Subsection 4.2 (with Details)

Fixing kernel matrices, the updating rules are given below:

$$\mathbf{V}^{(t+1)} = \arg\min_{||\mathbf{V}_i||_2^2<\tau,\forall i} \mathcal{L}(\tilde{\mathbf{H}}^{(t)}, \mathbf{V}; \mathbf{\Lambda}^{(t)}) + \frac{\xi^{(t)}}{2}||\mathbf{V}-\mathbf{V}^{(t)}||_F^2 \quad (8)$$

$$\tilde{\mathbf{H}}^{(t+1)} = \arg\min \mathcal{L}(\tilde{\mathbf{H}}, \mathbf{V}^{(t+1)}; \mathbf{\Lambda}^{(t)}) \quad (9)$$

$$\mathbf{\Lambda}^{(t+1)} = \mathbf{\Lambda}^{(t)} + \rho(\mathbf{V}^{(t+1)} - \tilde{\mathbf{H}}^{(t+1)}) \quad (10)$$

$$\xi^{(t+1)} = \tfrac{6}{\rho}\cdot \mathcal{J}(\tilde{\mathbf{H}}^{(t+1)}, \mathbf{B}, \mathbf{C}, \mathbf{V}^{(t+1)}) \quad (11)$$

$$\mathcal{L}(\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}; \mathbf{\Lambda}) = \mathcal{J}(\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}) + \tfrac{\rho}{2}||\mathbf{V}-\tilde{\mathbf{H}}+\Lambda/\rho||_F^2 \quad (12)$$

$\mathcal{L}(\tilde{\mathbf{H}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}, \mathbf{V}; \mathbf{\Lambda})$ is the augmented Lagrangian. Note that, proximal term $||\mathbf{V}-\mathbf{V}^{(t)}||_F^2$ and penalty parameter $\beta^{(t)}$ are added according to the study [Lu *et al.*, 2017]. The optimization w.r.t. $\mathbf{V}$ can be decomposed into k separable problems, each of which can be solved using gradient projection:

$$\begin{aligned}\mathbf{V}_{[i]}^{(r+1)} &= proj_{\mathbf{V}}(\mathbf{V}_{[i]}^{(r)} - \tau\nabla_{\mathbf{V}}[\mathcal{L} + \frac{\xi^{(t)}}{2}||\mathbf{V}-\mathbf{V}^{(t)}||_F^2]_{[i]}^{(r)}) \\ &= proj_{\mathbf{V}}(\mathbf{V}_{[i]}^{(r)} - \tau(\mathbf{A}_V^{(t)}\mathbf{V}_{[i]}^{(r)} - \mathbf{B}_{V}{}_{[i]}^{(t)})),\end{aligned} \quad (13)$$

$$\mathbf{A}_V^{(t)} = \tilde{\mathbf{B}}^T\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\tilde{\mathbf{B}} + \tilde{\mathbf{C}}^T\tilde{\mathbf{H}}^T\tilde{\mathbf{H}}\tilde{\mathbf{C}} + (\tfrac{\xi^{(t)}}{2}+1)\mathbf{I} \quad (14)$$

$$\mathbf{B}_V^{(t)} = \tilde{\mathbf{M}}^T\tilde{\mathbf{H}}\tilde{\mathbf{B}} + \tilde{\mathbf{S}}^T\tilde{\mathbf{H}}\tilde{\mathbf{C}} + \tilde{\mathbf{H}} + \tfrac{\xi^{(t)}}{2}\mathbf{H}^{(t)} - \mathbf{\Lambda}/\rho \quad (15)$$

$$proj_{\mathbf{V}}(\boldsymbol{w}) = \sqrt{\tau}\boldsymbol{w}/\max\{\sqrt{\tau}, ||\boldsymbol{w}||_2\},\ \forall \boldsymbol{w}\in\mathbb{R}^n \quad (16)$$

where r is the inner-iteration number, α is the step size and $\mathbf{V}_{[i]}$ denotes the i^{th} column of matrix $\mathbf{V}$. For a given vector $\mathbf{w}$, $proj_{\mathbf{V}}(\cdot)$ projects it onto the feasible set of $\mathbf{V}_{[i]}$. $\tilde{\mathbf{H}}$ can be solved via 1^{st}-order method, whose gradient is:

$$\begin{aligned}\nabla_{\tilde{\mathbf{H}}}\mathcal{L} &= \nabla_{\tilde{\mathbf{H}}}\mathcal{J} + \frac{\rho}{2}\nabla_{\tilde{\mathbf{H}}}||\mathbf{V}-\tilde{\mathbf{H}}+\Lambda/\rho||_F^2 \\ &= \tilde{\mathbf{H}}\tilde{\mathbf{B}}\mathbf{V}^{(t+1)^T}\mathbf{V}^{(t+1)}\tilde{\mathbf{B}}^T - \tilde{\mathbf{M}}\mathbf{V}^{(t+1)}\tilde{\mathbf{B}}^T + \tilde{\mathbf{E}}\tilde{\mathbf{D}}_p\tilde{\mathbf{E}}\tilde{\mathbf{H}}\tilde{\mathbf{D}}_p \\ &+ \tilde{\mathbf{H}}\tilde{\mathbf{C}}\mathbf{V}^{(t+1)^T}\mathbf{V}^{(t+1)}\tilde{\mathbf{C}}^T - \tilde{\mathbf{S}}\mathbf{V}^{(t+1)}\tilde{\mathbf{C}}^T + \tilde{\mathbf{E}}\tilde{\mathbf{D}}_q\tilde{\mathbf{E}}\tilde{\mathbf{H}}\tilde{\mathbf{D}}_q \\ &+ \rho(\tilde{\mathbf{H}} - \mathbf{V}^{(t+1)} - \mathbf{\Lambda}^{(t)}/\rho) + 2\tilde{\mathbf{E}}^T\tilde{\mathbf{E}}\tilde{\mathbf{H}} \\ &= (\tilde{\mathbf{H}}\tilde{\mathbf{B}}\mathbf{V}^{(t+1)^T} - \tilde{\mathbf{M}})\mathbf{V}^{(t+1)}\tilde{\mathbf{B}}^T + \tilde{\mathbf{E}}\,\mathcal{R}_p(\tilde{\mathbf{E}}\tilde{\mathbf{H}}) \\ &+ (\tilde{\mathbf{H}}\tilde{\mathbf{C}}\mathbf{V}^{(t+1)^T} - \tilde{\mathbf{S}})\mathbf{V}^{(t+1)}\tilde{\mathbf{C}}^T + \tilde{\mathbf{E}}\,\mathcal{R}_q(\tilde{\mathbf{E}}\tilde{\mathbf{H}}) \\ &+ \rho(\tilde{\mathbf{H}} - \mathbf{V}^{(t+1)} - \mathbf{\Lambda}^{(t)}/\rho) + 2\tilde{\mathbf{E}}^T\tilde{\mathbf{E}}\tilde{\mathbf{H}},\end{aligned} \quad (17)$$

as $\mathbf{D}_p = \mathbf{D}_p{}^{-1}$ and $\mathbf{D}_q = \mathbf{D}_q{}^{-1}$.

References

[Lu *et al.*, 2017] Songtao Lu, Mingyi Hong, and Zhengdao Wang. A nonconvex splitting method for symmetric nonnegative matrix factorization: Convergence analysis and optimality. *IEEE Trans. on Signal Processing*, 2017.