Notes for High-dimensional vector error-correction models

Sumanta Basu, Stephan Smeekes, Etienne Wijler

October 8, 2025

1 ADMM algorithm

Let $\boldsymbol{y}_t = (y_{1,t}, \dots, y_{N,t})'$ denote a vector time series generated by the model

$$\Delta \boldsymbol{y}_t = \boldsymbol{A}\boldsymbol{B}'\boldsymbol{y}_{t-1} + \boldsymbol{\epsilon}_t,$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are deterministic $(N \times r)$ matrices with typically $r << N$. Define $\boldsymbol{C} = (\boldsymbol{A}', \boldsymbol{B}')'$ as the $(2N \times r)$ matrix formed by vertically concatenating the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. Let $\boldsymbol{C}_{I_j}$ denote a matrix containing the columns of $\boldsymbol{C}$ indexed by I_j where $I_j = \{j, \dots, r\}$. Then, we define the loss function as

$$\begin{aligned}
\mathcal{L}(\boldsymbol{A}, \boldsymbol{B}; \boldsymbol{Y}) &= \frac{1}{T}\sum_{t=2}^{T} \|\Delta \boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{B}'\boldsymbol{y}_{t-1}\|_2^2 + P_{\boldsymbol{\lambda}}(\boldsymbol{A}, \boldsymbol{B}) \\
P_{\boldsymbol{\lambda}}(\boldsymbol{A}, \boldsymbol{B}) &= \lambda_1 (\|\boldsymbol{A}\|_1 + \|\boldsymbol{B}\|_1) + \lambda_2 \sum_{j=1}^{r} \omega_j \sqrt{\left\|\boldsymbol{A}_{I_j}\right\|_F^2 + \left\|\boldsymbol{B}_{I_j}\right\|_F^2} \qquad (1) \\
&\quad + \lambda_3 \left(\|\boldsymbol{A}\|_F^2 + \|\boldsymbol{B}\|_F^2\right),
\end{aligned}$$

where ω_j are predetermined scalar weights that will be defined later. We will decouple (1) by creating local copies of (subsets of) $\boldsymbol{A}$ and $\boldsymbol{B}$, which we can then estimate via ADMM (Boyd et al., 2011). Accordingly, we rewrite the optimization problem as follows:

$$\begin{aligned}
&\min_{\substack{\boldsymbol{A}^{(0)},\dots\boldsymbol{A}^{(N)},\boldsymbol{A},\\ \boldsymbol{B}^{(0)},\dots,\boldsymbol{B}^{(N)},\boldsymbol{B}}} \mathcal{L}(\boldsymbol{A}^{(0)}, \dots \boldsymbol{A}^{(N)}, \boldsymbol{A}, \boldsymbol{B}^{(0)}, \dots, \boldsymbol{B}^{(N)}, \boldsymbol{B}; \boldsymbol{Y}) \\
&= \min_{\substack{\boldsymbol{A}^{(0)},\dots\boldsymbol{A}^{(N)},\boldsymbol{A},\\ \boldsymbol{B}^{(0)},\dots,\boldsymbol{B}^{(N)},\boldsymbol{B}}} \left\{ \frac{1}{T}\sum_{t=2}^{T} \left\|\Delta \boldsymbol{y}_t - \boldsymbol{A}^{(0)}\boldsymbol{B}^{(0)\prime}\boldsymbol{y}_{t-1}\right\|_2^2 \right. \qquad (2) \\
&\quad + \lambda_G \sum_{j=1}^{N} \omega_j \sqrt{\left\|\boldsymbol{A}^{(j)}\right\|_F^2 + \left\|\boldsymbol{B}^{(j)}\right\|_F^2} + \lambda_I (\|\boldsymbol{A}\|_1 + \|\boldsymbol{B}\|_1)
\end{aligned}$$

subject to

$$\begin{aligned} \boldsymbol{A}^{(j)} &= \boldsymbol{A}_{I_j}, \quad j = 0, \ldots, N, \\ \boldsymbol{B}^{(j)} &= \boldsymbol{B}_{I_j}, \quad j = 0, \ldots, N, \end{aligned}$$

with $\boldsymbol{A}_{I_0} = \boldsymbol{A}$ and $\boldsymbol{B}_{I_0} = \boldsymbol{B}$. The augmented Lagrangian corresponding to (6) is given by

$$\begin{aligned} &\mathcal{L}^*(\boldsymbol{A}^{(0)}, \ldots \boldsymbol{A}^{(N)}, \boldsymbol{A}, \boldsymbol{B}^{(0)}, \ldots, \boldsymbol{B}^{(N)}, \boldsymbol{B}; \boldsymbol{Y}) \\ &= \frac{1}{T} \sum_{t=2}^{T} \left\| \Delta \boldsymbol{y}_t - \boldsymbol{A}^{(0)} \boldsymbol{B}^{(0)\prime} \boldsymbol{y}_{t-1} \right\|_2^2 \\ &\quad + \lambda_G \sum_{j=1}^{N} \omega_j \sqrt{\left\| \boldsymbol{A}^{(j)} \right\|_F^2 + \left\| \boldsymbol{B}^{(j)} \right\|_F^2} + \lambda_I \left(\|\boldsymbol{A}\|_1 + \|\boldsymbol{B}\|_1 \right) \\ &\quad + \sum_{j=0}^{N} \operatorname{tr} \left(\boldsymbol{M}_A^{(j)\prime} \left(\boldsymbol{A}^{(j)} - \boldsymbol{A}_{I_j} \right) + \boldsymbol{M}_B^{(j)\prime} \left(\boldsymbol{B}^{(j)} - \boldsymbol{B}_{I_j} \right) \right) \\ &\quad + \frac{\rho}{2} \sum_{j=0}^{N} \operatorname{tr} \left(\left(\boldsymbol{A}^{(j)} - \boldsymbol{A}_{I_j} \right)' \left(\boldsymbol{A}^{(j)} - \boldsymbol{A}_{I_j} \right) + \left(\boldsymbol{B}^{(j)} - \boldsymbol{B}_{I_j} \right)' \left(\boldsymbol{B}^{(j)} - \boldsymbol{B}_{I_j} \right) \right), \end{aligned} \tag{3}$$

where $M_A^{(j)}$ and $M_B^{(j)}$, $j = 0, \ldots, N$, are $(N \times N_j)$ Lagrange multipliers matrices. Then, the ADMM update equations are given by

$$\begin{aligned}
\boldsymbol{A}_{k+1}^{(0)} &= \arg\min_{\boldsymbol{A}} \frac{1}{T} \sum_{t=2}^{T} \left\| \Delta \boldsymbol{y}_t - \boldsymbol{A} \boldsymbol{B}_k^{(0)\prime} \boldsymbol{y}_{t-1} \right\|_2^2 + \operatorname{tr}\left(\boldsymbol{M}_{A,k}^{(0)\prime} \boldsymbol{A} \right) \\
&\quad + \frac{\rho}{2} \| \boldsymbol{A} - \boldsymbol{A}_k \|_F^2 \\
\boldsymbol{B}_{k+1}^{(0)} &= \arg\min_{\boldsymbol{A}} \frac{1}{T} \sum_{t=2}^{T} \left\| \Delta \boldsymbol{y}_t - \boldsymbol{A}_{k+1}^{(0)} \boldsymbol{B}' \boldsymbol{y}_{t-1} \right\|_2^2 + \operatorname{tr}\left(\boldsymbol{M}_{B,k}^{(0)\prime} \boldsymbol{B} \right) \\
&\quad + \frac{\rho}{2} \| \boldsymbol{B} - \boldsymbol{B}_k \|_F^2 \\
\left(\boldsymbol{A}_{k+1}^{(j)}, \boldsymbol{B}_{k+1}^{(j)} \right) &= \arg\min_{\boldsymbol{A},\boldsymbol{B}} \lambda_G \omega_j \sqrt{\|\boldsymbol{A}\|_F^2 + \|\boldsymbol{B}\|_F^2} + \operatorname{tr}\left(M_{A,k}^{(j)\prime} \boldsymbol{A} \right) + \operatorname{tr}\left(M_{B,k}^{(j)\prime} \boldsymbol{B} \right) \\
&\quad + \frac{\rho}{2} \left(\left\| \boldsymbol{A} - \boldsymbol{A}_{k,I_j} \right\|_F^2 + \left\| \boldsymbol{B} - \boldsymbol{B}_{k,I_j} \right\|_F^2 \right) \\
(\boldsymbol{A}_{k+1}, \boldsymbol{B}_{k+1}) &= \arg\min_{\boldsymbol{A},\boldsymbol{B}} \lambda_1 \left(\|\boldsymbol{A}\|_1 + \|\boldsymbol{B}\|_1 \right) - \sum_{j=0}^{N} \operatorname{tr}\left(\boldsymbol{M}_{A,k}^{(j)\prime} \boldsymbol{A}_{I_j} + \boldsymbol{M}_{B,k}^{(j)\prime} \boldsymbol{B}_{I_j} \right) \\
&\quad + \frac{\rho}{2} \sum_{j=0}^{N} \left(\left\| \boldsymbol{A}_{k+1}^{(j)} - \boldsymbol{A}_{I_j} \right\|_F^2 + \left\| \boldsymbol{B}_{k+1}^{(j)} - \boldsymbol{B}_{I_j} \right\|_F^2 \right) \\
\boldsymbol{M}_{A,k+1}^{(j)} &= \boldsymbol{M}_{A,k}^{(j)} + \rho \left(\boldsymbol{A}_{k+1}^{(j)} - \boldsymbol{A}_{I_j,k+1} \right) \\
\boldsymbol{M}_{B,k+1}^{(j)} &= \boldsymbol{M}_{B,k}^{(j)} + \rho \left(\boldsymbol{B}_{k+1}^{(j)} - \boldsymbol{B}_{I_j,k+1} \right)
\end{aligned}$$

Next, we provide analytic expressions for each of the steps in the ADMM algorithm.

Solving $\boldsymbol{A}_{k+1}^{(0)}$

First, letting $\Delta \boldsymbol{Y} = (\Delta \boldsymbol{y}_2, \ldots, \Delta \boldsymbol{y}_T)'$ and $\boldsymbol{Y}_{-1} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{T-1})'$, we note that $\boldsymbol{A}_{k+1}^{(0)}$ is the minimizer of

$$\operatorname{tr}\left(\boldsymbol{A} \boldsymbol{B}_k^{(0)\prime} \left(\frac{1}{T} \boldsymbol{Y}_{-1}' \boldsymbol{Y}_{-1} \right) \boldsymbol{B}_k^{(0)} \boldsymbol{A}' - 2 \boldsymbol{A} \boldsymbol{B}_k^{(0)\prime} \left(\frac{1}{T} \boldsymbol{Y}_{-1}' \Delta \boldsymbol{Y} \right) + \boldsymbol{M}_{A,k}^{(0)\prime} \boldsymbol{A} + \frac{\rho}{2} \boldsymbol{A}' \boldsymbol{A} - \rho \boldsymbol{A}' \boldsymbol{A}_k \right).$$

Using standard results for the derivative of the trace operator, it follows that the first order conditions for $\hat{\boldsymbol{A}}$ to be a minimizer are

$$\begin{aligned}
&2 \hat{\boldsymbol{A}} \boldsymbol{B}_k^{(0)\prime} \left(\frac{1}{T} \boldsymbol{Y}_{-1}' \boldsymbol{Y}_{-1} \right) \boldsymbol{B}_k^{(0)} - 2 \left(\frac{1}{T} \Delta \boldsymbol{Y}' \boldsymbol{Y}_{-1} \right) \boldsymbol{B}_k^{(0)} + \boldsymbol{M}_{A,k}^{(0)} + \rho \hat{\boldsymbol{A}} - \rho \boldsymbol{A}_k \\
&= \hat{\boldsymbol{A}} \left(\boldsymbol{B}_k^{(0)\prime} \left(\frac{1}{T} \boldsymbol{Y}_{-1}' \boldsymbol{Y}_{-1} \right) \boldsymbol{B}_k^{(0)} + \frac{\rho}{2} \boldsymbol{I} \right) - \left(\left(\frac{1}{T} \Delta \boldsymbol{Y}' \boldsymbol{Y}_{-1} \right) \boldsymbol{B}_k^{(0)} + \frac{\rho}{2} \boldsymbol{A}_k - \frac{1}{2} \boldsymbol{M}_{A,k}^{(0)} \right) \\
&= 0.
\end{aligned}$$

Hence, we obtain

$$\boldsymbol{A}_{k+1}^{(0)}=\left(\left(\frac{1}{T}\Delta\boldsymbol{Y}'\boldsymbol{Y}_{-1}\right)\boldsymbol{B}_k^{(0)}+\frac{\rho}{2}\boldsymbol{A}_k-\frac{1}{2}\boldsymbol{M}_{A,k}^{(0)}\right)\left(\boldsymbol{B}_k^{(0)\prime}\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)\boldsymbol{B}_k^{(0)}+\frac{\rho}{2}\boldsymbol{I}\right)^{-1}.$$

Solving $\boldsymbol{B}_{k+1}^{(0)}$

Note that $\boldsymbol{B}_{k+1}^{(0)}$ is the minimizer of

$$\operatorname{tr}\left(\boldsymbol{A}_{k+1}^{(0)}\boldsymbol{B}'\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)\boldsymbol{B}\boldsymbol{A}_{k+1}^{(0)\prime}-2\boldsymbol{A}_{k+1}^{(0)}\boldsymbol{B}'\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\Delta\boldsymbol{Y}\right)+\boldsymbol{M}_{B,k}^{(0)\prime}\boldsymbol{B}+\frac{\rho}{2}\boldsymbol{B}'\boldsymbol{B}-\rho\boldsymbol{B}'\boldsymbol{B}_k\right).$$

Using standard matrix calculus, the first order conditions can be derived to be

$$2\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)\boldsymbol{B}\boldsymbol{A}_{k+1}^{(0)\prime}\boldsymbol{A}_{k+1}^{(0)}-2\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\Delta\boldsymbol{Y}\right)\boldsymbol{A}_{k+1}^{(0)}+\boldsymbol{M}_{B,k}^{(0)}+\rho\boldsymbol{B}-\rho\boldsymbol{B}_k\overset{s}{=}\boldsymbol{0}$$
$$\Rightarrow\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)\hat{\boldsymbol{B}}\boldsymbol{A}_{k+1}^{(0)\prime}\boldsymbol{A}_{k+1}^{(0)}+\frac{\rho}{2}\hat{\boldsymbol{B}}=\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\Delta\boldsymbol{Y}\right)\boldsymbol{A}_{k+1}^{(0)}+\frac{1}{2}\left(\rho\boldsymbol{B}_k-\boldsymbol{M}_{B,k}^{(0)}\right).$$

Vectorizing both sides gives

$$\left[\left(\boldsymbol{A}_{k+1}^{(0)\prime}\boldsymbol{A}_{k+1}^{(0)}\otimes\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)+\frac{\rho}{2}\boldsymbol{I}\right]\operatorname{vec}\left(\hat{\boldsymbol{B}}\right)=\operatorname{vec}\left(\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\Delta\boldsymbol{Y}\right)\boldsymbol{A}_{k+1}^{(0)}+\frac{1}{2}\left(\rho\boldsymbol{B}_k-\boldsymbol{M}_{B,k}^{(0)}\right)\right)$$
$$\operatorname{vec}\left(\hat{\boldsymbol{B}}\right)=\left[\left(\boldsymbol{A}_{k+1}^{(0)\prime}\boldsymbol{A}_{k+1}^{(0)}\otimes\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)+\frac{\rho}{2}\boldsymbol{I}\right]^{-1}\operatorname{vec}\left(\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\Delta\boldsymbol{Y}\right)\boldsymbol{A}_{k+1}^{(0)}+\frac{1}{2}\left(\rho\boldsymbol{B}_k-\boldsymbol{M}_{B,k}^{(0)}\right)\right).$$

Hence,

$$\boldsymbol{B}_{k+1}^{(0)}=\operatorname{vec}^{-1}\left(\left[\left(\boldsymbol{A}_{k+1}^{(0)\prime}\boldsymbol{A}_{k+1}^{(0)}\otimes\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)+\frac{\rho}{2}\boldsymbol{I}\right]^{-1}\operatorname{vec}\left(\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\Delta\boldsymbol{Y}\right)\boldsymbol{A}_{k+1}^{(0)}+\frac{1}{2}\left(\rho\boldsymbol{B}_k-\boldsymbol{M}_{B,k}^{(0)}\right)\right)\right).$$

Solving $\left(\boldsymbol{A}_{k+1}^{(j)},\boldsymbol{B}_{k+1}^{(j)}\right)$

Define $\boldsymbol{c}_{k+1}^{(j)}=\operatorname{vec}\left(\boldsymbol{A}_{k+1}^{(j)},\boldsymbol{B}_{k+1}^{(j)}\right)$, $\boldsymbol{c}_{j,k}=\operatorname{vec}\left(\boldsymbol{A}_{I_j,k},\boldsymbol{B}_{I_j,k}\right)$ and $\boldsymbol{m}_k^{(j)}=\operatorname{vec}\left(\boldsymbol{M}_{A,k}^{(j)},\boldsymbol{M}_{B,k}^{(j)}\right)$. Then,

$$\begin{aligned}
\boldsymbol{c}_{k+1}^{(j)}&=\arg\min_{\boldsymbol{c}}\ \lambda\omega_j\|\boldsymbol{c}\|_2+\frac{\rho}{2}\|\boldsymbol{c}-\boldsymbol{c}_{j,k}\|_2^2+\boldsymbol{m}_k^{(j)\prime}\boldsymbol{c}\\
&=\arg\min_{\boldsymbol{c}}\ \lambda\omega_j\|\boldsymbol{c}\|_2+\frac{\rho}{2}\left\|\boldsymbol{c}-\left(\boldsymbol{c}_{j,k}-\frac{1}{\rho}\boldsymbol{m}_k^{(j)}\right)\right\|_2^2\\
&=\arg\min_{\boldsymbol{c}}\ \|\boldsymbol{c}\|_2+\frac{\rho}{2\lambda\omega_j}\left\|\boldsymbol{c}-\left(\boldsymbol{c}_{j,k}-\frac{1}{\rho}\boldsymbol{m}_k^{(j)}\right)\right\|_2^2\\
&=\operatorname{Prox}_{\frac{\lambda\omega_j}{\rho}\|\cdot\|_2}\left(\boldsymbol{c}_{j,k}-\frac{1}{\rho}\boldsymbol{m}_k^{(j)}\right),
\end{aligned}$$

where the last equation represents the proximal operator of the scaled L_2-norm. The analytic expression of this operator is well-known (e.g. Parikh et al., 2014, Section 6.5.1) to be

$$\mathrm{Prox}_{\frac{\lambda\omega_j}{\rho}\|\cdot\|_2}\left(\boldsymbol{c}_{j,k}-\frac{1}{\rho}\boldsymbol{m}_k^{(j)}\right)=\left(1-\frac{\lambda\omega_j}{\rho\left\|\boldsymbol{c}_{j,k}-\boldsymbol{m}_k^{(j)}\right\|_2}\right)_+\left(\boldsymbol{c}_{j,k}-\boldsymbol{m}_k^{(j)}\right).$$

Solving $(\boldsymbol{A}_{k+1}, \boldsymbol{B}_{k+1})$

We derive element-wise analytic solutions for $\boldsymbol{A}_{K+1}$ and $\boldsymbol{B}_{K+1}$. Define element (i,j) of $\boldsymbol{A}_{k+1}$ as by $a_{k+1,i,j}$ and let $b_{k+1,i,j}$ be defined similarly. The subdifferential that defines the relevant first-order conditions is given by

$$\begin{aligned}
&\frac{\partial}{\partial a_{i,j}}\left\{\lambda_1\|\boldsymbol{A}\|_1-\sum_{j=0}^{N}\mathrm{tr}\left(\boldsymbol{M}_{A,k}^{(j)\prime}\boldsymbol{A}_{I_j}\right)+\frac{\rho}{2}\sum_{j=0}^{N}\left\|\boldsymbol{A}^{(j)}-\boldsymbol{A}_{I_j}\right\|_F^2\right\}\\
&=\lambda_1 S_{\|\cdot\|_1}(a_{i,j})-\sum_{l=0}^{j}\left[m_{A,k,i,j}^{(l)}+\rho a_{k+1,i,j}^{(l)}\right]+\rho(j+1)a_{i,j} \qquad (4)\\
&=\lambda_1 S_{\|\cdot\|_1}(a_{i,j})-\rho\sum_{l=0}^{j}a_{k+1,i,j}^{(l)}+\rho(j+1)a_{i,j},
\end{aligned}$$

where $S_{\|\cdot\|_1}(a_{i,j})$ is the subdifferential of the L_1-norm and the last equation follows from the fact that $\sum_{l=0}^{j} m_{A,k,i,j}^{(l)}=0$, as derived in section 7.1 of Boyd et al. (2011). Hence, by setting (4) to zero and solving for $a_{i,j}$, it follows that

$$a_{k+1,i,j}=\mathrm{sign}\left(\sum_{l=0}^{j}a_{k+1,i,j}^{(l)}\right)\left(\frac{\left|\sum_{l=0}^{j}a_{k+1,i,j}^{(l)}\right|}{j+1}-\frac{\lambda_1}{\rho(j+1)}\right)_+$$

The same line of reasoning shows that

$$b_{k+1,i,j}=\mathrm{sign}\left(\sum_{l=0}^{j}b_{k+1,i,j}^{(l)}\right)\left(\frac{\left|\sum_{l=0}^{j}b_{k+1,i,j}^{(l)}\right|}{j+1}-\frac{\lambda}{\rho(j+1)}\right)_+.$$

Etienne says: "We may consider replacing $\lambda_1\|\boldsymbol{A}\|$ by $\lambda_1\sum_{i,j=1}^{N}(j+1)|a_{i,j}|$ such that the central collector will be updated as

$$b_{k+1,i,j}=\mathrm{sign}\left(\sum_{l=0}^{j}b_{k+1,i,j}^{(l)}\right)\left(\frac{\left|\sum_{l=0}^{j}b_{k+1,i,j}^{(l)}\right|}{j+1}-\frac{\lambda}{\rho}\right)_+.$$

"

2 ADMM Algorithm 2: nonconvex restriction

Let $\boldsymbol{a}_j$ and $\boldsymbol{b}_j$ denote the j-th columns of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Define $\|\cdot\|_*$ as the Nuclear norm. Consider for now the simplified estimator

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{A}, \boldsymbol{B}; \boldsymbol{Y}) &= \frac{1}{T} \sum_{t=2}^{T} \|\Delta \boldsymbol{y}_t - \boldsymbol{A}\boldsymbol{B}' \boldsymbol{y}_{t-1}\|_2^2 + P_{\boldsymbol{\lambda}}(\boldsymbol{A}, \boldsymbol{B}) \\
P_{\boldsymbol{\lambda}}(\boldsymbol{A}, \boldsymbol{B}) &= \lambda_* \|\boldsymbol{A}\boldsymbol{B}'\|_* + \lambda_2 \sum_{j=1}^{r} \omega_j \sqrt{\|\boldsymbol{a}_j\|_2^2 + \|\boldsymbol{b}_j\|_2^2},
\end{aligned} \tag{5}
$$

where $\{\omega_j\}_{j=1,\ldots,N}$ is a predetermined sequence of increasing weights. We will cast (5) into the ADMM framework (Boyd et al., 2011) by rewriting the optimization problem as

$$
\begin{aligned}
&\min_{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}} \mathcal{L}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}; \boldsymbol{Y}) \\
&= \min_{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}} \frac{1}{T} \sum_{t=2}^{T} \|\Delta \boldsymbol{y}_t - \boldsymbol{C}\boldsymbol{y}_{t-1}\|_2^2 + \lambda_* \|\boldsymbol{C}\|_* + \lambda_{L_2} \sum_{j=1}^{N} \omega_j \sqrt{\|\boldsymbol{a}_j\|_2^2 + \|\boldsymbol{b}_j\|_2^2}.
\end{aligned} \tag{6}
$$

subject to

$$
\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}'.
$$

The augmented Lagrangian corresponding to (6) is given by

$$
\begin{aligned}
&\mathcal{L}^*(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}; \boldsymbol{Y}) \\
&= \frac{1}{T} \sum_{t=2}^{T} \|\Delta \boldsymbol{y}_t - \boldsymbol{C}\boldsymbol{y}_{t-1}\|_2^2 + \lambda_* \|\boldsymbol{C}\|_* + \lambda_{L_2} \sum_{j=1}^{N} \omega_j \sqrt{\|\boldsymbol{a}_j\|_2^2 + \|\boldsymbol{b}_j\|_2^2} \\
&\quad + \operatorname{tr}\left(\boldsymbol{M}' \left(\boldsymbol{C} - \boldsymbol{A}\boldsymbol{B}'\right)\right) + \frac{\rho}{2} \|\boldsymbol{C} - \boldsymbol{A}\boldsymbol{B}'\|_F^2,
\end{aligned} \tag{7}
$$

where $\boldsymbol{M}$ is an $(N \times N)$ matrix of Lagrange multipliers. Then, the ADMM update equations are given by

$$
\begin{aligned}
\boldsymbol{C}_{k+1} &= \arg\min_{\boldsymbol{C}} \frac{1}{T} \sum_{t=2}^{T} \|\Delta \boldsymbol{y}_t - \boldsymbol{C}\boldsymbol{y}_{t-1}\|_2^2 + \lambda_* \|\boldsymbol{C}\|_* \\
&\quad + \operatorname{tr}(\boldsymbol{M}_k' \boldsymbol{C}) + \frac{\rho}{2} \|\boldsymbol{C} - \boldsymbol{A}_k \boldsymbol{B}_k'\|_F^2. \\
(\boldsymbol{A}_{k+1}, \boldsymbol{B}_{k+1}) &= \arg\min_{\boldsymbol{A},\boldsymbol{B}} \lambda_{L_2} \sum_{j=1}^{N} \omega_j \sqrt{\|\boldsymbol{a}_j\|_2^2 + \|\boldsymbol{b}_j\|_2^2} \\
&\quad - \operatorname{tr}(\boldsymbol{M}_k' \boldsymbol{A}\boldsymbol{B}') + \frac{\rho}{2} \|\boldsymbol{C}_{k+1} - \boldsymbol{A}\boldsymbol{B}'\|_F^2 \\
\boldsymbol{M}_{k+1} &\stackrel{?}{=} \boldsymbol{M}_k + \rho \left(\boldsymbol{C}_{k+1} - \boldsymbol{A}_{k+1} \boldsymbol{B}_{k+1}'\right).
\end{aligned}
$$

Updating $\boldsymbol{C}_{k+1}$

While there is no closed-form solution for this problem, it is easily computable via proximal gradient descent. Define the smooth part of the function to minimize as

$$\begin{aligned}G(\boldsymbol{C}) &= \operatorname{tr}\left(\frac{1}{T}\left(\Delta\boldsymbol{Y}-\boldsymbol{Y}_{-1}\boldsymbol{C}'\right)'\left(\Delta\boldsymbol{Y}-\boldsymbol{Y}_{-1}\boldsymbol{C}'\right)+\boldsymbol{M}_k'\boldsymbol{C}\right)\\ &\quad+\operatorname{tr}\left(\frac{\rho}{2}\left(\boldsymbol{C}-\boldsymbol{A}_k\boldsymbol{B}_k'\right)'\left(\boldsymbol{C}-\boldsymbol{A}_k\boldsymbol{B}_k'\right)\right)\\ &= \operatorname{tr}\left(\frac{1}{T}\Delta\boldsymbol{Y}'\boldsymbol{Y}+\boldsymbol{C}\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)\boldsymbol{C}'-2\boldsymbol{C}\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\Delta\boldsymbol{Y}\right)+\boldsymbol{C}'\boldsymbol{M}_k\right)\\ &\quad+\frac{\rho}{2}\operatorname{tr}\left(\boldsymbol{C}'\boldsymbol{C}-2\boldsymbol{C}'\boldsymbol{A}_k\boldsymbol{B}_k'+\boldsymbol{B}_k\boldsymbol{A}_k'\boldsymbol{A}_k\boldsymbol{B}_k'\right)\end{aligned}$$

from which it is easily seen that

$$\nabla\boldsymbol{G}(\boldsymbol{C}) = 2\boldsymbol{C}\left(\frac{1}{T}\boldsymbol{Y}_{-1}'\boldsymbol{Y}_{-1}\right)-2\left(\frac{1}{T}\Delta\boldsymbol{Y}'\boldsymbol{Y}_{-1}\right)+\boldsymbol{M}_k+\rho\left(\boldsymbol{C}-\boldsymbol{A}_k\boldsymbol{B}_k'\right).$$

Then, we iterate

$$\boldsymbol{C}_{l+1} = \operatorname{prox}_{\mathrm{tr},t\lambda_*}\left(\boldsymbol{C}_l - t\nabla\boldsymbol{G}(\boldsymbol{C}_l)\right)$$

until convergence, where the proximal operator soft-thresholds the singular values of its argument by $t\lambda_*$.

Updating $(\boldsymbol{A}_{k+1}, \boldsymbol{B}_{k+1})$

We can again tackle this optimization problem via gradient descent. First, note we can equivalently define $(\boldsymbol{A}_{k+1}, \boldsymbol{B}_{k+1})$ as

$$\begin{aligned}&\underset{\boldsymbol{A},\boldsymbol{B}}{\arg\min}\ \lambda_{L_2}\sum_{j=1}^{N}\omega_j\sqrt{\|\boldsymbol{a}_j\|_2^2+\|\boldsymbol{b}_j\|_2^2}+\frac{\rho}{2}\left\|\left(\boldsymbol{C}_{k+1}+\frac{1}{\rho}\boldsymbol{M}_k\right)-\boldsymbol{A}\boldsymbol{B}'\right\|_F^2\\ &= \underset{\boldsymbol{A},\boldsymbol{B}}{\arg\min}\ \lambda_{L_2}\sum_{j=1}^{N}\omega_j\sqrt{\|\boldsymbol{a}_j\|_2^2+\|\boldsymbol{b}_j\|_2^2}+\frac{\rho}{2}\|\boldsymbol{S}_k-\boldsymbol{A}\boldsymbol{B}'\|_F^2,\end{aligned}$$

where $\boldsymbol{S}_k = \boldsymbol{C}_{k+1}+\frac{1}{\rho}\boldsymbol{M}_k$. Define

$$g(\boldsymbol{A},\boldsymbol{B}) = \frac{\rho}{2}\|\boldsymbol{S}_k-\boldsymbol{A}\boldsymbol{B}'\|_F^2 \quad \text{and} \quad h(\boldsymbol{A},\boldsymbol{B}) = \lambda_{L_2}\sum_{j=1}^{N}\omega_j\sqrt{\|\boldsymbol{a}_j\|_2^2+\|\boldsymbol{b}_j\|_2^2}.$$

Using elementary matrix calculus, it follows that

$$\begin{aligned}\nabla_A\ g(\boldsymbol{A},\boldsymbol{B}) &= \frac{\rho}{2}\nabla_A\ \operatorname{tr}\left(\boldsymbol{A}\boldsymbol{B}'\boldsymbol{B}\boldsymbol{A}'-2\boldsymbol{A}'\boldsymbol{S}_k\boldsymbol{B}\right)\\ &= \rho(\boldsymbol{A}\boldsymbol{B}'\boldsymbol{B}-\boldsymbol{S}_k\boldsymbol{B}) = -\rho(\boldsymbol{S}_k-\boldsymbol{A}\boldsymbol{B}')\boldsymbol{B}\end{aligned}$$

and

$$\begin{aligned}\nabla_B \ g(\boldsymbol{A},\boldsymbol{B}) &= \frac{\rho}{2}\nabla_B \ \mathrm{tr}\left(\boldsymbol{B}\boldsymbol{A}'\boldsymbol{A}\boldsymbol{B}' - 2\boldsymbol{B}\boldsymbol{A}'\boldsymbol{S}_k\right) \\ &= \rho(\boldsymbol{B}\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{S}_k'\boldsymbol{A}) = -\rho(\boldsymbol{S}_k - \boldsymbol{A}\boldsymbol{B}')'\boldsymbol{A}\end{aligned}$$

Define $\tilde{\boldsymbol{a}}_{k,j}$ and $\tilde{\boldsymbol{b}}_{k,j}$ as the j-th column of $\nabla_{A_k} \ g(\boldsymbol{A}_k, \boldsymbol{B}_k)$ and $\nabla_{B_k} \ g(\boldsymbol{A}_k, \boldsymbol{B}_k)$, respectively, and let $\tilde{\boldsymbol{c}}_j = (\tilde{\boldsymbol{a}}_j', \tilde{\boldsymbol{b}}_j')'$. Denote the vectorized parameters and gradient as

$$\boldsymbol{c}_k = \mathrm{vec}\left(\begin{bmatrix}\boldsymbol{A}_k \\ \boldsymbol{B}_k\end{bmatrix}\right) \quad \text{and} \quad \nabla_{\boldsymbol{c}_k} = \mathrm{vec}\left(\begin{bmatrix}\nabla_{A_k} g(\boldsymbol{A}_k, \boldsymbol{B}_k) \\ \nabla_{B_k} g(\boldsymbol{A}_k, \boldsymbol{B}_k)\end{bmatrix}\right)$$

respectively. Then, we set $\boldsymbol{c}^{(0)} = \boldsymbol{c}_k$ and we iterate

$$\boldsymbol{c}^{(l+1)} = \mathrm{prox}_{h,t\lambda}\left(\boldsymbol{c}^{(l)} - t\nabla_{\boldsymbol{c}^{(l)}}\right) = \left(\left(1 - \frac{t\lambda\omega_j}{\left\|\boldsymbol{c}_{I_j}^{(l)}\right\|_2}\right)_+ \boldsymbol{c}_{I_j}^{(l)}\right)_{j=1,\ldots,N}$$

until convergence, say at $\boldsymbol{c}^{(L)}$. Then, we update

$$\mathrm{vec}\left(\begin{bmatrix}\boldsymbol{A}_{k+1} \\ \boldsymbol{B}_{k+1}\end{bmatrix}\right) = \mathrm{vec}^{-1}\left(\boldsymbol{c}^{(L)}\right).$$

References

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122.

Parikh, N., Boyd, S., et al. (2014). Proximal algorithms. *Foundations and trends® in Optimization*, 1(3):127–239.