SubKit: Subspace Clustering Library

Stephen Tierney, Yi Guo and Junbin Gao

August 19, 2014

Contents

1 Motivation

SubKit is a library with implementations for subspace clustering learning algorithms. This library was created during the writing of the OSC journal article to ensure fair comparison between different methods. In the original OSC CVPR paper we used implementations provided by authors of SSC and LRR. However these implementations used different noise models to what we suggested with OSC. For example the original LRR implementation expects column wise noise while we expect Gaussian noise in all columns. Therefore we required new implementations with the same noise model for fair comparison in our journal article.

2 Preliminaries

We introduce required notation. We assume the following noise model

$$\mathbf{X} = \mathbf{A} + \mathbf{N} \tag{1}$$

where $\mathbf{A}$ is the noise free data (column wise data samples), $\mathbf{N}$ is some noise and $\mathbf{X}$ is the observed data. We assume that each data sample in $\mathbf{A}$ lies exactly on its corresponding subspace. In an ideal case subspace clustering uses the self expressive model with the noise free data i.e.

$$\mathbf{A} = \mathbf{A}\mathbf{Z} \tag{2}$$

where $\mathbf{Z}$ is the matrix of coefficients. However this is rarely the case since we do not observe noise free data therefore the self expressive model becomes

$$\mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E} \tag{3}$$

where $\mathbf{E}$ is a fitting error. Once $\mathbf{Z}$ has been estimated one performs spectral clustering to obtain the final segmentation.

3 Function Listing

Objective	Function	Section
$\min_{\mathbf{Z}} \ \|\mathbf{Z}\|_1$ s.t. $\mathbf{A} = \mathbf{AZ}$	ssc_noisefree	4.1
$\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{XZ}\|_F^2 + \lambda\|\mathbf{Z}\|_1$ s.t. $\mathbf{X} = \mathbf{XZ} + \mathbf{E}$, $\mathrm{diag}(\mathbf{Z}) = \mathbf{0}$	ssc_relaxed ssc_relaxed_lin ssc_relaxed_lin_ext ssc_relaxed_lin_acc	4.2 4.3.1 4.3.2 4.3.3
$\min_{\mathbf{Z},\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_F^2 + \lambda\|\mathbf{Z}\|_1$ s.t. $\mathbf{X} = \mathbf{XZ} + \mathbf{E}$, $\mathrm{diag}(\mathbf{Z}) = \mathbf{0}$	ssc_exact_fro	4.4
$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{E}\|_1 + \lambda\|\mathbf{Z}\|_1$ s.t. $\mathbf{X} = \mathbf{XZ} + \mathbf{E}$, $\mathrm{diag}(\mathbf{Z}) = \mathbf{0}$	ssc_exact_l1	4.4
$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{E}\|_{1,2} + \lambda\|\mathbf{Z}\|_1$ s.t. $\mathbf{X} = \mathbf{XZ} + \mathbf{E}$, $\mathrm{diag}(\mathbf{Z}) = \mathbf{0}$	ssc_exact_l1l2	4.4
$\min_{\mathbf{Z}} \ \|\mathbf{Z}\|_*$ s.t. $\mathbf{A} = \mathbf{AZ}$	lrr_noisefree	5.1
$\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{XZ}\|_F^2 + \lambda\|\mathbf{Z}\|_*$	lrr_relaxed lrr_relaxed_lin lrr_relaxed_lin_ext lrr_relaxed_lin_acc	5.2 5.3 5.3 5.3
$\min_{\mathbf{Z},\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_F^2 + \lambda\|\mathbf{Z}\|_*$ s.t. $\mathbf{X} = \mathbf{XZ} + \mathbf{E}$	lrr_exact_fro	5.4
$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{E}\|_1 + \lambda\|\mathbf{Z}\|_*$ s.t. $\mathbf{X} = \mathbf{XZ} + \mathbf{E}$	lrr_exact_l1	5.4
$\min_{\mathbf{Z},\mathbf{E}} \|\mathbf{E}\|_{1,2} + \lambda\|\mathbf{Z}\|_*$ s.t. $\mathbf{X} = \mathbf{XZ} + \mathbf{E}$	lrr_exact_l1l2	5.4
$\min_{\mathbf{Z},\mathbf{A},\mathbf{N}} \ \lambda\|\mathbf{N}\|_q + \|\mathbf{Z}\|_*$ s.t. $\mathbf{A} = \mathbf{AZ}, \mathbf{X} = \mathbf{A} + \mathbf{N}$	r_lrr_fro r_lrr_l1 r_lrr_l2	5.5 5.5 5.5
$\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{XZ}|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{ZR}\|_1$ s.t. $\mathrm{diag}(\mathbf{Z}) = \mathbf{0}$	spatsc_noisefree	6.1
$\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{XZ}|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{ZR}\|_{1,2}$ s.t. $\mathrm{diag}(\mathbf{Z}) = \mathbf{0}$	osc_noisefree	7.1

4 Sparse Subspace Clustering

4.1 Noise Free

$$\begin{aligned} &\min_{\mathbf{Z}} \lambda \|\mathbf{Z}\|_1 \\ \text{s.t.} \quad &\mathbf{A} = \mathbf{A}\mathbf{Z} \end{aligned} \tag{4}$$

The solution to the above objective has a fast approximate solution, given by the Shape Interaction Matrix. See the low-rank clustering section for further details. Alternatively one can calculate the sample correlation matrix $\mathbf{X}^T\mathbf{X}$ and select points with the maximum correlation as neighbours on the graph.

4.2 Relaxed SSC ADMM

$$\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_1 \tag{5}$$

To solve the relaxation via ADMM one must incorporate an auxiliary variable

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{J}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{J}\|_F^2 + \lambda\|\mathbf{Z}\|_1 \\ &\text{s.t.} \quad \mathbf{Z} = \mathbf{J} \end{aligned} \tag{6}$$

$$\min_{\mathbf{Z},\mathbf{J}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{J}\|_F^2 + \lambda\|\mathbf{Z}\|_1 + \langle \mathbf{Y}, \mathbf{Z} - \mathbf{J}\rangle + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{J}\|_F^2 \tag{7}$$

Then iterate the following

1. Fix others and solve for $\mathbf{J}$

$$\min_{\mathbf{J}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{J}\|_F^2 - \langle \mathbf{Y}, \mathbf{J}\rangle + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{J}\|_F^2$$

$$\min_{\mathbf{J}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{J}\|_F^2 + \frac{\mu}{2}\|(\mathbf{Z} + \frac{1}{\mu}\mathbf{Y}) - \mathbf{J}\|_F^2$$

$$\mathbf{X}^T(\mathbf{X} - \mathbf{X}\mathbf{J}) + \mu((\mathbf{Z} + \frac{1}{\mu}\mathbf{Y}) - \mathbf{J})$$

$$\mathbf{X}^T\mathbf{X}\mathbf{J} + \mu\mathbf{J} = \mathbf{X}^T\mathbf{X} + \mu\mathbf{Z} + \mathbf{Y}$$

$$(\mathbf{X}^T\mathbf{X} + \mu\mathbf{I})\mathbf{J} = \mathbf{X}^T\mathbf{X} + \mu\mathbf{Z} + \mathbf{Y}$$

$$\mathbf{J} = (\mathbf{X}^T\mathbf{X} + \mu\mathbf{I})^{-1}(\mathbf{X}^T\mathbf{X} + \mu\mathbf{Z} + \mathbf{Y})$$

2. Fix others and solve for $\mathbf{Z}$

$$\min_{\mathbf{Z}} \lambda\|\mathbf{Z}\|_1 + \langle \mathbf{Y}, \mathbf{Z}\rangle + \frac{\mu}{2}\|\mathbf{Z} - \mathbf{J}\|_F^2$$

$$\min_{\mathbf{Z}} \lambda\|\mathbf{Z}\|_1 + \frac{\mu}{2}\|\mathbf{Z} - (\mathbf{J} - \frac{1}{\mu}\mathbf{Y})\|_F^2$$

3. Update $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{Y} + \mu(\mathbf{Z} - \mathbf{J})$$

4.3 Relaxed SSC Linearised

Instead of using ADMM one can linearise the Frobenius norm term instead. This removes the need to do an expensive matrix inversion.

$$\min_{\mathbf{Z}} L = \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}\|_F^2 + \lambda\|\mathbf{Z}\|_1 \tag{8}$$

Let $F = \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}\|_F^2$ and $\partial F = -\mathbf{X}^T\mathbf{X} + \mathbf{X}^T\mathbf{X}\mathbf{Z}$.

4.3.1 Gradient Descent

This version consists on iterating the following until convergence

$$\min_{\mathbf{Z}} \widetilde{L}_\rho(\mathbf{Z}, \mathbf{Z_k}) = \lambda\|\mathbf{Z}\|_1 + \frac{\rho}{2}\|\mathbf{Z} - (\mathbf{Z}_{k-1} - \frac{1}{\rho}\partial F(\mathbf{Z}_{k-1}))\|_F^2$$

4.3.2 Extended Gradient Algorithm

It has been shown that through the correct choice of step size ρ the gradient method can achieve a convergence rate of $\mathrm{O}\left(\frac{1}{k}\right)$ [**?**]. More specifically we require that $\rho \geq C$ where C is the Lipschitz constant. However we do not know the value of C in advance. Fortunately we know that if $L(\mathbf{B}) \leq \widetilde{L}_\rho(\mathbf{B}, \mathbf{Z}_{k-1})$ where $\mathbf{B} = \mathcal{S}_{\frac{\lambda}{\rho}}(\mathbf{Z}_{k-1} - \frac{1}{\lambda}\partial L(\mathbf{Z}_{k-1}))$ then $\rho \geq C$.

Algorithm 1 Extended Gradient Descent for Robust MC

Require: $k = 1$, $r_0 = \infty$, $\mathbf{Z}_0 = \mathbf{0}$, λ, ρ, γ, ϵ
- **while** $r_k - r_{k-1} \geq \epsilon$ **do**
  - **while** $L(\mathbf{B}) \geq \widetilde{L}_\rho(\mathbf{B}, \mathbf{Z}_{k-1})$ **do**
    - $\rho = \gamma\rho$
  - **end while**
  - $\mathbf{Z}_k = \mathcal{S}_{\frac{\lambda}{\rho}}(\mathbf{Z}_{k-1} - \frac{1}{\rho}\partial F(\mathbf{Z}_{k-1}))$
  - $r_k = \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}_k\|_F^2 + \lambda\|\mathbf{Z}_k\|_1$
  - $k = k + 1$
- **end while**

4.3.3 Accelerated Gradient Algorithm

We can further improve our convergence rate to $\mathrm{O}\left(\frac{1}{k^2}\right)$ by adopting Nesterov's accelerated gradient algorithm. This is similar to the algorithm described in [**?**]. Let $\mathbf{B} = \mathcal{S}_{\frac{\lambda}{\rho}}(\mathbf{J}_{k-1} - \frac{1}{\rho}\partial L(\mathbf{J}_{k-1}))$.

Algorithm 2 Accelerated Gradient Descent for Robust MC

Require: $k = 1$, $r_0 = \infty$, $\mathbf{Z}_0 = \mathbf{0}$, $\mathbf{J}_0 = \mathbf{0}$, $\alpha_0 = 1$, λ, ρ, γ, ϵ
  while $r_k - r_{k-1} \geq \epsilon$ **do**
    while $L(\mathbf{B}) \geq \widetilde{L}_\rho(\mathbf{B}, \mathbf{J}_{k-1})$ **do**
      $\rho = \gamma\rho$
    end while
    $\mathbf{Z}_k = \mathcal{S}_{\frac{\lambda}{\rho}}(\mathbf{J}_{k-1} - \frac{1}{\rho}\partial F(\mathbf{J}_{k-1}))$
    $\alpha_k = \frac{1+\sqrt{1+4\alpha_{k-1}^2)}}{2}$
    $\mathbf{J}_k = \mathbf{Z}_k + \left(\frac{\alpha_{k-1}-1}{\alpha_k}\right)(\mathbf{Z}_k - \mathbf{Z}_{k-1})$
    $r_k = \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}_k\|_F^2 + \lambda\|\mathbf{Z}_k\|_1$
    $k = k + 1$
  end while

4.4 Exact SSC LADM

Using LADM and exact constraints allows us to modify the noise term from Gaussian noise only to any other norm. The objective becomes

$$\begin{aligned} \min_{\mathbf{E},\mathbf{Z}} \frac{1}{2}\|\mathbf{E}\|_q^2 + \lambda\|\mathbf{Z}\|_1 \\ \text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E} \end{aligned} \tag{9}$$

where q is a placeholder for norms such as $\ell_1, F, \ell_{1,2}$ etc.

$$\min_{\mathbf{E},\mathbf{Z}} \frac{1}{2}\|\mathbf{E}\|_q^2 + \lambda\|\mathbf{Z}\|_1 + \langle \mathbf{Y}, \mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\rangle + \frac{\mu}{2}\|\mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\|_F^2$$

Iterate the following

1. Fix others solve for $\mathbf{Z}$

$$\min_{\mathbf{Z}} \lambda\|\mathbf{Z}\|_1 + \langle \mathbf{Y}, \mathbf{X}\mathbf{Z}\rangle + \frac{\mu}{2}\|\mathbf{X}\mathbf{Z} - (\mathbf{X} - \mathbf{E})\|_F^2$$

$$\min_{\mathbf{Z}} \lambda\|\mathbf{Z}\|_1 + \frac{\mu}{2}\|\mathbf{X}\mathbf{Z} - (\mathbf{X} - \mathbf{E} - \frac{1}{\mu}\mathbf{Y})\|_F^2$$

Let $F = \frac{\mu}{2}\|\mathbf{X}\mathbf{Z} - (\mathbf{X} - \mathbf{E} - \frac{1}{\mu}\mathbf{Y})\|_F^2$ and $\partial F = \mu\mathbf{X}^T(\mathbf{X}\mathbf{Z} - (\mathbf{X} - \mathbf{E} - \frac{1}{\mu}\mathbf{Y}))$

$$\min_{\mathbf{Z}} \lambda\|\mathbf{Z}\|_1 + \frac{\rho}{2}\|\mathbf{Z} - (\mathbf{Z}_k - \frac{1}{\rho}\partial F(\mathbf{Z}_k)\|_F^2$$

2. Fix others and solve for $\mathbf{E}$

$$\min_{\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_q^2 + \langle \mathbf{Y}, \mathbf{E}\rangle + \frac{\mu}{2}\|\mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\|_F^2$$

$$\min_{\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_q^2 + \langle \mathbf{Y}, \mathbf{E}\rangle + \frac{\mu}{2}\|\mathbf{E} - (\mathbf{X} - \mathbf{X}\mathbf{Z})\|_F^2$$

$$\min_{\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_q^2 + \frac{\mu}{2}\|\mathbf{E} - (\mathbf{X} - \mathbf{X}\mathbf{Z} - \frac{1}{\mu}\mathbf{Y})\|_F^2$$

3. Update $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{Y} + \mu(\mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E})$$

4. Update μ

$$\mu = \min(\mu_{\max}, \gamma\mu)$$

where γ is defined as

$$\gamma = \begin{cases} \gamma_0 & \text{if } \mu_k \frac{\max(\sqrt{\rho}\|\mathbf{Z}_{k+1} - \mathbf{Z}_k\|_F, \|\mathbf{E}_{k+1} - \mathbf{E}_k\|_F)}{\|\mathbf{X}\|_F} < \epsilon \\ 1 & \text{otherwise,} \end{cases}$$

and $\rho > \|\mathbf{X}\|_F^2$, $\mu_{\max} >> \mu_0$ and $\epsilon > 0$.

4.5 Diagonal Constraint

In some cases it may be desirable to enforce the constraint $\text{diag}(\mathbf{Z}) = \mathbf{0}$ i.e. we should not allow each data point to be represented by itself. To enforce such a constraint it is not necessary to significantly alter the aforementioned optimisation schemes. This constraint only affects the step involving $\mathbf{Z}$. Since this step is the soft shrinkage operator and is separable at the element level one can simply set the diagonal entries to 0 afterwards.

5 Low-Rank Subspace Clustering

5.1 Noise Free

$$\begin{aligned} &\min_{\mathbf{Z}} \ \|\mathbf{Z}\|_* \\ \text{s.t.} \quad &\mathbf{A} = \mathbf{A}\mathbf{Z} \end{aligned} \tag{10}$$

The solution to the above objective is given by the Shape Interaction Matrix (SIM) which is defined as $\mathbf{V}\mathbf{V}^{\mathbf{T}}$ where $\mathbf{V}$ is the right singular vectors of $\mathbf{A}$ i.e.

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}, \ \ \mathbf{\Sigma} = \mathrm{diag}(\{\sigma_i\}_{i=1}^r)$$

5.2 Relaxed LRR ADMM

$$\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{X}\mathbf{Z}\|_F^2 + \tau\|\mathbf{Z}\|_* \tag{11}$$

This is solved similarly to Relaxed SSC ADMM. Instead of the ℓ_1 shrinking operator in the update step for $\mathbf{Z}$ we must use the singular value shrinking operator which is defined as

$$\mathcal{D}_\tau(\mathbf{Y}) = \mathbf{U}S_\tau(\mathbf{\Sigma})\mathbf{V}^T, \ \ S_\tau(\mathbf{\Sigma}) = \mathrm{diag}(\{\max(\sigma_i - \tau, 0)\}). \tag{12}$$

5.3 Relaxed LRR Linearised

This is solved similarly to Relaxed SSC ADMM, again replacing the ℓ_1 with the singular value shrinking operator.

5.4 Exact LRR LADM

This is solved similarly to Exact SSC LADM, again replacing the ℓ_1 with the singular value shrinking operator.

5.5 Robust LRR

An extension of LRR called Robust-LRR (R-LRR) aims to directly deal with noisy data using the previously described data generation model

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{A},\mathbf{N}} \ \lambda\|\mathbf{N}\|_q + \|\mathbf{Z}\|_* \\ \text{s.t.} \quad &\mathbf{A} = \mathbf{A}\mathbf{Z}, \mathbf{X} = \mathbf{A} + \mathbf{N} \end{aligned} \tag{13}$$

This objective can be decomposed into two steps. The first step solves a variant of RPCA [CITE] and the second uses the estimated $\mathbf{A}$ to compute $\mathbf{Z}$ using the SIM. However it is unclear as to whether such a procedure will be capable of exactly removing all noise, thus the SIM may give poor results. Fortunately further work in nuclear norm based subspace clustering by Vidal and Favaro [**?**] demonstrated a number of other closed form solutions for noisy data.

6 Spatial Subspace Clustering

In [**?**] the following spatial subspace clustering (SpatSC) objective was proposed

$$\begin{aligned} &\min_{\mathbf{Z},\mathbf{E}} \frac{1}{2}\|\mathbf{E}|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{Z}\mathbf{R}\|_1 \\ &\text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E},\ \mathrm{diag}(\mathbf{Z}) = \mathbf{0} \end{aligned} \tag{14}$$

6.1 Implementation

SpatSC can be implemented similarly to OSC. However the $\ell_{1,2}$ shrinkage operator is replaced with ℓ_1 shrinkage operator. Please see the following section for details.

7 Ordered Subspace Clustering

$$\min_{\mathbf{Z},\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{ZR}\|_{1,2} \quad (15)$$
$$\text{s.t.} \quad \mathbf{X} = \mathbf{XZ} + \mathbf{E}$$

7.1 Relaxed Constraints

First we remove the variable $\mathbf{E}$ by using the constraint and thus the objective (15) can be re-written as follows,

$$\min_{\mathbf{Z}} \frac{1}{2}\|\mathbf{X} - \mathbf{XS}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{J}\|_{1,2} \quad (16)$$
$$\text{s.t.} \quad \mathbf{S} = \mathbf{Z}, \mathbf{J} = \mathbf{SR}$$

Then the Augmented Lagrangian for the two introduced constraints is

$$\begin{aligned}\mathcal{L}(\mathbf{Z},\mathbf{S},\mathbf{U}) =& \frac{1}{2}\|\mathbf{X} - \mathbf{XS}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{U}\|_{1,2} \\ &+ \langle \mathbf{Y}_1, \mathbf{Z} - \mathbf{S}\rangle + \frac{\mu^k}{2}\|\mathbf{Z} - \mathbf{S}\|_F^2 \\ &+ \langle \mathbf{Y}_2, \mathbf{J} - \mathbf{SR}\rangle + \frac{\mu^k}{2}\|\mathbf{J} - \mathbf{SR}\|_F^2 \end{aligned} \quad (17)$$

We can solve (17) for $\mathbf{Z}, \mathbf{S}, \mathbf{U}$ in an alternative manner when fixing the others, respectively.

1. Set $\mathbf{S} = \mathbf{S}^k$ and $\mathbf{J} = \mathbf{J}^k$, solve for $\mathbf{Z}^{k+1}$ by

$$\min_{\mathbf{Z}} \lambda_1\|\mathbf{Z}\|_1 + \langle \mathbf{Y}_1^k, \mathbf{Z} - \mathbf{S}\rangle + \frac{\mu^k}{2}\|\mathbf{Z} - \mathbf{S}\|_F^2$$

which is equivalent to

$$\min_{\mathbf{Z}} \lambda_1\|\mathbf{Z}\|_1 + \frac{\mu^k}{2}||\mathbf{Z} - (\mathbf{S} - \frac{\mathbf{Y}_1^k}{\mu^k})||_F^2 \quad (18)$$

2. Set $\mathbf{Z} = \mathbf{Z}^k$ and $\mathbf{J} = \mathbf{J}^k$, solve for $\mathbf{S}^{k+1}$ by

$$\begin{aligned}\min_{\mu^k S} \frac{1}{2}\|\mathbf{X} - \mathbf{XS}\|_F^2 + \langle \mathbf{Y}_1^k, \mathbf{Z} - \mathbf{S}\rangle + \frac{\mu^k}{2}\|\mathbf{Z} - \mathbf{S}\|_F^2 \\ + \langle \mathbf{Y}_2^k, \mathbf{J} - \mathbf{SR}\rangle + \frac{\mu^k}{2}\|\mathbf{J} - \mathbf{SR}\|_F^2\end{aligned}$$

Setting the derivative of the above objective with respect to $\mathbf{S}$ to zero gives

$$\begin{aligned}\mathbf{X}^T(\mathbf{XS} - \mathbf{X}) + \mu^k(\mathbf{S} - \mathbf{Z}) + \mu^k(\mathbf{SR} - \mathbf{J})\mathbf{R}^T \\ -\mathbf{Y}_1^k - \mathbf{Y}_2^k\mathbf{R}^T = 0\end{aligned} \quad (19)$$

hence the solution is

$$\begin{aligned}&(\mathbf{X}^T\mathbf{X}+\mu^k I)\mathbf{S}+\mu^k\mathbf{S}\mathbf{R}\mathbf{R}^T\\=&\mathbf{X}^T\mathbf{X}+\mu^k\mathbf{J}\mathbf{R}^T+\mu^k\mathbf{Z}+\mathbf{Y}_1^k+\mathbf{Y}_2^k\mathbf{R}^T\end{aligned}\tag{20}$$

We can vectorize the above linear matrix equation into

$$\begin{aligned}&[I\otimes(\mathbf{X}^T\mathbf{X}+\mu^k I)+\mu^k\mathbf{R}\mathbf{R}^T\otimes I]\mathrm{vec}(\mathbf{S})\\=&\ \mathrm{vec}(\mathbf{X}^T\mathbf{X}+\mu^k\mathbf{J}\mathbf{R}^T+\mu^k\mathbf{Z}+\mathbf{Y}_1^k+\mathbf{Y}_2^k\mathbf{R}^T)\end{aligned}$$

where $\otimes$ is the tensor product. However this solution is infeasible for large matrices in terms of memory and computation time. Fortunately (20) is a Sylvester equation of the form

$$\mathbf{A}\mathbf{S}+\mathbf{S}\mathbf{B}=\mathbf{C}$$

and has a number of efficient solutions.

3. Set $\mathbf{Z}=\mathbf{Z}^k$ and $\mathbf{S}=\mathbf{S}^k$, solve for $\mathbf{J}^{k+1}$ by

$$\min_{\mathbf{J}}\lambda_2\|\mathbf{J}\|_{1,2}+\langle\mathbf{Y}_2^k,\mathbf{J}-\mathbf{S}\mathbf{R}\rangle+\frac{\mu^k}{2}\|\mathbf{J}-\mathbf{S}\mathbf{R}\|_F^2$$

which is equivalent to

$$\min_{\mathbf{J}}\lambda_2\|\mathbf{J}\|_{1,2}+\frac{\mu^k}{2}||\mathbf{J}-(\mathbf{S}\mathbf{R}-\frac{1}{\mu^k}\mathbf{Y}_2^k)||_F^2$$

4. Update $\mathbf{Y}_1$ and $\mathbf{Y}_2$ by

$$\begin{aligned}\mathbf{Y}_1^{k+1}&=\mathbf{Y}_1^k+\mu^k(\mathbf{Z}-\mathbf{S})\\\mathbf{Y}_2^{k+1}&=\mathbf{Y}_2^k+\mu^k(\mathbf{J}-\mathbf{S}\mathbf{R})\end{aligned}$$

5. Update μ

$$\mu^{k+1}=\min(\mu_{\max_1},\gamma\mu^k)$$

where γ is defined as

$$\gamma=\begin{cases}\gamma^0 & \text{if } \ \mu^k\frac{\max(\|\mathbf{Z}^{k+1}-\mathbf{Z}^k\|_F,\|\mathbf{S}^{k+1}-\mathbf{S}^k\|,\|\mathbf{J}^{k+1}-\mathbf{J}^k\|_F,\|\mathbf{S}^{k+1}\mathbf{R}-\mathbf{S}^k\mathbf{R}\|)}{\|\mathbf{X}\|}<\epsilon_2\\1 & \text{otherwise,}\end{cases}$$

and $\mu^{\max}>>\mu^0$ and $\epsilon>0$.

6. Check stopping criteria

$$\frac{\|\mathbf{Z}^{k+1}-\mathbf{S}^{k+1}\|_F}{\|\mathbf{X}\|}<\epsilon_1,\frac{\|\mathbf{J}^{k+1}-\mathbf{S}^{k+1}\mathbf{R}\|_F}{\|\mathbf{X}\|}<\epsilon_1,$$

$$\mu^k\frac{\max(\|\mathbf{Z}^{k+1}-\mathbf{Z}^k\|_F,\|\mathbf{S}^{k+1}-\mathbf{S}^k\|,\|\mathbf{J}^{k+1}-\mathbf{J}^k\|_F,\|\mathbf{S}^{k+1}\mathbf{R}-\mathbf{S}^k\mathbf{R}\|)}{\|\mathbf{X}\|}<\epsilon_2$$

7.2 Exact Constraints

Similar to the relaxed version we begin by introducing auxiliary variables

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{E},\mathbf{J}} \frac{1}{2}\|\mathbf{E}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{J}\|_{1,2} \\ \text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \mathbf{J} = \mathbf{Z}\mathbf{R} \end{aligned} \tag{21}$$

We then form the Augmented Lagrangian to incorporate our constraints

$$\begin{aligned} \mathcal{L}(\mathbf{E},\mathbf{Z},\mathbf{J}) =& \frac{1}{2}\|\mathbf{E}\|_F^2 + \lambda_1\|\mathbf{Z}\|_1 + \lambda_2\|\mathbf{J}\|_{1,2} \\ &+ \langle \mathbf{Y}_1, \mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\rangle + \frac{\mu}{2}\|\mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\|_F^2 \\ &+ \langle \mathbf{Y}_2, \mathbf{J} - \mathbf{Z}\mathbf{R}\rangle + \frac{\mu}{2}\|\mathbf{J} - \mathbf{Z}\mathbf{R}\|_F^2 \end{aligned} \tag{22}$$

Then we solve for $\mathbf{E}, \mathbf{Z}, \mathbf{J}$ alternatively while fixing others in the following way

1. Set $\mathbf{E} = \mathbf{E}^k$ and $\mathbf{J} = \mathbf{J}^k$, solve for $\mathbf{Z}^{k+1}$ by

$$\begin{aligned} \min_{\mathbf{Z}} \lambda_1\|\mathbf{Z}\|_1 + \langle \mathbf{Y}_1^k, \mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\rangle + \frac{\mu^k}{2}\|\mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\|_F^2 \\ + \langle \mathbf{Y}_2^k, \mathbf{J} - \mathbf{Z}\mathbf{R}\rangle + \frac{\mu^k}{2}\|\mathbf{J} - \mathbf{Z}\mathbf{R}\|_F^2 \end{aligned}$$

$$\begin{aligned} \min_{\mathbf{Z}} \lambda_1\|\mathbf{Z}\|_1 + \frac{\mu^k}{2}\|\mathbf{X}\mathbf{Z} - (\mathbf{X} - \mathbf{E} - \frac{1}{\mu^k}\mathbf{Y}_1^k)\|_F^2 \\ + \frac{\mu^k}{2}\|\mathbf{Z}\mathbf{R} - (\mathbf{J} - \frac{1}{\mu^k}\mathbf{Y}_2^k)\|_F^2 \end{aligned}$$

   We linearise the last two terms which we call F

$$\min_{\mathbf{Z}} \lambda_1\|\mathbf{Z}\|_1 + \frac{\rho}{2}\|\mathbf{Z} - (\mathbf{Z}_{k-1} - \frac{1}{\rho}\partial F)\|_F^2$$

   and $\partial F = \mu^k \mathbf{X}^T(\mathbf{X}\mathbf{Z}_{\mathbf{k-1}} - (\mathbf{X} - \mathbf{E} - \frac{1}{\mu^k}\mathbf{Y}_1^k)) + \mu^k(\mathbf{Z}_{\mathbf{k-1}}\mathbf{R} - (\mathbf{J} - \frac{1}{\mu^k}\mathbf{Y}_2^k))\mathbf{R}^T$.

2. Set $\mathbf{Z} = \mathbf{Z}^k$ and $\mathbf{J} = \mathbf{J}^k$ solve for $\mathbf{E}^{k+1}$ by

$$\begin{aligned} \min_{\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_F^2 + \langle \mathbf{Y}_1^k, \mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\rangle \\ + \frac{\mu^k}{2}\|\mathbf{X}\mathbf{Z} - \mathbf{X} + \mathbf{E}\|_F^2 \end{aligned}$$

$$\min_{\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_F^2 + \frac{\mu^k}{2}\|\mathbf{E} - (\mathbf{X}\mathbf{Z} - \mathbf{X} + \frac{1}{\mu^k}\mathbf{Y}_1^k)\|_F^2$$

3. Set $\mathbf{Z} = \mathbf{Z}^k$ and $\mathbf{E} = \mathbf{E}^k$ solve for $\mathbf{J}^{k+1}$ by

$$\min_{\mathbf{J}} \lambda_2 \|\mathbf{J}\|_{1,2} + \langle \mathbf{Y}_2^k, \mathbf{J} - \mathbf{ZR} \rangle + \frac{\mu^k}{2} \|\mathbf{J} - \mathbf{ZR}\|_F^2$$

$$\min_{\mathbf{J}} \lambda_2 \|\mathbf{J}\|_{1,2} + \frac{\mu^k}{2} \|\mathbf{J} - (\mathbf{ZR} - \frac{1}{\mu^k} \mathbf{Y}_2^k)\|_F^2$$

4. Update $\mathbf{Y}_1$ and $\mathbf{Y}_2$

$$\begin{aligned} \mathbf{Y}_1^{k+1} =& \mathbf{Y}_1^k + \mu^k (\mathbf{XZ} - \mathbf{X} + \mathbf{E}) \\ \mathbf{Y}_2^{k+1} =& \mathbf{Y}_2^k + \mu^k (\mathbf{J} - \mathbf{ZR}) \end{aligned}$$

5. Update μ

$$\mu^{k+1} = \min(\mu_{\max_1}, \gamma \mu^k)$$

where γ is defined as

$$\gamma_1 = \begin{cases} \gamma^0 & \text{if } \mu^k \sqrt{\rho} \frac{\max(\|\mathbf{Z}^{k+1} - \mathbf{Z}^k\|_F, \|\mathbf{E}^{k+1} - \mathbf{E}^k\|, \|\mathbf{J}^{k+1} - \mathbf{J}^k\|_F, \|\mathbf{Z}^{k+1}\mathbf{R} - \mathbf{Z}^k\mathbf{R}\|_F)}{\|\mathbf{X}\|_F} < \epsilon_2 \\ 1 & \text{otherwise,} \end{cases}$$

where $\rho > \|\mathbf{X}\|_F^2$, $\mu^{\max} >> \mu^0$ and $\epsilon > 0$.

6. Check stopping criteria

$$\frac{\|\mathbf{XZ^{k+1}} - \mathbf{X} + \mathbf{E^{k+1}}\|_F}{\|\mathbf{X}\|_F} < \epsilon_1, \frac{\|\mathbf{J^{k+1}} - \mathbf{Z^{k+1}R}\|_F}{\|\mathbf{X}\|_F} < \epsilon_1$$

$$\mu^k \sqrt{\rho} \frac{\max(\|\mathbf{Z}^{k+1} - \mathbf{Z}^k\|_F, \|\mathbf{E}^{k+1} - \mathbf{E}^k\|, \|\mathbf{J}^{k+1} - \mathbf{J}^k\|_F, \|\mathbf{Z}^{k+1}\mathbf{R} - \mathbf{Z}^k\mathbf{R}\|_F)}{\|\mathbf{X}\|_F} < \epsilon_2$$