Closed form

Sohyeon Kim

$$
\begin{aligned}
L_\delta(\theta,\eta,\alpha,e,u,w) = &\sum_{g=1}^{m}\sum_{\ell=1}^{b}\sum_{i=1}^{n}\frac{1}{nb}\rho_{\tau_\ell}(e_i^{(\ell)(g)}) + \sum_{g=1}^{m}\sum_{\ell=1}^{b}\boldsymbol{u}^{(\ell)(g)T}(\boldsymbol{Y}^{(g)} - \boldsymbol{Z}^{(g)} - \boldsymbol{V}^{(\ell)}\boldsymbol{\eta}^{(g)} - \boldsymbol{e}^{(\ell)(g)}) \\
&+ \frac{\delta}{2}\sum_{g=1}^{m}\sum_{\ell=1}^{b}||\boldsymbol{Y}^{(g)} - \boldsymbol{Z}^{(g)} - \boldsymbol{V}^{(\ell)}\boldsymbol{\eta}^{(g)} - \boldsymbol{e}^{(\ell)(g)}||_2^2 \\
&+ \sum_{g=1}^{m}\boldsymbol{w}^{(g)T}(\boldsymbol{\theta}^{(g)} - \boldsymbol{\eta}^{(g)}) + \frac{\delta}{2}\sum_{g=1}^{m}||\boldsymbol{\theta}^{(g)} - \boldsymbol{\eta}^{(g)}||_2^2 \\
&+ \lambda_1||\mathbf{Z}||_{*,\psi} + \lambda_2\sum_{g=1}^{m}\sum_{j=1}^{p}\phi_j^{(g)}||\boldsymbol{\theta}_j^{(g)}||_2,
\end{aligned} \tag{1}
$$

where $||\cdot||_{*,\psi}$ is adaptive nuclear norm.
Let (1) be Q.

1 update η

$$
\begin{aligned}
\frac{\partial Q}{\partial \boldsymbol{\eta}^{(g)T}} = &-\sum_{\ell}^{b}\boldsymbol{V}^{(\ell)T}\boldsymbol{u}^{(\ell)} - \delta\sum_{\ell=1}^{b}V^{(\ell)T}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{e}^{(\ell)}) \\
&+ \delta\sum_{\ell}^{b}\boldsymbol{V}^{(\ell)T}\boldsymbol{V}^{(\ell)}\boldsymbol{\eta}^{(\boldsymbol{g})} - \boldsymbol{w} + \delta\boldsymbol{\eta}^{(g)} - \delta\boldsymbol{\theta}^{(g)}
\end{aligned}
$$

$$
\delta\Big(\sum_{\ell=1}^{b}V^{(\ell)T}V^{(\ell)} + I\Big)\eta^{(g)} = \sum_{\ell}^{b}\boldsymbol{V}^{(\ell)T}\boldsymbol{u}^{(\ell)} + \delta\sum_{\ell=1}^{b}V^{(\ell)T}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{e}^{(\ell)}) + \mathbf{w} + \delta\boldsymbol{\theta}^{(g)}
$$

$$
\therefore \hat{\boldsymbol{\eta}}^{(g)} = \frac{1}{\delta}(\sum_{\ell}^{b}\boldsymbol{V}^{(\ell)T}\boldsymbol{V}^{(\ell)} + \boldsymbol{I})^{-1}(\boldsymbol{w} + \delta\boldsymbol{\theta}^{(g)} + \sum_{\ell}^{b}\boldsymbol{V}^{(\ell)T}\boldsymbol{u} + \delta\sum_{\ell}^{b}\boldsymbol{V}^{(\ell)T}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{e}))
$$

2 update θ

$$\theta_j^{(g)^{k+1}} := argmin_{\theta_j^{(g)} \in \mathbb{R}^K} \sum_{g=1}^{m} \boldsymbol{w}^{(g)^k T} \theta_j^{(g)} + \frac{\delta}{2} \sum_{g=1}^{m} ||\theta_j^{(g)} - \eta_j^{(g)^{k+1}}||_2^2 + \lambda_2 \sum_{g=1}^{m} \phi_j^{(g)} ||\theta_j^{(g)}||_2$$

$$\frac{\partial Q}{\partial \theta_j^{(g)T}} = w_j^{(g)} + \delta \theta_j^{(g)} - \delta \eta_j^{(g)} + \lambda_2 \cdot \phi_j^{(g)} \frac{\partial ||\theta_j^{(g)}||_2}{\partial \theta_j^{(g)}}$$

$$\frac{\partial ||\theta_j^{(g)}||_2}{\partial \theta_j^{(g)}} = \begin{cases} \theta_j^{(g)} / ||\theta_j^{(g)}||_2 & \text{if } \theta_j^{(g)} \neq 0 \\ \{u : ||u||_2 \leq 1\} & \text{if } \theta_j^{(g)} = 0 \end{cases}$$

2.1 if $\theta_j^{(g)} \neq 0$

$$w_j^{(g)} + \delta \theta_j^{(g)} - \delta \eta_j^{(g)} + \lambda_2 \cdot \phi_j^{(g)} \frac{\theta_j^{(g)}}{||\theta_j^{(g)}||_2} = 0$$

$$(1 + \frac{\lambda_2 \cdot \phi_j^{(g)}}{\delta \cdot ||\theta_j^{(g)}||_2}) \theta_j^{(g)} = \underbrace{\eta_j^{(g)} - \frac{w_j^{(g)}}{\delta}}_{r_j^{(g)}}$$

$$||r_j^{(g)}||_2 = (1 + \frac{\lambda_2 \cdot \phi_j^{(g)} / \delta}{||\theta_j^{(g)}||_2}) ||\theta_j^{(g)}||_2$$

$$\rightarrow ||\theta_j^{(g)}||_2 = ||r_j^{(g)}||_2 - \lambda_2 \cdot \phi_j^{(g)} / \delta \geq 0$$

$$\therefore \theta_j^{(g)} = r_j^{(g)} (1 + \frac{\lambda_2 \cdot \phi_j^{(g)} / \delta}{||\theta_j^{(g)}||_2})^{-1}$$

$$= (1 - \frac{\lambda_2 \cdot \phi_j^{(g)} / \delta}{||r_j^{(g)}||_2}) r_j^{(g)}$$

2.2 if $\theta_j^{(g)} = 0$

$$w_j^{(g)} + \delta \underbrace{\theta_j^{(g)}}_{=0} - \delta \eta_j^{(g)} + \lambda_2 \cdot \phi_j^{(g)} u = 0$$

$$\frac{\lambda_2 \cdot \phi_j^{(g)}}{\delta} u = \eta_j^{(g)} - \frac{w_j^{(g)}}{\delta} = r_j^{(g)}$$

$$\frac{\lambda_2 \cdot \phi_j^{(g)}}{\delta} ||u||_2 = ||r_j^{(g)}||_2 \leq \frac{\lambda_2 \cdot \phi_j^{(g)}}{\delta}$$

$$\therefore \hat{\theta}_j^{(g)} = \begin{cases} (1 - \frac{\lambda_2 \cdot \phi_j^{(g)} / \delta}{||r_j^{(g)}||_2}) r_j^{(g)} & \text{if } ||r_j^{(g)}||_2 \geq \lambda_2 \cdot \phi_j^{(g)} / \delta, \\ 0 & \text{if } ||r_j^{(g)}||_2 < \lambda_2 \cdot \phi_j^{(g)} / \delta \end{cases}$$

3 update $\mathbf{Z} = \mathbf{X}\boldsymbol{\alpha}$

$$
\begin{aligned}
(\boldsymbol{XA})^{k+1} &= \operatorname{argmin}_{\boldsymbol{XA}} \underbrace{-\sum_{\ell}^{b} \operatorname{tr}(\boldsymbol{U}^{(\ell)^T}\boldsymbol{XA}) + \frac{\delta}{2}\sum_{\ell=1}^{b} ||\boldsymbol{Y} - \boldsymbol{XA} - \boldsymbol{V}^{(\ell)}\boldsymbol{H}^{k+1} - \boldsymbol{E}^{(\ell)^k}||_F^2}_{(*)} + \lambda_1||\boldsymbol{XA}||_{*\boldsymbol{\psi}} \\
&= \operatorname{argmin}_Z \frac{1}{2}||\frac{1}{b}\sum_{\ell}^{b}(\boldsymbol{Y} - \boldsymbol{V}^{(\ell)}\boldsymbol{H}^{k+1} - \boldsymbol{E}^{(\ell)^k} + \frac{1}{\delta}\boldsymbol{U}^{(\ell)}) - \boldsymbol{Z}||_F^2 + \frac{\lambda_1}{\delta \cdot b}||\boldsymbol{Z}||_{*\boldsymbol{\psi}} \\
&= \boldsymbol{PD}_{\lambda_1/\delta\cdot b,\boldsymbol{\psi}}(\Sigma)\boldsymbol{Q}^T,
\end{aligned}
$$

where $\boldsymbol{P\Sigma Q}^T$ is the SVD of $\frac{1}{b}\sum_{\ell}^{b}(\boldsymbol{Y} - \boldsymbol{V}^{(\ell)}\boldsymbol{H}^{k+1} - \boldsymbol{E}^{(\ell)^k} + \frac{1}{\delta}\boldsymbol{U}^{(\ell)})$,
and $\boldsymbol{D}_{\lambda_1/\delta\cdot b,\boldsymbol{\psi}}(\Sigma) = \operatorname{diag}(\{\sigma_i - \frac{\lambda_1\boldsymbol{\psi}_i}{(\delta\cdot b)}\}_+)$.

$$
\begin{aligned}
(*) &= \sum_{\ell}\left\{\operatorname{tr}\left(\frac{\delta}{2}(Y - Z - V^{(\ell)}H - E^{(\ell)})^T(Y - Z - V^{(\ell)}H - E^{(\ell)}) - U^{(\ell)T}Z\right)\right\} \\
&= \frac{\delta}{2}\sum_{\ell}\left\{\operatorname{tr}\left((Y - Z - V^{(\ell)}H - E^{(\ell)})^T(Y - Z - V^{(\ell)}H - E^{(\ell)}) - \frac{2}{\delta}U^{(\ell)T}Z\right)\right\} \\
&\propto \frac{\delta}{2}\sum_{\ell}\left\{\operatorname{tr}\left(Z^TZ - 2(Y - V^{(\ell)}H - E^{(\ell)} + \frac{1}{\delta}U^{(\ell)})^TZ\right)\right\} \\
&= \frac{\delta\cdot b}{2}\left\{\operatorname{tr}\left(Z^TZ - 2(\frac{1}{b}\sum_{\ell} Y - V^{(\ell)}H - E^{(\ell)} + \frac{1}{\delta}U^{(\ell)})^TZ\right)\right\} \\
&\propto \frac{\delta\cdot b}{2}\left\{\operatorname{tr}\left(\left[Z - \frac{1}{b}\sum_{\ell}(Y - V^{(\ell)}H - E^{(\ell)} + \frac{1}{\delta}U^{(\ell)})\right]^T\left[Z - \frac{1}{b}\sum_{\ell}(Y - V^{(\ell)}H - E^{(\ell)} + \frac{1}{\delta}U^{(\ell)})\right]\right)\right\} \\
&= \frac{\delta\cdot b}{2}||\frac{1}{b}\sum_{\ell}(Y - V^{(\ell)}H - E^{(\ell)} + \frac{1}{\delta}U^{(\ell)}) - Z||_F^2
\end{aligned}
$$

4 update e

$$
e_i^{(\ell)(g)} = \operatorname{argmin}\ \frac{1}{nb}\rho_{\tau_\ell}(e_i^{(\ell)(g)}) - u_i^{(\ell)(g)}e_i^{(\ell)(g)} + \frac{\delta}{2}(Y_i^{(g)} - X_i\alpha^{(g)} - V_i^{(\ell)}\eta^{(g)} - e_i^{(\ell)(g)})^2
$$

$$
\frac{\partial Q}{\partial e_i^{(\ell)(g)}} = -\delta(Y_i^{(g)} - X_i\alpha^{(g)} - V_i^{(\ell)}\eta^{(g)} - e_i^{(\ell)(g)}) - u_i^{(\ell)(g)} + \frac{1}{nb}\frac{\partial\rho_{\tau_\ell}(e_i^{(\ell)(g)})}{\partial e_i^{(\ell)(g)}}
$$

$$
\frac{\partial\rho_{\tau_\ell}(e_i^{(\ell)(g)})}{\partial e_i^{(\ell)(g)}} = \begin{cases} \tau_\ell - 1 & \text{if } e_i^{(\ell)(g)} < 0 \\ \{c \in \mathbb{R} : \tau_\ell - 1 \le c \le \tau_\ell\} & \text{if } e_i^{(\ell)(g)} = 0 \\ \tau_\ell & \text{if } e_i^{(\ell)(g)} > 0 \end{cases}
$$

Let $\epsilon_i^{(\ell)(g)} = Y_i^{(g)} - X_i^T\alpha^{(g)} - V_i^{(\ell)}\eta^{(g)}$.

4.1 if $e_i^{(\ell)(g)} < 0$

$$-\delta(\epsilon_i^{(\ell)(g)} - e_i^{(\ell)(g)}) - u_i^{(\ell)(g)} + \frac{1}{nb}(\tau_\ell - 1) = 0$$

$$e_i^{(\ell)(g)} = \epsilon_i^{(\ell)(g)} + \frac{u_i^{(\ell)(g)}}{\delta} - \frac{\tau_\ell - 1}{nb\delta} < 0 \quad \text{by hypo}$$

4.2 if $e_i^{(\ell)(g)} > 0$

$$-\delta(\epsilon_i^{(\ell)(g)} - e_i^{(\ell)(g)}) - u_i^{(\ell)(g)} + \frac{1}{nb}\tau_\ell = 0$$

$$e_i^{(\ell)(g)} = \epsilon_i^{(\ell)(g)} + \frac{u_i^{(\ell)(g)}}{\delta} - \frac{\tau_\ell}{nb\delta} > 0 \quad \text{by hypo}$$

4.3 if $e_i^{(\ell)(g)} = 0$

$$-\delta\epsilon_i^{(\ell)(g)} - u_i^{(\ell)(g)} + \frac{c}{nb} = 0$$

$$\frac{\tau_\ell - 1}{nb\delta} \leq (\epsilon_i^{(\ell)(g)}) + \frac{u_i^{(\ell)(g)}}{\delta} = \frac{c}{nb\delta} \leq \frac{\tau_\ell}{nb\delta}$$

$$\therefore e_i^{(\ell)(g)^{k+1}} = \begin{cases} \epsilon_i^{(\ell)(g)} + u_i^{(\ell)(g)}/\delta - (\tau_\ell - 1)/nb\delta & \text{if } \epsilon_i^{(\ell)(g)} + u_i^{(\ell)(g)}/\delta < (\tau_\ell - 1)/(nb\delta) \\ 0 & \text{if } (\tau_\ell - 1)/(nb\delta) \leq \epsilon_i^{(\ell)(g)} + u_i^{(\ell)(g)}/\delta \leq \tau_\ell/(nb\delta) \\ \epsilon_i^{(\ell)(g)} + u_i^{(\ell)(g)}/\delta - \tau_\ell/nb\delta & \text{if } \epsilon_i^{(\ell)(g)} + u_i^{(\ell)(g)}/\delta > \tau_\ell/(nb\delta) \end{cases}$$

5 update multiplier $\mathbf{u}, \mathbf{w}$

$$\begin{aligned} \boldsymbol{u}^{(\ell)(g)^{k+1}} :=& \boldsymbol{u}^{(\ell)(g)^k} + \delta(\boldsymbol{Y}^{(g)} - \boldsymbol{Z}[\quad, g] - \boldsymbol{V}^{(\ell)}\boldsymbol{\eta}^{(g)^{k+1}} - \boldsymbol{e}^{(\ell)(g)^{k+1}}) \\ \boldsymbol{w}^{(g)^{k+1}} :=& \boldsymbol{w}^{(g)^k} + \delta(\boldsymbol{\theta}^{(g)^{k+1}} - \boldsymbol{\eta}^{(g)^{k+1}}) \end{aligned}$$