Failure of Coordinate Descent in Quantile Regression: An Illustrative Example

Swapnamoy Samadder

Abstract

Coordinate descent is a well-known and widely used algorithm for optimization, valued for its scalability and faster rate of convergence compared to standard algorithms such as simplex or majorization–minimization. However, a serious issue lies in its lack of guaranteed convergence. It is well known (see Powell) that coordinate descent may fail due to cycling.

In the context of quantile regression on $\mathbb{R}^2$, the check-loss function

$$\rho_\tau(t) = \tau|t|\mathbb{I}(t \geq 0) + (1-\tau)|t|\mathbb{I}(t < 0)$$

is convex. Thus, coordinate descent is often applied without hesitation.

In this paper, we find a region $R \subset \mathbb{R}^2$with infinite lebesgue measure, such that if $\beta^{(0)} \in R$, coordinate descent fails to find the global minimum of the check-loss function for given $(Z_n, (x_n, y_n))$, $n \in \{1, \dots, k\}$.

Introduction: Coordinate Descent and Convergence Issues in Quantile Regression

Coordinate descent is a standard algorithm used to find global optima of functions—particularly convex functions that often exhibit favorable properties ensuring convergence to a global minimum under certain regularity conditions (see Tseng, 2001).

In $\mathbb{R}^2$, the algorithm proceeds as follows: Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a convex (or pseudoconvex) function, bounded below. Choose initial values

$$\beta^{(0)} = (\beta_1^{(0)}, \beta_2^{(0)}).$$

At iteration $n-1$,

$$\beta^{(n-1)} = (\beta_1^{(n-1)}, \beta_2^{(n-1)}).$$

To obtain the next iterate:

$$\beta_1^{(n)} = \arg\min_\beta f(\beta, \beta_2^{(n-1)}),$$

$$\beta_2^{(n)} = \arg\min_\beta f(\beta_1^{(n)}, \beta),$$

$$\Rightarrow \beta^{(n)} = \left(\beta_1^{(n)}, \beta_2^{(n)}\right).$$

The iterations continue until a stopping criterion is met.

Assumptions

(i) $\{(x_n, y_n)\}_{n=1}^k$ are all not parallel, i.e.,

$$\{(x_n, y_n)\}_{n=1}^k \subset \mathbb{R}^2 \setminus \{(0, x) \cup (x, 0)\}.$$

Without loss of generality, define the objective function:

$$f(\beta_1, \beta_2) = \sum_{n=1}^{k-1} |Z_n - x_n\beta_1 - y_n\beta_2| + |\beta_1 - \beta_2|.$$

Assume that after transformation:

$$x \to c_1 x, \qquad y \to c_2 y + d,$$

and that

$$\min(x_1, \ldots, x_{k-1}) \geq 0.$$

Then, after transformation:

$$\sum_{n=1}^{k-1} x_n + 1 > 0, \qquad \sum_{n=1}^{k-1} y_n + 1 > 0,$$

but

$$-\left(\sum_{n=1}^{k-1} x_n + \sum_{n=1}^{k-1} y_n\right) < 0.$$

Lemma 1

Statement:

$$\exists M \in \mathbb{R}^+ : \text{if } t > M, \text{ then } (t,t) \text{ is an irregular point of } f(\beta_1, \beta_2).$$

Proof:

$$\min(x_1, \ldots, x_{k-1}, y_1, \ldots, y_{k-1}) \geq 0, \quad \max(x_1, \ldots, x_{k-1}) > 0.$$

$$\Rightarrow \exists M \in \mathbb{R} : \forall t > M,\ Z_n < x_n t + y_n t \quad \forall n = 1, \ldots, k-1.$$

Choose $t > M$ and $\delta > 0$. Then:

$$f(t+\delta, t) = \sum_{n=1}^{k-1} |Z_n - x_n(t+\delta) - y_n t| = f(t,t) + (1 + \sum x_n)\delta.$$

Thus,

$$f^+((t,t), e_1) = \lim_{\delta \to 0^+} \frac{f(t+\delta, t) - f(t,t)}{\delta} = \sum x_n + 1 > 0,$$

$$f^-((t,t), e_1) = \sum x_n > 0.$$

Similarly,

$$f^+((t,t), e_2) = \sum y_n + 1 > 0, \quad f^-((t,t), e_2) = \sum y_n + 1 > 0.$$

Continuing,

$$f(t+\delta, t+\delta) = f(t,t) - (\sum x_n + \sum y_n)\delta.$$

Hence,

$$f^-((t,t), e_1 + e_2) = -(\sum x_n + \sum y_n) < 0.$$

Therefore, (t,t) is an irregular point if $t > M$. ■

Lemma 2

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be convex. Fix $(x_0, y_0) \in \mathbb{R}^2$, and define

$$f_1(x) = f(x, y_0), \qquad f_2(y) = f(x_0, y).$$

Then both f_1 and f_2 are convex functions $\mathbb{R} \to \mathbb{R}$.

Proof:

$$f_1(\theta x_1 + (1-\theta)x_2) = f(\theta x_1 + (1-\theta)x_2, y_0) \leq \theta f(x_1, y_0) + (1-\theta) f(x_2, y_0) = \theta f_1(x_1) + (1-\theta) f_1(x_2).$$

■

Lemma 3

Let $(t_1, t_2) \in \mathbb{R}^2$ with $\min(t_1, t_2) > M$. Let $\beta^{(0)} = (t_1, t_2)$. If the assumptions are satisfied, then:

- $f(t_1, t_2)$ and $f(t, t)$ are strictly convex, and - each has a unique coordinate-wise minimum at t_2 and t_1, respectively.

Proof: From Lemma 1,

$$f_1(t) = f(t, t_2), \quad f_1^+(t_2) = f^+((t_2, t_2), e_1) = \sum x_n + 1 > 0, \quad f_1^-(t_2) = \sum x_n > 0.$$

Hence t_2 is a unique local (and thus global) minimum of f_1. Similarly,

$$\arg\min_t f(t_1, t) = \{t_1\}.$$

■

Theorem

If $\beta^{(0)}$ satisfies the conditions of Lemma 3 and the *assumptions* hold, then coordinate descent from any direction will converge to an **irregular point**, which is **not** the global minimum of f.

Proof: A minimum is by definition a regular point since

$$f(m, d) \geq 0 \quad \forall d \in \mathbb{R}^2.$$

Now, if

$$\beta^{(0)} = (t_1, t_2),$$

then by Lemma 3,

$$\beta_1^{(1)} = \arg\min_t f(t, t_2) = \{t_2\}, \quad \beta_2^{(1)} = \arg\min_t f(t_2, t) = \{t_2\}.$$

Therefore,

$$\beta^{(n)} = (t_2, t_2), \quad \forall n \geq 1.$$

But (t_2, t_2) is irregular by Lemma 1. Hence, coordinate descent converges to a non-optimal irregular point.

■