4. Solution of RPCA–EWC

For convenience, the subscripts of parameters are ignored. Specifically, let $\boldsymbol{P}_0^* = \boldsymbol{P}_{\mathcal{M}_{K-1}}$, $\boldsymbol{X}_2 = \tilde{\boldsymbol{X}}_{2,K}$, $\boldsymbol{\Omega} = \zeta_K \boldsymbol{\Omega}_{\mathcal{M}_{K-1}}$. The formula (12) in the paper can be described as follows:

$$\begin{aligned}\mathcal{J}(\boldsymbol{P}) =& \mathcal{J}_K(\boldsymbol{P}) + \zeta_K \mathcal{J}_{loss}(\boldsymbol{P}, \boldsymbol{P}_0^*) \\ =& \|\boldsymbol{X}_2 - \boldsymbol{X}_2 \boldsymbol{P}\boldsymbol{P}^T\|_F^2 + \|\boldsymbol{P} - \boldsymbol{P}_0^*\|_{\boldsymbol{\Omega}}^2\end{aligned} \tag{S9}$$

The formula (S9) can be reformulated as

$$\mathcal{J}(\boldsymbol{P}) = tr(\boldsymbol{P}^T\boldsymbol{\Omega}\boldsymbol{P}) - tr(\boldsymbol{P}^T\boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{P}) - 2tr(\boldsymbol{P}^T\boldsymbol{\Omega}\boldsymbol{P}_0^*) + \underbrace{\{tr(\boldsymbol{X}_2^T\boldsymbol{X}_2) + tr(\boldsymbol{P}_0^{*T}\boldsymbol{\Omega}\boldsymbol{P}_0^*)\}}_{constant} \tag{S10}$$

Let $G(\boldsymbol{P}) = tr(\boldsymbol{P}^T\boldsymbol{\Omega}\boldsymbol{P}) - 2tr(\boldsymbol{P}^T\boldsymbol{\Omega}\boldsymbol{P}_0^*)$, $H(\boldsymbol{P}) = tr(\boldsymbol{P}^T\boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{P})$. Thus, $\mathcal{J}(\boldsymbol{P}) = G(\boldsymbol{P}) - H(\boldsymbol{P}) + constant$. The minimization of (S10) is equivalent to

$$\begin{aligned}\min_{\boldsymbol{P}} \quad & G(\boldsymbol{P}) - H(\boldsymbol{P}) \\ s.t. \quad & \boldsymbol{P}^T\boldsymbol{P} = \boldsymbol{I} \in \mathbb{R}^{l\times l}\end{aligned} \tag{S11}$$

Because $G(\boldsymbol{P})$ and $H(\boldsymbol{P})$ are convex, the objective function (S11) is actually DC programming problem [3], [4]. DC programming includes linearizing and solving the convex functions as follows [5].

Assume that $\boldsymbol{P}_i$ is the solution at ith iteration, we approximate the second part $H(\boldsymbol{P})$ by linearizing

$$H_l(\boldsymbol{P}) = H(\boldsymbol{P}_i) + \langle \boldsymbol{P} - \boldsymbol{P}_i, \boldsymbol{U}_i \rangle$$

As the subgradient $\boldsymbol{U} \in \partial H(\boldsymbol{P}) = 2\boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{P}$, let $\boldsymbol{U}_i = 2\boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{P}_i$. Then, (S11) is approximated by

$$\boldsymbol{P}_{i+1} \doteq \arg\min_{\boldsymbol{P}^T\boldsymbol{P}=\boldsymbol{I}} \quad G(\boldsymbol{P}) - < \boldsymbol{P}, \boldsymbol{U}_i >$$

Because $\boldsymbol{\Omega}$ is semidefinite, let $\boldsymbol{\Omega} = \boldsymbol{L}^T\boldsymbol{L}$ and $\boldsymbol{L}$ is the triangle matrix [5]. Thus, we can get

$$\begin{aligned}& G(\boldsymbol{P}) - < \boldsymbol{P}, \boldsymbol{U}_i > \\ =& tr(\boldsymbol{P}^T\boldsymbol{\Omega}\boldsymbol{P}) - 2tr(\boldsymbol{P}^T\boldsymbol{\Omega}\boldsymbol{P}_0^*) - 2tr(\boldsymbol{P}^T\boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{P}_i) \\ =& \langle \boldsymbol{L}\boldsymbol{P}, \boldsymbol{L}\boldsymbol{P} \rangle - 2\langle \boldsymbol{L}\boldsymbol{P}, \boldsymbol{L}\boldsymbol{P}_0^* + (\boldsymbol{L}^T)^{-1}\boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{P}_i \rangle \\ =& \|\boldsymbol{Z}_i - \boldsymbol{L}\boldsymbol{P}\|_F^2 - \|\boldsymbol{Z}_i\|_F^2\end{aligned} \tag{S12}$$

where $\boldsymbol{Z}_i = \boldsymbol{L}\boldsymbol{P}_0^* + (\boldsymbol{L}^T)^{-1}\boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{P}_i$ is constant at $(i+1)$th iteration [5]. Then, (S12) is equivalent to [5], [6]

$$\boldsymbol{P}_{i+1} = \arg\min_{\boldsymbol{P}^T\boldsymbol{P}=\boldsymbol{I}} \|\boldsymbol{P} - \boldsymbol{L}^T\boldsymbol{Z}_i\|_F^2 \tag{S13}$$

Let $\boldsymbol{Y}_i = \boldsymbol{L}^T\boldsymbol{Z}_i = \boldsymbol{\Omega}\boldsymbol{P}_0^* + \boldsymbol{X}_2^T\boldsymbol{X}_2\boldsymbol{P}_i$. According to the lemma in [6], we can obtain that $\boldsymbol{P}_{i+1} = \boldsymbol{W}_i\boldsymbol{I}_{m_2,l}\boldsymbol{V}_i^T$, where $\boldsymbol{W}_i \in \mathbb{R}^{m_2\times m_2}$ and $\boldsymbol{V}_i \in \mathbb{R}^{l\times l}$ are left and right singular values of the singular vector decomposition of $\boldsymbol{Y}_i$ [5]. m_2 is the dimension of $\boldsymbol{X}_2$ and l is the number of principal components. The procedure is summarized in Algorithm 2 in the paper.