1 Original Problem

The original problem is :

$$\begin{aligned} &\min_{S,Y} \parallel X - ASY \parallel_F^2 \\ &s.t.\ \ S^T S = I,\ S_{ij} \in \{0,1\},\ Y_i \in \Omega \end{aligned} \tag{1}$$

Ω is a set in $\mathbf{R}^n$,it equals to $e_1, e_2, ..., e_n$,where e_i is the i-th column of the identity matrix.
where Y_i is a column of Matrix $\mathbf{Y}$,$\mathbf{X} \in \mathbf{R}^{Dim\times n}$, $\mathbf{A} \in \mathbf{R}^{Dim\times m}$, $\mathbf{Y} \in \mathbf{R}^{k\times n}$.And $\mathbf{S}$ is the selection matrix, $\mathbf{S} \in \mathbf{R}^{m\times k}$,$m \geq k$.

2 Relaxed Problem

The original problem is strongly NP-hard and also difficult to solve approximately.It is quite similar to the below one by relax it's constraint :

$$\begin{aligned} &\min_{S,Y} \parallel X - ASY \parallel_F^2 \\ &s.t.\ \ S^T S = I,\ S \geq 0\ Y_i \in \Omega \end{aligned} \tag{2}$$

The problem above is equal to :

$$\begin{aligned} &\min_{Y,S,} \qquad \parallel X - DM \parallel_F^2 \\ &s.t.\ \ A \cdot S = D,\ Y = M,\ S = Q,\ S = J,\ Q^T Q = I,\ J \geq 0,\ Y_i \in \Omega \end{aligned} \tag{3}$$

where Y_i is a column of Matrix $\mathbf{Y}$, $\mathbf{X} \in \mathbf{R}^{Dim\times n}$, $\mathbf{A} \in \mathbf{R}^{Dim\times m}$, $\mathbf{Y}, \mathbf{M} \in \mathbf{R}^{k\times n}$.And $\mathbf{S}$ is the selection matrix, $\mathbf{S} \in \mathbf{R}^{m\times k}$,$m \geq k$.

3 Solution

A popular method for this above problem is the Augmented Lagrange Multiplier Method(ALM).The original problem is transformed below :

$$\begin{aligned} \min_{Y,S,D,Q,J} \quad & \parallel X - DM \parallel_F^2 + < Z_1, S - Q > + < Z_2, S - J > \\ & + < Z_3, AS - D > + < Z_4, Y - M > \\ & + \tfrac{1}{2}\mu \parallel S - Q \parallel_F^2 + \tfrac{1}{2}\mu \parallel S - J \parallel_F^2 \\ & + \tfrac{1}{2}\mu \parallel AS - D \parallel_F^2 + \tfrac{1}{2}\mu \parallel Y - M \parallel_F^2 \\ s.t.\ \ & Q^T Q = I,\ J \geq 0,\ Y_i \in \Omega \end{aligned} \tag{4}$$

where Y_i is a column of Matrix $\mathbf{Y}$, $\mathbf{D} \in \mathbf{R}^{Dim \times k}$ and $\mathbf{S}, \mathbf{Q}, \mathbf{J} \in \mathbf{R}^{m \times k}$, $\mathbf{M} \in \mathbf{R}^{k \times n}$.

The ALM can be decomposed into 6 subproblems.All this 6 subproblems have close-formed solutions.So we can do the optimization by 6 steps.

Steps 1: Fix $S, D, Q, J, M, Z_1, Z_2, Z_3, Z_4$ to optimize Y and ignore the constant items, We can get:

$$\begin{aligned} &\min_{Y} < Z_4, Y - M > + \tfrac{1}{2}\mu \| \ Y - M \ \|_F^2 \\ &s.t. \ \ Y_i \in \Omega \end{aligned} \tag{5}$$

$\Longrightarrow$

$$\begin{aligned} &\min_{Y} \| \ Y - (M - \tfrac{Z_4}{\mu}) \ \|_F^2 \\ &s.t. \ \ Y_i \in \Omega \end{aligned} \tag{6}$$

$\Longrightarrow$

$$Y_{ij} = \begin{cases} 1 & \text{j = t} \\ 0 & \text{j} \neq \text{t} \end{cases} \tag{7}$$

where $\mathbf{t}$ is the index of max member of the i-th column of matrix $\mathbf{M} - \frac{\mathbf{Z_4}}{\mu}$ and $\mathbf{Y_{ij}}$ is the j-th member of i-th column of matrix $\mathbf{Y}$.
Now we get Y.

Steps 2: Fix $S, M, Y, Q, J, Z_1, Z_2, Z_3, Z_4$ to optimize D and ignore the constant items, We can get:

$$\min_{D} \quad \| \ X - DY \ \|_F^2 + < Z_3, AS - D > + \ \tfrac{1}{2}\mu \| \ AS - D \ \|_F^2 \tag{8}$$

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

$$\begin{aligned} &2(DY - X)Y^T - Z_3 + \mu(D - AS) = 0 \\ &2DY \cdot Y^T + \mu D = 2XY^T + \mu AS + Z_3 \\ &D(2YY^T + \mu I) = 2XY^T + \mu AS + Z_3 \\ &D = (2XY^T + \mu AS + Z_3) \cdot (2YY^T + \mu I)^{-1} \end{aligned}$$

Now we can get D.

Steps 3: Fix $D, Y, M, Q, J, Z_1, Z_2, Z_3, Z_4$ to optimize S and ignore the constant items, We can get:

$$\begin{aligned} &\min_S < Z_1, S-Q > + < Z_2, S-J > + < Z_3, AS-D > \\ &+ \tfrac{1}{2}\mu \| S-Q \|_F^2 + \tfrac{1}{2}\mu \| S-J \|_F^2 + \tfrac{1}{2}\mu \| AS-D \|_F^2 \end{aligned} \tag{9}$$

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

$$Z_1 + Z_2 + A^T \cdot Z_3 + \mu(S-Q) + \mu(S-J) + \mu A^T(AS-D) = 0$$

$$2\mu \cdot S + \mu \cdot A^T AS = \mu(A^T D + Q + J) - A^T \cdot Z_3 - Z_1 - Z_2$$

$$S = (2I + A^T A)^{-1} \cdot (A^T D + Q + J) - \tfrac{1}{\mu} \cdot (A^T \cdot Z_3 + Z_1 + Z_2))$$

Now we can get S.

Steps 4: Fix $D, Y, S, Q, J, Z_1, Z_2, Z_3, Z_4$ to optimize M and ignore the constant items, We can get:

$$\min_M \| X - DM \|_F^2 + < Z_4, Y-M > + \tfrac{1}{2}\mu \| Y-M \|_F^2 \tag{10}$$

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

$$2D^T(DM - X) - Z_4 + \mu(M-Y) = 0$$

$$2D^T D \cdot M + \mu M = Z_4 + \mu Y + 2D^T X$$

$$M = (2D^T D + \mu I)^{-1} \cdot (Z_4 + \mu Y + 2D^T X)$$

Now we can get M.

Steps 5: Fix $D, Y, M, S, J, Z_1, Z_2, Z_3, Z_4$ to optimize Q and ignore the constant items, We can get:

$$\begin{aligned} &\min_Q < Z_1, S-Q > + \tfrac{1}{2}\mu \| S-Q \|_F^2 \\ &s.t.\ \ Q^T Q = I \end{aligned} \tag{11}$$

It equals to:

$$\begin{aligned} &\min_Q < Z_1, -Q > - \mu < S, Q > \\ &s.t.\ \ Q^T Q = I \end{aligned} \tag{12}$$

$\Longleftrightarrow$

$$\begin{aligned} &\max_Q < Q, \mu S + Z_1 > \\ &s.t.\ \ Q^T Q = I \end{aligned} \tag{13}$$

Use the proposition of matrix:

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

$$\| X \|_* := \max_Y \{< X, Y > \ : \ \| Y \| \leq 1\} \tag{14}$$

Now let $X = U\Sigma V^T$ be a **thin** singular value decomposition of the $m \times n$ matrix X,where U is an $m \times n$ 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 $n \times n$ 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 6: Fix $D, Y, S, M, Q, Z_1, Z_2, Z_3, Z_4$ to optimize J and ignore the constant items, We can get:

$$\begin{aligned} &\min_J < Z_2, S - J > + \tfrac{1}{2}\mu \| S - J \|_F^2 \\ &s.t. \ \ J \geq 0 \end{aligned} \tag{15}$$

$\Longleftrightarrow$

$$\begin{aligned} &\min_J \tfrac{2}{\mu} < Z_2, S - J > + \| S - J \|_F^2 + \tfrac{1}{\mu} \| Z_2 \|_F^2 \\ &s.t. \ \ J \geq 0 \end{aligned} \tag{16}$$

$\Longleftrightarrow$

$$\begin{aligned} &\min_J \| J - (S + \tfrac{Z_2}{\mu}) \|_F^2 \\ &s.t. \ \ J \geq 0 \end{aligned} \tag{17}$$

$\Longleftrightarrow$

$$J = \tfrac{1}{2} \cdot \ [abs(S + \tfrac{Z_2}{\mu}) + (S + \tfrac{Z_2}{\mu})] \tag{18}$$

Steps 7: Do some updates:

$$\begin{aligned} &Z_1 = Z_1 + \mu \cdot (S - Q) \\ &Z_2 = Z_2 + \mu \cdot (S - J) \\ &Z_3 = Z_3 + \mu(AS - D) \\ &Z_4 = Z_4 + \mu(Y - M) \\ &\mu = \begin{cases} \rho \times \mu & \mu \leq 10^{\,30} \\ 10^{\,30} & \text{else} \end{cases} \end{aligned} \tag{19}$$

where $\rho = 1.1$.

4 Termination Conditions

The algorithm will stop when the conditions are both satisfied:

$$\begin{aligned}
&\| S - Q \|_{\infty} < \varepsilon \\
&\| S - J \|_{\infty} < \varepsilon \\
&\| Y - M \|_{\infty} < \varepsilon \\
&\| A \cdot S - D \|_{F} < \varepsilon \times \| A \|_{F}
\end{aligned} \tag{20}$$

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