1. Minimization Over $B$

Given the current fixed $A^{(k)}$ and $\Lambda^{(k)}$, $B$ can be updated by minimizing the augmented Lagrangian:

$$
B^{(k+1)} := \arg \min_B L_\rho(A^{(k)}, B, \Lambda^{(k)})
:= \arg \min_B \ell(B) + \frac{\gamma}{q} \| B \|_{p,q}^{q}
$$

(10)

where $\ell(B)$ is the smooth part of function is

$$
\ell(B) = \frac{1}{2} \| A^{(k)\top} C_s B - C_t \|_F^2 + \frac{\alpha}{2} \| X_0^0 B - X_1^0 \|_F^2 - \text{tr}(\Lambda^{(k)\top} B) + \frac{\rho}{2} \| B \|_F^2
$$

This minimization problem is a convex quadratic programing with a non-smooth sparsity regularizer. We solve it using a fast proximal gradient descent method with a quadratic convergence rate (Beck and Teboulle, 2009), which tackles Eq.(10) by solving a sequence of intermediate problems iteratively with proximity operators. The algorithm is given in Algorithm 1 below. The convergence of the algorithm is proved in (Beck and Teboulle, 2009).

**Algorithm 1 Fast Proximal Gradient Descent Algorithm**

| Initialization: $Q^{(1)} = B^{(0)}$=starting point, $\beta_1 = 1$, $t = 0$. |
| For iter = 1:maxiters |
| 1. Set $t = t + 1$ |
| 2. Update: $B^{(t)} = P_\eta(Q^{(t)})$, $\beta_{t+1} = \frac{1 + \sqrt{1 + 4 \beta_t^2}}{2},$ |
| $Q^{(t+1)} = B^{(t)} + \left( \frac{\beta_t - 1}{\beta_{t+1}} \right) (B^{(t)} - B^{(t-1)})$ |
| End For |

For the $t$-th iteration, the intermediate problem at point $Q^{(t)}$ is in the following form:

$$
P_\eta(Q^{(t)}) = \arg \min_B \left\{ \frac{1}{2} \| B - \hat{Q}^{(t)} \| + \frac{\gamma}{q\eta} \| B \|_{p,q}^{q} \right\}
$$

(11)

where $\hat{Q}^{(t)}$ is derived from the gradient of $\ell(Q^{(t)})$ such that

$$
\hat{Q}^{(t)} = Q^{(t)} - \frac{1}{\eta} \nabla \ell(Q^{(t)})
$$
and $\eta$ is the Lipschitz constant of the general gradient function $\nabla \ell(B)$. The gradient can be computed as

$$
\nabla \ell(B) = \left( C_s^\top A^{(k)} A^{(k)\top} C_s + \alpha X_s^{0\top} X_s^0 + \rho I \right) B - \left( C_s^\top A^{(k)} C_t + \alpha X_s^{0\top} X_t^0 + \Lambda^{(k)} + \rho A^{(k)} \right)
$$

A Lipschitz constant $\eta$ of $\nabla \ell(B)$ needs to satisfy the property

$$
\| \nabla \ell(B) - \nabla \ell(B') \|_F \leq \eta \| B - B' \|_F, \text{ for any feasible } B, B'.
$$

**Lemma 1** Let

$$
\eta = \sigma_{\text{max}} \left( C_s^\top A^{(k)} A^{(k)\top} C_s + \alpha X_s^{0\top} X_s^0 + \rho I \right),
$$

where $\sigma_{\text{max}}(\cdot)$ denotes the largest singular value of the corresponding matrix. Then $\eta$ is a Lipschitz constant of $\nabla \ell(B)$.

**Proof** Let $H = C_s^\top A^{(k)} A^{(k)\top} C_s + \alpha X_s^{0\top} X_s^0 + \rho I$. We have the following derivations

$$
\| \nabla \ell(B) - \nabla \ell(B') \|_F
= \left\| (C_s^\top A^{(k)} A^{(k)\top} C_s + \alpha X_s^{0\top} X_s^0 + \rho I)(B - B') \right\|_F
= \| H(B - B') \|_F
= \left( \sum_j \| H(B:j - B'_j) \|_2^2 \right)^{1/2}
\leq \left( \| H \|_2^2 \sum_j \| B:j - B'_j \|_2^2 \right)^{1/2}
= \| H \|_F \| B - B' \|_F
= \sigma_{\text{max}}(H) \| B - B' \|_F
$$

where $\| \cdot \|_2$ denotes the spectral norm of the corresponding matrix or the Euclidean norm of a vector; $B:j$ denotes the $j$-th column of matrix $B$.

The nice property about the intermediate problem in Eq.(11) is that it allows us to exploit closed-form solutions for the proximity operator $P_\eta(Q^{(t)})$ with either the $\ell_1$-norm regularizer ($p = 1$ and $q = 1$) or the $\ell_1,2$-norm regularizer ($p = 1$ and $q = 2$). According to (Kowalski et al., 2009), we have the following closed-form solution for the proximity operations:

If $p = 1$ and $q = 1$ ($\ell_1$-norm), we have

$$
P_\eta(Q^{(t)}) = \text{sign}(\widehat{Q}^{(t)}) \circ \left( |\widehat{Q}^{(t)}| - \frac{\gamma}{\eta} \right)_+
$$

where $(\cdot)_+ = \max(0, \cdot)$ and $\circ$ denotes the entrywise Hadamard product operator.
If \( p = 1 \) and \( q = 2 \) (\( \ell_{1,2}\)-norm), we have

\[
P_{\eta}(Q^{(t)}) = \tilde{Q}
\]

such that

\[
\tilde{Q}_{i,j} = \text{sign}(\hat{Q}_{i,j}^{(t)}) \left( |\hat{Q}_{i,j}^{(t)}| - \frac{\gamma \sum_{r=1}^{m_j} \tilde{Q}_{r,j}}{(\eta + \gamma m_j)\|\hat{Q}_{j}^{(t)}\|_2} \right)_+
\]

where \( \tilde{Q}_{j} \) denotes a reordered \( j \)-th column \( |\tilde{Q}_{j}^{(t)}| \) with a descending order of the entries, and the corresponding \( m_j \) is the number such that

\[
\tilde{Q}_{m_j+1,j} \leq \frac{\gamma}{\eta} \sum_{r=1}^{m_j} (\tilde{Q}_{r,j} - \tilde{Q}_{m_j+1,j})
\]

\[
\tilde{Q}_{m_j,j} > \frac{\gamma}{\eta} \sum_{r=1}^{m_j} (\tilde{Q}_{r,j} - \tilde{Q}_{m_j,j})
\]

References
