1 Original Problem

The original problem is :

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

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 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 :

$$\begin{aligned}&\min_{S,Y} \parallel X - ASY \parallel_F^2 + \alpha \parallel Y \parallel_F^2\\&s.t.\ \ S^TS = I,\ S \geq 0\end{aligned} \tag{2}$$

The problem above is equal to :

$$\min_{Y,S,K,Q,J,Z_1,Z_2,Z_3} \qquad \parallel X - KY \parallel_F^2 + \alpha \parallel Y \parallel_F^2 \tag{3}$$

$$s.t.\ \ A \cdot S = K,\ S = Q,\ S = J,\ Q^TQ = I,\ J \geq 0$$

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 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 :

$$\begin{aligned}\min_{Y,S,k,Q,J,Z_1,Z_2,Z_3} \qquad &\parallel X - KY \parallel_F^2 + \alpha \parallel Y \parallel_F^2 + < Z_1, S - Q >\\&+ < Z_2, S - J > + < Z_3, X \cdot S - K >\\&+ \tfrac{1}{2}\mu \parallel S - Q \parallel_F^2 + \tfrac{1}{2}\mu \parallel S - J \parallel_F^2\\&+ \tfrac{1}{2}\mu \parallel X \cdot S - K \parallel_F^2\\s.t.\ \ A \cdot S = K,\ S = Q,\ &S = J,\ Q^TQ = I,\ J \geq 0\end{aligned} \tag{4}$$

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

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 $S, K, Q, J, Z_1, Z_2, Z_3$ to optimize Y and ignore the constant items, We can get:

$$\min_{Y} \| X - KY \|_F^2 + \alpha \| Y \|_F^2 \tag{5}$$

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

$$K^T \cdot (KY - X) + \alpha Y = 0$$

$$(K^T K + \alpha I) \cdot Y = K^T \cdot X$$

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

Now we can get Y.

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

$$\min_{K} \quad \| X - KY \|_F^2 + < Z_3, AS - K > + \frac{1}{2}\mu \| AS - K \|_F^2 \tag{6}$$

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

$$2(KY - X)Y^T - Z_3 + \mu(K - AS) = 0$$

$$2K \cdot Y \cdot Y^T + \mu K = 2XY^T + \mu AS + Z_3$$

$$K(2YY^T + \mu I) = 2XY^T + \mu AS + Z_3$$

$$K = (2XY^T + \mu AS + Z_3) \cdot (2YY^T + \mu I)^{-1}$$

Now we can get K.

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

$$\begin{aligned} &\min_{S} < Z_1, S - Q > + < Z_2, S - J > + < Z_3, AS - K > \\ &+ \frac{1}{2}\mu \| S - Q \|_F^2 + \frac{1}{2}\mu \| S - J \|_F^2 + \frac{1}{2}\mu \| AS - K \|_F^2 \end{aligned} \tag{7}$$

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

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

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

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

Now we can get S.

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

$$\begin{aligned} & \min_{Q} < Z_1, S - Q > + \tfrac{1}{2}\mu \| S - Q \|_F^2 \\ & s.t. \ \ Q^T Q = I \end{aligned} \tag{8}$$

It equals to:

$$\begin{aligned} & \min_{Q} < Z_1, -Q > -\mu < S, Q > \\ & s.t. \ \ Q^T Q = I \end{aligned} \tag{9}$$

$\Longleftrightarrow$

$$\begin{aligned} & \max_{Q} < Q, \mu S + Z_1 > \\ & s.t. \ \ Q^T Q = I \end{aligned} \tag{10}$$

Use the proposition of matrix:

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

$$\| X \|_* := \max_{Y} \{ < X, Y > \ : \ \| Y \| \leq 1 \} \tag{11}$$

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

$$\begin{aligned} & \min_{J} < Z_2, S - J > + \tfrac{1}{2}\mu \| S - J \|_F^2 \\ & s.t. \ \ J \geq 0 \end{aligned} \tag{12}$$

$\Longleftrightarrow$

$$\begin{aligned} & \min_{J} \tfrac{2}{\mu} < Z_2, S - J > + \| S - J \|_F^2 + \tfrac{1}{\mu} \| Z_2 \|_F^2 \\ & s.t. \ \ J \geq 0 \end{aligned} \tag{13}$$

$\Longleftrightarrow$

$$\begin{aligned} &\min_{J} \parallel J - (S + \tfrac{Z_2}{\mu}) \parallel_F^2 \\ &s.t.\ \ J \geq 0 \end{aligned} \tag{14}$$

$\Longleftrightarrow$

$$J = \tfrac{1}{2} \cdot\ \ [abs(S + \tfrac{Z_2}{\mu}) + (S + \tfrac{Z_2}{\mu})] \tag{15}$$

Steps 6: Do some updates:

$$\begin{aligned} &Z_1 = Z_1 + \mu \cdot (S - Q) \\ &Z_2 = Z_2 + \mu \cdot (S - J) \\ &Z_3 = Z_3 + \mu(X \cdot S - K) \\ &\mu = \rho \times \mu \end{aligned} \tag{16}$$

where $\rho = 1.1$.

4 Termination Conditions

The algorithm will stop when the conditions are both satisfied:

$$\begin{aligned} &\parallel S - Q \parallel_\infty < \varepsilon \\ &\parallel S - J \parallel_\infty < \varepsilon \\ &\parallel X \cdot S - K \parallel_F < \varepsilon \times \parallel X \parallel_F \end{aligned} \tag{17}$$

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