1 Original Problem

The original problem is :

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

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 a select 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 :

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

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, $\mathbf{S} \in \mathbf{R}^{m\times k}$,$m \geq k$.

3 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:

$$
\begin{aligned}
\min_{Y,S,Q,J,Z_1,Z_2} \quad & \| X - ASY \|_F^2 + \alpha \| Y \|_F^2 + < Z_1, S - Q > \\
& + \tfrac{1}{2}\mu \| S - Q \|_F^2 + < Z_2, S - J > + \tfrac{1}{2}\mu \| S - J \|_F^2
\end{aligned}
\tag{3}
$$

$$
s.t.\ \ S = Q,\ S = J,\ Q^T Q = I,\ J \geq 0
$$

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

4 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:

$$
\min_{Y} \| X - ASY \|_F^2 + \alpha \| Y \|_F^2 \tag{4}
$$

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:

$$\begin{aligned} \min_S \quad & \| X - ASY \|_F^2 + < Z_1, S - Q > + \tfrac{1}{2}\mu \| S - Q \|_F^2 \\ & + < Z_2, S - J > + \tfrac{1}{2}\mu \| S - J \|_F^2 \end{aligned} \tag{5}$$

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

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

$$2A^T ASYY^T + 2\mu S = 2A^T XY^T + \mu(Q + J)$$

Use the property of $KroneckerProduct$.

$$Vec(C \cdot S \cdot B) = (B^T \bigotimes C) \cdot Vec(S)$$

We can get:

$$2(YY^T \bigotimes A^T A) \cdot Vec(S) + 2\mu Vec(S) = Vec(2A^T XY^T + \mu(Q + J))$$

$$2(YY^T \bigotimes A^T A + \mu I) \cdot Vec(S) = Vec(2A^T XY^T + \mu(Q + J))$$

$$Vec(S) = 2(YY^T \bigotimes A^T A + \mu I)^{-1} \cdot Vec(2A^T XY^T + \mu(Q + J))$$

Now We need to convert $Vector(S)$ 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:

$$\begin{aligned} & \min_Q < Z_1, S - Q > + \tfrac{1}{2}\mu \| S - Q \|_F^2 \\ & s.t. \;\; Q^T Q = I \end{aligned} \tag{6}$$

It equals to:

$$\begin{aligned} & \min_Q < Z_1, -Q > + \mu < S, Q > \\ & s.t. \;\; Q^T Q = I \end{aligned} \tag{7}$$

$\Longleftrightarrow$

$$\begin{aligned} & \max_Q < Q, \mu S + Z_1 > \\ & s.t. \;\; Q^T Q = I \end{aligned} \tag{8}$$

Use the proposition of matrix:

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

$$\| X \|_* := \max_Y \{< X, Y > \ : \ \| Y \| \leq 1\} \tag{9}$$

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

$$\begin{aligned} &\min_J < Z_2, S - J > + \tfrac{1}{2}\mu \| S - J \|_F^2 \\ &s.t. \ \ J \geq 0 \end{aligned} \tag{10}$$

$\Longleftrightarrow$

$$\begin{aligned} &\min_J \tfrac{2}{\mu} < Z_2, S - J > + \| S - J \|_F^2 + \tfrac{1}{\mu} \| Z_2 \|_F^2 \\ &s.t. \ \ J \geq 0 \end{aligned} \tag{11}$$

$\Longleftrightarrow$

$$\begin{aligned} &\min_J \| J - (S + \tfrac{Z_2}{\mu}) \|_F^2 \\ &s.t. \ \ J \geq I \end{aligned} \tag{12}$$

$\Longleftrightarrow$

$$J = \tfrac{1}{2} \cdot \ \ [abs(S + \tfrac{Z_2}{\mu}) + (S + \tfrac{Z_2}{\mu})] \tag{13}$$

Steps 5: Do some updates:

$$\begin{aligned} &Z_1 = Z_1 + \mu \cdot (S - Q) \\ &Z_2 = Z_2 + \mu \cdot (S - J) \\ &\mu = \rho \times \mu \end{aligned} \tag{14}$$

where $\rho = 1.1$.

5 Termination Conditions

The algorithm will stop when the conditions are both satisfied:

$$
\begin{aligned}
&\tfrac{1}{m \times k} \cdot \parallel S - J \parallel_F < \varepsilon \\
&\tfrac{1}{m \times k} \cdot \parallel S - Q \parallel_F < \varepsilon
\end{aligned}
\tag{15}
$$

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