1 Introduction

The purpose of this document is to explain some technical details about the PLQ Composite Decomposition toolkit.

Given a general regularized empirical risk minimization (ERM) problem based on a convex piecewise linear-quadratic(PLQ) loss with linear constraints as (1) below.

$$\min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} L_i(\mathbf{x}_i^\intercal \beta) + \frac{1}{2}\|\beta\|_2^2, \qquad \text{s.t. } \mathbf{A}\beta + \mathbf{b} \geq \mathbf{0}, \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^d$ is the feature vector for the i-th observation, $\beta \in \mathbb{R}^d$ is an unknown coefficient vector, $\mathbf{A} \in \mathbb{R}^{\mathbf{K} \times d}$ and $\mathbf{b} \in \mathbb{R}^{\mathbf{K}}$ are defined as linear inequality constraints for β, and $L_i(\cdot) \geq 0$ is a convex PLQ loss function.

Let $z_i = \mathbf{x}_i^\intercal \beta$, then $L_i(z_i)$ is a univariate PLQ function. By Theorem 1 in [1], any convex PLQ loss function can be expressed in the form (2).

$$L(z) = \sum_{l=1}^{L} \mathrm{ReLU}(u_l z + v_l) + \sum_{h=1}^{H} \mathrm{ReHU}_{\tau_h}(s_h z + t_h) \tag{2}$$

where u_l, v_l and s_h, t_h, τ_h are the ReLU-ReHU loss parameters. The ReLU and ReHU functions are defined as

$$\mathrm{ReLU}(z) = \max(z, 0)$$

and

$$\mathrm{ReHU}_\tau(z) = \begin{cases} 0, & z \leq 0 \\ z^2/2, & 0 < z \leq \tau \\ \tau(z - \tau/2), & z > \tau \end{cases}.$$

The core of the PLQ Composite Decomposition toolkit is a two-step method.
1) Decompose a convex PLQ function to form (2)
2) Broadcast (2) to all data

2 Formulation

2.1 Input

PLQ Composite Decomposition accepts three types of input for the PLQ function. One is the coefficients of each piece with cutoffs (3), another is the coefficients only and takes the maximum of each piece (4), the other is the linear version based on a series of given points (5).

$$L(z) = \begin{cases} a_1 z^2 + b_1 z + c_1, & \text{if } z \leq d_1, \\ a_i z^2 + b_i z + c_i, & \text{if } d_{i-1} < z \leq d_i, i = 2, ..., n-1 \\ a_n z^2 + b_n z + c_n, & \text{if } z > d_{n-1}. \end{cases} \tag{3}$$

or

$$L(z) = max\{a_i z^2 + b_i z + c_i\}.i = 1, 2, ..., n \tag{4}$$

or

$$L(z) = \begin{cases} y_1 + \frac{y_2 - y_1}{x_2 - x_1}(z - x_1), & \text{if } z \leq x_1, \\ y_{i-1} + \frac{y_i - y_{i-1}}{x_i - x_{i-1}}(z - x_{i-1}), & \text{if } x_{i-1} < z \leq x_i, i = 2, ..., n \\ y_{n-1} + \frac{y_{n-1} - y_n}{x_{n-1} - x_n}(z - x_n), & \text{if } z > x_n. \end{cases} \tag{5}$$

where $\{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$ is a series of given points and $n \geq 2$

From the formula above, we can know that (5) is a specical case of (3). For (4), we can first find out all the intersection points of each two pieces, denoted as $p_1 < p_2 < ... < p_m$. Then $L(z) = max\{a_i z^2 + b_i z + c_i\}$if $p_j < z \leq p_{j+1}$, which is also a special case of (3). Thus we only discuss $L(z)$ with form (3) in the remaining parts.

2.2 Decompose Stage

In decompose stage, the main task is to convert a single convex PLQ loss function $L(z)$ with form (3) into the form (2)

Without loss of generality, if the minimum value of $L(z)$ is not equal to 0, we can subtract it to make the minimum value equal to zero. Suppose the minimum value was achieved when $z = z^*$ and $L(z) = 0$, $z^* \in \{d_1, ..., d_{n-1}\}$. Separate $L(z)$ from $z = 0$ to $f_l(z)$ and $f_r(z)$ below.

$$f_l(z) = \begin{cases} L(z), & \text{if } z \leq z^* \\ 0, & \text{if } z > z^* \end{cases} \qquad f_r(z) = \begin{cases} 0, & \text{if } z < z^* \\ L(z), & \text{if } z \geq z^* \end{cases}$$

We can know $L(z) = f_l(z) + f_r(z), z \in \mathbb{R}$ and $f_l(z)$,$f_r(z)$ are also PLQ functions. In the following, we decompose $f_l(z)$,$f_r(z)$ into ReLU-ReHU piece by piece.

For $f_r(z)$, work from each cut point from left to right piece by piece. Subtract the left tangent line from the cutpoint, and then we have

$$\begin{aligned} g_{r,i}(z) &= f_{r,i}(z) - \left[f'_{r,i-1}(d_{i-1})(z - d_{i-1}) + f_{r,i-1}(d_{i-1})\right] \\ &= a_i z^2 + b_i z + c_i - \left[(2a_{i-1}d_{i-1} + b_{i-1})(z - d_{i-1}) + a_i d_{i-1}^2 + b_i d_{i-1} + c_i\right] \\ &= \frac{(\sqrt{2a_i}z - \sqrt{2a_i}d_{i-1})^2}{2} + \left[2d_{i-1}(a_i - a_{i-1}) + (b_i - b_{i-1})\right](z - d_{i-1}) \\ &= \text{ReHU}_\tau(sz + t) + \text{ReLU}(uz + v) \\ \tau &= \sqrt{2a_i(d_i - d_{i-1})},\ s = \sqrt{2a_i},\ t = -d_{i-1}s, \\ u &= 2d_{i-1}(a_i - a_{i-1}) + (b_i - b_{i-1}),\ v = -d_{i-1}u \end{aligned} \tag{6}$$

where d_{i-1} are the $(i-1)^{th}$ cutpoint, $a_i, b_i, c_i, a_{i-1}, b_{i-1}, c_{i-1}$ are the coefficients of the piece i and $i - 1$, then

$$f_r(z) = \sum_{i=1}^{n_r} g_{r,i}(z) = \sum_{l=1}^{L_r} \text{ReLU}(u_l z + v_l) + \sum_{h=1}^{H_r} \text{ReHU}_{\tau_h}(s_h z + t_h)$$

Similarly, we can get the result for $f_l(z)$

$$f_l(z) = \sum_{i=1}^{n_l} g_{l,i}(z) = \sum_{l=1}^{L_l} \mathrm{ReLU}(u_l z + v_l) + \sum_{h=1}^{H_l} \mathrm{ReHU}_{\tau_h}(s_h z + t_h)$$

Add $f_l(z)$ and $f_r(z)$ together, we will get a ReLU-ReHU form of $L(z)$ like (2)

2.3 Broadcast Stage

In broadcast stage, the main task is to broadcast $L(z)$ in the form (2) in the decomposition stage to all data. i.e. generate $L_i(z_i)$ from the $L(z)$ above.

Usually, there exists a special relationship (If relationship not holds, you have to mannully do the decomposition stage n times)

$$L_i(z_i) = c_i L(p_i z_i + q_i) \tag{7}$$

By Proposition 1 in [1] the composite (ReLU-ReHU) function class is closed under affine transformations.

Proposition 1 (Closure under affine transformation). If $L(z)$ is a composite (ReLU-ReHU) function as in (ReLU-ReHU), then for any $c > 0,\ p \in \mathbb{R},\ and\ q \in \mathbb{R},\ cL(pz + q)$ it is also a composite (ReLU-ReHU) function, that is,

$$cL(pz+q) = \sum_{l=1}^{L} \mathrm{ReLU}(u'_l z + v'_l) + \sum_{h=1}^{H} \mathrm{ReHU}_{\tau'_h}(s'_h z + t'_h), \tag{8}$$

where $u'_l = cpu_l,\ v'_l = cu_l q + cv_l,\ \tau'_h = \sqrt{c}\tau_h,\ s'_h = \sqrt{c}ps_h,\ and\ t'_h = \sqrt{c}(s_h q + t_h).$
Combine (7) and (8), then

$$\begin{aligned} L_i(z_i) &= c_i L(p_i z_i + q_i) \\ &= \sum_{l=1}^{L} \mathrm{ReLU}(u'_{li} z_i + v'_{li}) + \sum_{h=1}^{H} \mathrm{ReHU}_{\tau'_{hi}}(s'_{hi} z_i + t'_{hi}), \end{aligned} \tag{9}$$

Substitute (9) into (1) then

$$\min_{\beta \in \mathbb{R}^d} \sum_{i=1}^{n} \sum_{l=1}^{L} \mathrm{ReLU}(u_{li} \mathbf{x}_i^\intercal \beta + v_{li}) + \sum_{i=1}^{n} \sum_{h=1}^{H} \mathrm{ReHU}_{\tau_{hi}}(s_{hi} \mathbf{x}_i^\intercal \beta + t_{hi}) + \frac{1}{2}\|\beta\|_2^2, \quad \text{s.t. } \mathbf{A}\beta + \mathbf{b} \geq \mathbf{0}, \tag{10}$$

where $\mathbf{U} = (u_{li}), \mathbf{V} = (v_{li}) \in \mathbb{R}^{L \times n}$ and $\mathbf{S} = (s_{hi}), \mathbf{T} = (t_{hi}), \tau = (\tau_{hi}) \in \mathbb{R}^{H \times n}$ are the ReLU-ReHU loss parameters, and $(\mathbf{A}, \mathbf{b})$ are the constraint parameters.

With the ReLU-ReHU loss and constraint parameters above, we can utilize the rehline package to solve the ERM problem.

References

[1] Dai, B., & Qiu, Y. (2023). ReHLine: Regularized Composite ReLU-ReHU Loss Minimization with Linear Computation and Linear Convergence. In *the Thirty-Seventh Conference on Neural Information Processing Systems*.