ADMM Derivation in FGST

1 Problem Setting

$$
\begin{aligned}
\min_W L = & \sum_{k=1}^{K}\sum_{n=1}^{N}(X_n^k {W_{in}}_n^k - {Y_{in}}_n^k)^2 + \frac{1}{2}\sum_{k=1}^{K}\|{W_{in}}^k P\|_1 + \sum_{n=1}^{N}\|{W_{in}}_n Q\|_1 \\
& + \sum_{k=1}^{K}\sum_{n=1}^{N}(X_n^k {W_{out}}_n^k - {Y_{out}}_n^k)^2 + \frac{1}{2}\sum_{k=1}^{K}\|{W_{out}}^k P\|_1 + \sum_{n=1}^{N}\|{W_{out}}_n Q\|_1 \\
& + \sum_{k=1}^{K}\sum_{m=1}^{M} h(m)|\sum_{i\in C_m} X_i^k({W_{in}}_i^k - {W_{out}}_i^k)| + \theta\left(\sum_{k=1}^{K}\sum_{n=1}^{N}(\|{W_{in}}_n^k\|^2) + \sum_{k=1}^{K}\sum_{n=1}^{N}(\|{W_{out}}_n^k\|^2)\right)
\end{aligned}
\tag{1.1}
$$

2 Optimization Method

We introduce auxiliary variable matrices:

$$
\begin{aligned}
E^k &= {W_{in}}^k P, \quad k = 1, \dots, K \\
F_n &= {W_{in}}_n Q, \quad n = 1, \dots, N \\
G^k &= {W_{out}}^k P, \quad k = 1, \dots, K \\
H_n &= {W_{out}}_n Q, \quad n = 1, \dots, N
\end{aligned}
$$

We can rewrite the original problem into

$$
\begin{aligned}
\min_W L = & \sum_{k=1}^{K}\sum_{n=1}^{N}(X_n^k {W_{in}}_n^k - {Y_{in}}_n^k)^2 + \frac{1}{2}\sum_{k=1}^{K}\|E^k\|_1 + \sum_{n=1}^{N}\|F_n\|_1 \\
& + \sum_{k=1}^{K}\sum_{n=1}^{N}(X_n^k {W_{out}}_n^k - {Y_{out}}_n^k)^2 + \frac{1}{2}\sum_{k=1}^{K}\|G^k\|_1 + \sum_{n=1}^{N}\|H_n\|_1 \\
& + \sum_{k=1}^{K}\sum_{m=1}^{M} h(m)\left|\sum_{i\in C_m} X_i^k({W_{in}}_i^k - {W_{out}}_i^k)\right| + \theta\left(\sum_{k=1}^{K}\sum_{n=1}^{N}(\|{W_{in}}_n^k\|^2) + \sum_{k=1}^{K}\sum_{n=1}^{N}(\|{W_{out}}_n^k\|^2)\right) \\
\text{s.t. } & E^k = {W_{in}}^k P, \quad k = 1,\dots,K \\
& F_n = {W_{in}}_n Q, \quad n = 1,\dots,N \\
& G^k = {W_{out}}^k P, \quad k = 1,\dots,K \\
& H_n = {W_{out}}_n Q, \quad n = 1,\dots,N
\end{aligned}
\tag{2.1}
$$

The ADMM objective function of Eq(2.1) can be written as

$$
\begin{aligned}
\min\ & L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R) \\
= & \sum_{k=1}^{K}\sum_{n=1}^{N}(X_n^k {W_{in}}_n^k - {Y_{in}}_n^k)^2 \\
& + \frac{1}{2}\sum_{k=1}^{K}\|E^k\|_1 + \frac{\rho}{2}\sum_{k=1}^{K}\|{W_{in}}^k P - E^k + U^k\|_F^2 \\
& + \sum_{n=1}^{N}\|F_n\|_1 + \frac{\rho}{2}\sum_{n=1}^{N}\|{W_{in}}_n Q - F_n + V_n\|_F^2 \\
& + \sum_{k=1}^{K}\sum_{n=1}^{N}(X_n^k {W_{out}}_n^k - {Y_{out}}_n^k)^2 \\
& + \frac{1}{2}\sum_{k=1}^{K}\|G^k\|_1 + \frac{\rho}{2}\sum_{k=1}^{K}\|{W_{out}}^k P - G^k + S^k\|_F^2 \\
& + \sum_{n=1}^{N}\|H_n\|_1 + \frac{\rho}{2}\sum_{n=1}^{N}\|{W_{out}}_n Q - H_n + R_n\|_F^2 \\
& + \sum_{k=1}^{K}\sum_{m=1}^{M} h(m)\left|\sum_{i\in C_m} X_i^k({W_{in}}_i^k - {W_{out}}_i^k)\right| + \theta\left(\sum_{k=1}^{K}\sum_{n=1}^{N}(\|{W_{in}}_n^k\|^2) + \sum_{k=1}^{K}\sum_{n=1}^{N}(\|{W_{out}}_n^k\|^2)\right)
\end{aligned}
\tag{2.2}
$$

The $t+1$-th iteration of ADMM of (2.2) consists of the following procedures:

$$
\begin{aligned}
{W_{in}}_n^k(t+1) &= \operatorname*{argmin}_{{W_{in}}_n^k} L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R) && (2.3)\\
E^k(t+1) &= \mathcal{S}_{\frac{1}{2}/\rho}\left({W_{in}}^k(t+1)P + U^k(t)\right) && (2.4)\\
F_n(t+1) &= \mathcal{S}_{1/\rho}\left({W_{in}}_n(t+1)Q + V_n(t)\right) && (2.5)\\
{W_{out}}_n^k(t+1) &= \operatorname*{argmin}_{{W_{out}}_n^k} L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R) && (2.6)\\
G^k(t+1) &= \mathcal{S}_{\frac{1}{2}/\rho}\left({W_{out}}^k(t+1)P + S^k(t)\right) && (2.7)\\
H_n(t+1) &= \mathcal{S}_{1/\rho}\left({W_{out}}_n(t+1)Q + R_n(t)\right) && (2.8)\\
U^k(t+1) &= U^k(t) + {W_{in}}^k(t+1)P - E^k(t+1) && (2.9)\\
V_n(t+1) &= V_n(t) + {W_{in}}^k(t+1)Q - F_n(t+1) && (2.10)\\
S^k(t+1) &= S^k(t) + {W_{out}}^k(t+1)P - G^k(t+1) && (2.11)\\
R_n(t+1) &= R_n(t) + {W_{out}}^k(t+1)Q - H_n(t+1) && (2.12)
\end{aligned}
$$

where the soft thresholding operator $\mathcal{S}$ is defined as

$$
\mathcal{S}_\alpha(x) = \begin{cases} x-\alpha & \text{if } x > \alpha \\ 0 & \text{if } |x| \le \alpha \\ x+\alpha & \text{if } x < -\alpha \end{cases} \tag{2.13}
$$

In the update of ${W_{in}}_n^k$ and ${W_{out}}_n^k$, we use SGD to get the update formula. The gradient of $L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)$ with respect to ${W_{in}}_n^k$ is :

$$
\begin{aligned}
&\frac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{in}}_n^k}\\
&= 2(X_n^k {W_{in}}_n^k - {Y_{in}}_n^k)\cdot (X_n^k)^T + \rho({W_{in}}^k(t)P - E^k(t) + U^k(t))\cdot (P_n)^T \qquad (2.14)\\
&\quad + \rho({W_{in}}_n(t)Q - F_n(t) + V_n(t))\cdot (Q^k)^T + 2\theta {W_{in}}_n^k + \frac{\partial g(W)}{\partial {W_{in}}_n^k}
\end{aligned}
$$

where $g(W) = \sum_{k=1}^K \sum_{m=1}^M h(m)\left|\sum_{i\in C_m} X_i^k({W_{in}}_i^k - {W_{out}}_i^k)\right|$ and

$$
\frac{\partial g(W)}{\partial {W_{in}}_n^k} = \operatorname{sgn}\left(\sum_{i\in C_m} X_i^k({W_{in}}_i^k - {W_{out}}_i^k)\right)\cdot (X_n^k)^T \tag{2.15}
$$

where

$$
\operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}
$$

Algorithm 1 The ADMM Framework

Input:

Output:

Initialize

while Not Convergent **do**

  Select integer $k \in [1, K]$ and $n \in [1, N]$ randomly

  while Not Convergent **do**

    Calculate $\dfrac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{in}}_n^k}$ according to Eq.(2.14)

    Update ${W_{in}}_n^k \leftarrow {W_{in}}_n^k - \gamma \dfrac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{in}}_n^k}$

  end while

  Update E^k according to Eq.(2.4)

  Update F_n according to Eq.(2.5)

  while Not Convergent **do**

    Calculate $\dfrac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{out}}_n^k}$

    Update ${W_{out}}_n^k \leftarrow {W_{out}}_n^k - \gamma \dfrac{\partial L_\rho(W_{in}, W_{out}, E, F, G, H, U, V, S, R)}{\partial {W_{out}}_n^k}$

  end while

  Update G^k according to Eq.(2.7)

  Update H_n according to Eq.(2.8)

  Update U^k according to Eq.(2.9)

  Update V_n according to Eq.(2.10)

  Update S^k according to Eq.(2.11)

  Update R_n according to Eq.(2.12)

end while