1 Goal

The goal is to optimize the following:

$$\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \mathbf{PLQ}(y_i,\ \mathbf{x}_i^\top \boldsymbol{\beta}) + \frac{1}{2}||\boldsymbol{\beta}||_2^2 + \rho||\boldsymbol{\beta}||_1$$
$$\textbf{s. t. } \mathbf{A}\boldsymbol{\beta} + \mathbf{b} \geq \mathbf{0}$$

After ReLU-ReHU decomposition on the PLQ loss:

$$\min_{\boldsymbol{\beta}} \sum_{l,i} \mathbf{ReLU}(u_{li}\mathbf{x}_i^\top \boldsymbol{\beta} + v_{li}) + \sum_{h,i} \mathbf{ReHU}_{\tau_{hi}}(s_{hi}\mathbf{x}_i^\top \boldsymbol{\beta} + t_{hi}) + \frac{1}{2}||\boldsymbol{\beta}||_2^2 + \rho \sum_{j=1}^{d} |\beta_j|$$
$$\textbf{s. t. } \mathbf{A}\boldsymbol{\beta} + \mathbf{b} \geq \mathbf{0}$$

$$\iff \quad \min_{\boldsymbol{\beta},\Pi,\Sigma,\Theta} \sum_{l,i} \pi_{li} + \sum_{h,i} \frac{1}{2}\theta_{hi}^2 + \sum_{h,i} \tau_{hi}\sigma_{hi} + \frac{1}{2}||\boldsymbol{\beta}||_2^2 + \rho \sum_{j=1}^{d} |\beta_j|$$
$$\textbf{s. t. } \mathbf{A}\boldsymbol{\beta} + \mathbf{b} \geq \mathbf{0},\ \pi_{li} - u_{li}\mathbf{x}_i^\top \boldsymbol{\beta} - v_{li} \geq 0,$$
$$\theta_{hi} + \sigma_{hi} - s_{hi}\mathbf{x}_i^\top \boldsymbol{\beta} - t_{hi} \geq 0,$$
$$\pi_{li} \geq 0,\ \theta_{hi} \geq 0$$

Then we substitute ℓ_1 norm terms:

$$\min_{\boldsymbol{\beta},\Pi,\Sigma,\Theta} \sum_{l,i} \pi_{li} + \sum_{h,i} \frac{1}{2}\theta_{hi}^2 + \sum_{h,i} \tau_{hi}\sigma_{hi} + \frac{1}{2}||\boldsymbol{\beta}||_2^2 + \rho \sum_{j=1}^{d} \omega_j$$
$$\textbf{s. t. } \mathbf{a}_k^\top \boldsymbol{\beta} + b_k \geq \mathbf{0},$$
$$\pi_{li} - u_{li}\mathbf{x}_i^\top \boldsymbol{\beta} - v_{li} \geq 0,$$
$$\theta_{hi} + \sigma_{hi} - s_{hi}\mathbf{x}_i^\top \boldsymbol{\beta} - t_{hi} \geq 0,$$
$$\pi_{li} \geq 0,$$
$$\theta_{hi} \geq 0,$$
$$\omega_j - \beta_j \geq 0,$$
$$\omega_j + \beta_j \geq 0.$$

2 Prime Problem

$$
\begin{aligned}
\mathcal{L}_P = &\sum_{l,i} \pi_{li} + \sum_{h,i} \frac{1}{2}\theta_{hi}^2 + \sum_{h,i} \tau_{hi}\sigma_{hi} + \frac{1}{2}||\boldsymbol{\beta}||_2^2 + \rho \sum_{j=1}^{d} \omega_j \\
&- \sum_{k} \xi_k \mathbf{a}_k^\top \boldsymbol{\beta} + \sum_{k} \xi_k b_k - \sum_{l,i} \lambda_{l,i}(\pi_{li} - u_{li}\mathbf{x}_i^\top \boldsymbol{\beta} - v_{li}) \\
&- \sum_{h,i} \gamma_{hi}(\theta_{hi} + \sigma_{hi} - s_{hi}\mathbf{x}_i^\top \boldsymbol{\beta} - t_{hi}) \\
&- \sum_{l,i} \delta_{li}\pi_{li} - \sum_{h,i} \psi_{hi}\sigma_{hi} \\
&- \sum_{j=1}^{d} \eta_j(\omega_j - \beta_j) - \sum_{j=1}^{d} \mu_j(\omega_j + \beta_j)
\end{aligned}
$$

KKT Condition gives:

$$\theta_{hi} = \gamma_{hi}$$

$$\delta_{li} = 1 - \lambda_{li}$$

$$\psi_{hi} = \tau_{hi} - \gamma_{hi}$$

$$\eta_j = \rho - \mu_j$$

$$\boldsymbol{\beta} = \sum_{k=1}^{K} \xi_k \mathbf{a}_k - \sum_{i=1}^{n} \mathbf{x}_i (\sum_{l=1}^{L} \lambda_{li}u_{li} + \sum_{h=1}^{H} \gamma_{hi}s_{hi}) + 2\boldsymbol{\mu} - \rho \mathbf{1}_d$$

$$\beta_j = \sum_{k=1}^{K} \xi_k a_{kj} - \sum_{i=1}^{n} x_{ij} (\sum_{l=1}^{L} \lambda_{li}u_{li} + \sum_{h=1}^{H} \gamma_{hi}s_{hi}) + 2\mu_j - \rho$$

3 Negative Dual Problem

$$\mathcal{L}_D = \frac{1}{2}\left(\sum_{k=1}^{K}\sum_{k'=1}^{K}\xi_k\xi_{k'}\mathbf{a}_k^\top\mathbf{a}_{k'} + \sum_{l,i}\sum_{l',i'}\lambda_{li}\lambda_{l'i'}u_{li'}u_{l'i'}\mathbf{x}_i^\top\mathbf{x}_{i'} + \sum_{h,i}\sum_{h',i'}\gamma_{hi}\gamma_{h'i'}s_{hi}s_{h'i'}\mathbf{x}_i^\top\mathbf{x}_{i'}\right)$$
$$- \sum_{k=1}^{K}\sum_{l,i}\xi_k\lambda_{li}u_{li}\mathbf{a}_k^\top\mathbf{x}_i - \sum_{k=1}^{K}\sum_{h,i}\xi_k\gamma_{hi}s_{hi}\mathbf{a}_k^\top\mathbf{x}_i + \sum_{l,i}\sum_{h',i'}\lambda_{li}u_{li}\gamma_{h'i'}s_{h'i'}\mathbf{x}_i^\top\mathbf{x}_{i'}$$
$$+ \frac{1}{2}\sum_{h,i}\gamma_{hi}^2 + \sum_{k=1}^{K}\xi_k b_k - \sum_{l,i}\lambda_{li}v_{li} - \sum_{h,i}\gamma_{hi}t_{hi}$$
$$+ \sum_{k=1}^{K}\xi_k\ (2\boldsymbol{\mu} - \rho\mathbf{1}_d)^\top\mathbf{a}_k + \sum_{l,i}\lambda_{li}u_{li}(\rho\mathbf{1}_d - 2\boldsymbol{\mu})^\top\mathbf{x}_i + \sum_{h,i}\gamma_{hi}s_{hi}(\rho\mathbf{1}_d - 2\boldsymbol{\mu})^\top\mathbf{x}_i$$
$$+ 2\boldsymbol{\mu}^\top\boldsymbol{\mu} - 2\boldsymbol{\mu}^\top\rho\mathbf{1}_d + \frac{d\rho^2}{2}$$

$$\begin{aligned} \textbf{s. t. } & \xi_k \geq 0, \\ & 0 \leq \lambda_{li} \leq 1, \\ & 0 \leq \gamma_{hi} \leq \tau_{hi}, \\ & 0 \leq \mu_d \leq \rho. \end{aligned}$$

Update μ_j and β_j with all other variables fixed:

$$\min_{\mu_j}\ \sum_{k=1}^{K}\xi_k a_{kj}\mu_j - \sum_{i=1}^{n}\sum_{l=1}^{L}\lambda_{li}u_{li}x_{ij}\mu_j - \sum_{i=1}^{n}\sum_{h=1}^{H}\gamma_{hi}s_{hi}x_{ij}\mu_j + \mu_j^2 - \rho\mu_j$$
$$\iff \quad \min_{\mu_j}\ \mu_j^2 + (\sum_{k=1}^{K}\xi_k a_{kj} - \sum_{i=1}^{n}\sum_{l=1}^{L}\lambda_{li}u_{li}x_{ij} - \sum_{i=1}^{n}\sum_{h=1}^{H}\gamma_{hi}s_{hi}x_{ij} - \rho)\mu_j$$

$$\begin{aligned} \nabla_{\mu_j}\mathcal{L}_{\mathcal{D}} &= 2\mu_j + \sum_{k=1}^{K}\xi_k a_{kj} - \sum_{i=1}^{n}\sum_{l=1}^{L}\lambda_{li}u_{li}x_{ij} - \sum_{i=1}^{n}\sum_{h=1}^{H}\gamma_{hi}s_{hi}x_{ij} - \rho \\ &= \beta_j \end{aligned}$$

$$\mu_j^{\mathbf{new}} = \mathcal{P}_{[0,\rho]}\left(\mu_j^{\mathbf{old}} - \frac{\beta_j^{\mathbf{old}}}{2}\right) \quad \overset{\textbf{parallelize}}{\iff} \quad \boldsymbol{\mu}^{\mathbf{new}} = \mathcal{P}_{[\mathbf{0},\rho\mathbf{1}]}\left(\boldsymbol{\mu}^{\mathbf{old}} - \frac{\boldsymbol{\beta}^{\mathbf{old}}}{2}\right)$$
$$\beta_j^{\mathbf{new}} = \beta_j^{\mathbf{old}} + 2(\mu_j^{\mathbf{new}} - \mu_j^{\mathbf{old}}) \quad \overset{\textbf{parallelize}}{\iff} \quad \boldsymbol{\beta}^{\mathbf{new}} = \boldsymbol{\beta}^{\mathbf{old}} + 2(\boldsymbol{\mu}^{\mathbf{new}} - \boldsymbol{\mu}^{\mathbf{old}})$$

Update ξ_k and $\boldsymbol{\beta}$with all other variables fixed:

$$\min_{\xi_k} \frac{1}{2}\mathbf{a}_k^\top \mathbf{a}_k \xi_k^2 + \sum_{k' \neq k} \xi_{k'} \mathbf{a}_{k'}^\top \mathbf{a}_k \xi_k - \sum_{l,i} \lambda_{li} u_{li} \xi_k \mathbf{a}_k^\top \mathbf{x}_i - \sum_{h,i} \gamma_{hi} s_{hi} \xi_k \mathbf{a}_k^\top \mathbf{x}_i + b_k \xi_k + (2\boldsymbol{\mu} - \mathbf{1}_d)^\top \mathbf{a}_k \xi_k$$

$$\begin{aligned}
\nabla_{\xi_k} \mathcal{L}_{\mathcal{D}} = & \ \mathbf{a}_k^\top \mathbf{a}_k \xi_k + \sum_{k' \neq k} \xi_{k'} \mathbf{a}_{k'}^\top \mathbf{a}_k - \sum_{l,i} \lambda_{li} u_{li} \mathbf{a}_k^\top \mathbf{x}_i - \sum_{h,i} \gamma_{hi} s_{hi} \mathbf{a}_k^\top \mathbf{x}_i + b_k + (2\boldsymbol{\mu} - \mathbf{1}_d)^\top \mathbf{a}_k \\
= & \sum_{k'=1}^{K} \xi_{k'} \mathbf{a}_{k'}^\top \mathbf{a}_k - \sum_{l,i} \lambda_{li} u_{li} \mathbf{a}_k^\top \mathbf{x}_i - \sum_{h,i} \gamma_{hi} s_{hi} \mathbf{a}_k^\top \mathbf{x}_i + \mathbf{a}_k^\top (2\boldsymbol{\mu} - \mathbf{1}_d) \ + b_k \\
= & \ \mathbf{a}_k^\top \boldsymbol{\beta} + b_k
\end{aligned}$$

$$\begin{aligned}
\xi_k^{\text{new}} &= \mathcal{P}_{[0,+\infty)} \left(\xi_k^{\text{old}} - \frac{\nabla_{\xi_k} \mathcal{L}(\xi^{\text{old}})}{\|\mathbf{a}_k\|_2^2} \right) = \max \left(0, \xi_k^{\text{old}} - \frac{\mathbf{a}_k^\intercal \boldsymbol{\beta}^{\text{old}} + b_k}{\|\mathbf{a}_k\|_2^2} \right) \\
\boldsymbol{\beta}^{\text{new}} &= \boldsymbol{\beta}^{\text{old}} + (\xi_k^{\text{new}} - \xi_k^{\text{old}}) \mathbf{a}_k.
\end{aligned}$$

Update λ_{li} and $\boldsymbol{\beta}$ with all other variables fixed:

$$\begin{aligned}
\min_{\lambda_{li}} \quad & \frac{1}{2} u_{li}^2 \lambda_{li}^2 \mathbf{x}_i^\top \mathbf{x}_i + \sum_{(l',i') \neq (l,i)} \lambda_{l'i'} u_{l'i'} u_{li} \mathbf{x}_{i'}^\top \mathbf{x}_i \lambda_{li} - \sum_{k=1}^{K} \xi_k u_{li} \mathbf{a}_k^\top \mathbf{x}_i \lambda_{li} + \sum_{h',i'} u_{li} \gamma_{h'i'} s_{h'i'} \mathbf{x}_i^\top \mathbf{x}_{i'} \lambda_{li} \\
& - v_{li} \lambda_{li} + \lambda_{li} u_{li} (\mathbf{1}_d - 2\boldsymbol{\mu})^\top \mathbf{x}_i
\end{aligned}$$

$$\begin{aligned}
\nabla_{\lambda_{li}} \mathcal{L}_{\mathcal{D}} &= u_{li}^2 \lambda_{li} \mathbf{x}_i^\top \mathbf{x}_i + \sum_{(l',i') \neq (l,i)} \lambda_{l'i'} u_{l'i'} u_{li} \mathbf{x}_{i'}^\top \mathbf{x}_i - \sum_{k=1}^{K} \xi_k u_{li} \mathbf{a}_k^\top \mathbf{x}_i + \sum_{h',i'} u_{li} \gamma_{h'i'} s_{h'i'} \mathbf{x}_i^\top \mathbf{x}_{i'} \\
& \quad - v_{li} + u_{li} \mathbf{x}_i^\top (\mathbf{1}_d - 2\boldsymbol{\mu}) \\
&= -u_{li} \mathbf{x}_i^\top \boldsymbol{\beta} - v_{li}
\end{aligned}$$

$$\begin{aligned}
\lambda_{li}^{\text{new}} &= \mathcal{P}_{[0,1]} \left(\lambda_{li}^{\text{old}} - \frac{\nabla_{\lambda_{li}} \mathcal{L}(\lambda^{\text{old}})}{u_{li}^2 \|\mathbf{x}_i\|_2^2} \right) = \max \left(0, \min \left(1, \lambda_{li}^{\text{old}} + \frac{u_{li} \mathbf{x}_i^\intercal \boldsymbol{\beta}^{\text{old}} + v_{li}}{u_{li}^2 \|\mathbf{x}_i\|_2^2} \right) \right) \\
\boldsymbol{\beta}^{\text{new}} &= \boldsymbol{\beta}^{\text{old}} - (\lambda_{li}^{\text{new}} - \lambda_{li}^{\text{old}}) u_{li} \mathbf{x}_i.
\end{aligned}$$

Update γ_{hi} and $\boldsymbol{\beta}$ with all other variables fixed:

$$
\begin{aligned}
\min_{\gamma_{hi}} \;\; & \frac{1}{2}(s_{hi}^2 \mathbf{x}_i^\top \mathbf{x}_i + 1)\gamma_{hi}^2 + \sum_{(h',i') \neq (h,i)} \gamma_{h'i'} s_{h'i'} s_{hi} \mathbf{x}_{i'}^\top \mathbf{x}_i \gamma_{hi} - \sum_{k=1}^{K} \xi_k s_{hi} \mathbf{a}_k^\top \mathbf{x}_i \gamma_{hi} \\
& + \sum_{l',i'} s_{hi} \lambda_{l'i'} u_{l'i'} \mathbf{x}_i^\top \mathbf{x}_{i'} \gamma_{hi} - t_{hi}\gamma_{hi} + \gamma_{hi} s_{hi} (\mathbf{1}_d - 2\boldsymbol{\mu})^\top \mathbf{x}_i
\end{aligned}
$$

$$
\begin{aligned}
\nabla_{\gamma_{hi}} \mathcal{L}_{\mathcal{D}} &= (s_{hi}^2 \mathbf{x}_i^\top \mathbf{x}_i + 1)\gamma_{hi} + \sum_{(h',i') \neq (h,i)} \gamma_{h'i'} s_{h'i'} s_{hi} \mathbf{x}_{i'}^\top \mathbf{x}_i - \sum_{k=1}^{K} \xi_k s_{hi} \mathbf{a}_k^\top \mathbf{x}_i \\
&\qquad + \sum_{l',i'} s_{hi} \lambda_{l'i'} u_{l'i'} \mathbf{x}_i^\top \mathbf{x}_{i'} - t_{hi} + s_{hi} (\mathbf{1}_d - 2\boldsymbol{\mu})^\top \mathbf{x}_i \\
&= \gamma_{hi} + \sum_{h,i} \gamma_{hi} s_{hi} \mathbf{x}_i^\top \mathbf{x}_i - \sum_{k=1}^{K} \xi_k s_{hi} \mathbf{a}_k^\top \mathbf{x}_i + \sum_{l,i} s_{hi} \lambda_{li} u_{li} \mathbf{x}_i^\top \mathbf{x}_i - t_{hi} \\
&\qquad + s_{hi} \mathbf{x}_i^\top (\mathbf{1}_d - 2\boldsymbol{\mu}) \\
&= \gamma_{hi} - s_{hi} \mathbf{x}_i^\top \boldsymbol{\beta} - t_{hi}
\end{aligned}
$$

$$
\begin{aligned}
\gamma_{hi}^{\text{new}} &= \mathcal{P}_{[0,\tau_{hi}]} \left(\gamma_{hi}^{\text{old}} - \frac{\nabla_{\gamma_{hi}} \mathcal{L}(\gamma^{\text{old}})}{s_{hi}^2 \|\mathbf{x}_i\|_2^2 + 1} \right) = \max \left(0, \min \left(\tau_{hi}, \gamma_{hi}^{\text{old}} + \frac{s_{hi} \mathbf{x}_i^\intercal \boldsymbol{\beta}^{\text{old}} + t_{hi} - \gamma_{hi}^{\text{old}}}{s_{hi}^2 \|\mathbf{x}_i\|_2^2 + 1} \right) \right), \\
\boldsymbol{\beta}^{\text{new}} &= \boldsymbol{\beta}^{\text{old}} - (\gamma_{hi}^{\text{new}} - \gamma_{hi}^{\text{old}}) s_{hi} \mathbf{x}_i.
\end{aligned}
\tag{B.11}
$$