Q^T Q = I \end{aligned} \quad (7)$$

Use the proposition of matrix:

The dual norm of the operator norm $\|\cdot\|$ in $R^{m \times n}$ is the nuclear norm $\|\cdot\|_*$.
 \Rightarrow

$$\|X\|_* := \max_Y \{ \langle X, Y \rangle : \|Y\| \leq 1 \} \quad (8)$$

Now let $X = U\Sigma V^T$ be a **thin** singular value decomposition of the $m \times n$ matrix X , where U is an $m \times m$ matrix and $U^T \cdot U = I$, but $U \cdot U^T \neq I$. V is an $n \times n$ orthogonal matrix which $V \cdot V^T = I$. Σ is an $m \times n$ matrix. Let $Y := U \cdot V^T$, $\|Y\| = 1$ and $Tr(Y^T \cdot X) = Tr(VU^T \cdot U\Sigma V^T) = Tr(\Sigma) = \|X\|_*$. And can get $Y^T \cdot Y = I$.

So from this derivation, We can also do a **thin** singular value decomposition of $\mu S + Z_1$:

$$\mu S + Z_1 = U\Sigma V^T$$

$$Q = U \cdot V^T$$

Steps 4: Fix S, Y, Q to optimize J and ignore the constant items, We can get:

$$\begin{aligned} \min_J & \langle Z_2, S - J \rangle + \frac{1}{2}\mu \|S - J\|_F^2 \\ \text{s.t. } & J \geq 0 \end{aligned} \quad (9)$$

\Leftrightarrow

$$\begin{aligned} \min_J & \frac{2}{\mu} \langle Z_2, S - J \rangle + \|S - J\|_F^2 + \frac{1}{\mu} \|Z_2\|_F^2 \\ \text{s.t. } & J \geq 0 \end{aligned} \quad (10)$$

\Leftrightarrow

$$\begin{aligned} \min_J & \|J - (S + \frac{Z_2}{\mu})\|_F^2 \\ \text{s.t. } & J \geq 0 \end{aligned} \quad (11)$$

\Leftrightarrow

$$J = \frac{1}{2} \cdot [\text{abs}(S + \frac{Z_2}{\mu}) + (S + \frac{Z_2}{\mu})] \quad (12)$$

Steps 5: Do some updates:

$$\begin{aligned} Z_1 &= Z_1 + \mu \cdot (S - Q) \\ Z_2 &= Z_2 + \mu \cdot (S - J) \\ \mu &= \rho \times \mu \end{aligned} \quad (13)$$

where $\rho = 1.1$.

4 Termination Conditions

The algorithm will stop when the conditions are both satisfied:

$$\begin{aligned} \|S - J\|_F &< \varepsilon \\ \|S - Q\|_F &< \varepsilon \end{aligned} \quad (14)$$

where $\varepsilon = 10^{-8}$.