A Log-Linear Time Sequential Optimal Calibration Algorithm for Quantized Isotonic L2 Regression

Kaan Gokcesu, Hakan Gokcesu

Abstract—We study the sequential calibration of estimations in a quantized isotonic L2 regression setting. We start by showing that the optimal calibrated quantized estimations can be acquired from the traditional isotonic L2 regression solution. We modify the traditional PAVA algorithm to create calibrators for both batch and sequential optimization of the quantized isotonic regression problem. Our algorithm can update the optimal quantized monotone mapping for the samples observed so far in linear space and logarithmic time per new unordered sample.

I. INTRODUCTION

The minimization of certain error or loss functions are heavily studied in the problems of machine learning, prediction and estimation [11–13]: especially in the applications pertaining to control theory [4], optimization [5], [6], scheduling [7], forecasting [8], decision theory [10], density estimation and anomaly detection [11–16], bandits [17–19], signal processing [20, 21] and game theory [22, 23]. In most of the applications, a learner or decision maker produces its estimates in the form of some score. Even though, for some applications, the ordinality implied by these estimates are sufficient; calibration is needed for increased accuracy, which is generally achieved with post-processing calibrators [24, 25]. Although calibrated outputs generally perform better, its unlikely to achieve perfect matching [26]. Nonetheless, well-calibrated predictions are important in many heavily studied fields such as business and medicine. These well-calibrated estimates are used not only to increase decision making performance but also for more accurate evaluation of learning models [27, 28] and even their more effective ensemble [29]. While lacking in comparison to the effort in creating models that utilize some regularization (against over-fit) [30].

The most straightforward calibration technique regularizes the output estimations by mapping the inputs using an optimized parametric function (e.g., a step function with optimized threshold for binary classification [30]). The most widely used technique involves a sigmoidal fit [31], where the input estimations are passed through a logistic function with appropriately tuned parameters [32]. It has been empirically shown that this parametric technique performs satisfactorily despite its efficiency in contrast to training another learning model [31]. It has been successfully utilized in the calibration of many different learning algorithms [33], [44] in addition to the original study on SVM [31]. However, in another empirical study, it has also been shown that the sigmoidal fitting technique may under-perform for different learning models [35]. The unsatisfactory performance arises because of the strong restriction imposed by the parametric shape [36]. The restrictiveness of the parametric calibration is relaxed by the binning approach [37], [38], which starts by partitioning the ordered input estimations in bins. The output estimations are created from the respective bins individually, which imposes the regularization of mapping similar estimations to similar outputs. Because of its less restrictive nature in contrast to the sigmoidal fitting, its parameters (e.g., the bin sizes and locations) are harder to optimize [35]. Incorrect determination of the bin parameters may lead to ineffective learning. To efficiently address this issue, confidence intervals are utilized to create the bins [36]. To address the inherent limitations of the binning approach, ensemble methods that use Bayesian mixture [39] are studied [26, 40]. Although the relaxed restriction of binning provides more freedom in the design, it may lead to over-fitting.

In the field of calibration, the inherent drawbacks of the sigmoid fitting and quantile binning approaches are addressed by the technique of isotonic regression [35], [41], [42], whose regularity originates from the monotonicity on the mapping, i.e., the calibrated estimations share the same ordinality with the input estimations. This regularity results in the mapping of several adjacent samples to the same calibrated value (hence the name isotonic). The number of samples mapped to a single value increases as the input estimates are incorrectly ranked (bad discriminative performance). From the perspective of restrictiveness, it is between sigmoid fitting and binning; and shares certain similarities with them [30, 35]. Many algorithms were designed to create isotonic regressors [43], [44]. It has been utilized to help quantile binning strategies [45, 46]; and to aggregate distinct input estimates [47].

Although the isotonic regression addresses many issues with the sigmoid fitting and quantile binning, it is traditionally hard to update [30]. Even though a sequential algorithm for isotonic mapping can have up to quadratic time complexity in general [48], we show that when focused on the problem of quantized regression, a log-linear complexity algorithm exists. To this end, we propose an efficient sequential algorithm with logarithmic time and linear space complexity for the problem of quantized isotonic L2 regression.

In Section II we formally define the quantized isotonic L2 regression. In Section III we prove that an optimal quantized monotone mapping can be extracted from an optimal unconstrained monotone mapping. In Section IV we design an algorithm that can sequentially update the optimal quantized monotone mapping in logarithmic time with each new sample when the scores arrive in order. In Section V we propose a sequential algorithm that can update the optimal quantized monotone mapping in logarithmic time with each new sample even when the samples scores are observed out of order.
II. PROBLEM DEFINITION

For $N$ number of samples indexed by $n \in \{1, 2, \ldots, N\}$, we have the target variables $y_n$ where

$$y_n \in \mathbb{R},$$

where $\mathbb{R}$ is the set of reals. For every $y_n$, we have unordered input estimations $x_n$ (i.e., the score of the $n^{th}$ sample) where

$$x_n \in \mathbb{R}.$$  

Moreover, we have the positive sample weights $\alpha_n$ where

$$\alpha_n > 0.$$  

Thus, the initial uncalibrated loss is given by

$$\sum_{n=1}^{N} \alpha_n (y_n - x_n)^2.$$  

In the isotonic regression problem, the traditional goal is to create calibrated estimations $z_n \in \mathbb{R}$ to decrease the weighted square error, where

$$z_n = C(x_n),$$

for some monotone transform $C(\cdot)$ such that

$$C(x_n) \leq C(x_{n'}) \quad \text{if} \quad x_n \leq x_{n'} \quad \forall n, n'.$$  

**Definition 1.** The objective of the traditional isotonic $L^2$ regression problem is given by

$$\arg\min_{C(\cdot) \in \Omega} \sum_{n=1}^{N} \alpha_n (y_n - C(x_n))^2,$$

where $\Omega$ is the class of all monotonically nondecreasing functions that map from $\mathbb{R}$ to $\mathbb{R}$.

In this paper, we study the quantized version of this isotonic regression problem, where

$$z_n = C_Q(x_n) \in \mathbb{Q},$$

for some quantized subset $\mathbb{Q} \subset \mathbb{R}$.

**Definition 2.** The revised objective of the quantized isotonic $L^2$ regression is given by

$$\arg\min_{C_Q(\cdot) \in \Omega_Q} \sum_{n=1}^{N} \alpha_n (y_n - C_Q(x_n))^2,$$

where $\Omega_Q$ is the class of all monotonically nondecreasing functions that map from $\mathbb{R}$ to a quantization subset $\mathbb{Q}$.

Let an optimizer for the objective in **Definition 2** be $C_N(\cdot)$ for $N$ number of samples. Our goal is to sequentially create these optimizers in an efficient manner. After we acquire $C_N(\cdot)$, we want to update the mapping and create $C_{N+1}(\cdot)$ using the new sample $x_{N+1}$. We will show that we can update the mapping in a time complexity that is logarithmic in the number of samples observed so far.

III. FROM AN OPTIMAL MONOTONE MAPPING TO AN OPTIMAL QUANTIZED MONOTONE MAPPING

Let us consider that the sample scores $x_n$ are ordered, i.e., $x_n \leq x_{n+1}$ for all $n$ (if not we can just order them). Let us have a minimizer $C_N(\cdot)$ for **Definition 2** given by

$$C_N(x_n) = z_n, \quad n \in \{1, \ldots, N\},$$

where $z_n$ be the corresponding mapped value.

A. Optimal Quantized Monotone Mapping

**Lemma 1.** There exists a minimizer $C_N(\cdot)$ for **Definition 2** such that

$$z_n = z_{n+1} + 1 \quad \text{if} \quad y_n \geq y_{n+1}.$$  

**Proof.** Let us have an initial quantized monotone mapping $\{z'_n\}_{n=1}^{N}$, where $z'_n \in \mathbb{Q}$ and $z'_n \leq z'_{n+1}$ for all $n$. Let the loss of interest be $L_n(z'_n, z_{n+1}) = \alpha_n (z'_n - y_n)^2 + \alpha_{n+1} (z'_{n+1} - y_{n+1})^2$. Given $y_n \geq y_{n+1}$, we have the following six cases:

1. $z'_n \leq z'_{n+1} \leq y_n$: $L_n(z'_n, z'_{n+1}) \leq L_n(z'_n, z_{n+1})$.
2. $z'_n \leq y_{n+1} \leq z'_{n+1} \leq y_n$: $L_n(z'_n, z'_{n+1}) \leq L_n(z'_n, z'_{n+1})$.
3. $z'_n \leq y_{n+1} \leq z'_{n+1} \leq y_n$: $L_n(z'_n, z'_{n+1}) \leq L_n(z'_n, z'_{n+1})$.
4. $y_{n+1} \leq z'_{n+1}$: $L_n(z'_n, z_{n+1}) \leq L_n(z'_n, z'_{n+1})$.
5. $y_{n+1} \leq z'_{n+1} \leq y_n$: $L_n(z'_n, z_{n+1}) \leq L_n(z'_n, z'_{n+1})$.
6. $y_{n+1} \leq y_n$: $L_n(z'_n, z_{n+1}) \leq L_n(z'_n, z'_{n+1})$.

Therefore, there exists a minimizer $C_N(\cdot)$ such that $z_n = z_{n+1}$ if $y_n \geq y_{n+1}$, which concludes the proof.

Hence, there are groups of samples with the same quantized mapping $z_n \in \mathbb{Q}$. Following [30, 48], let there be $I$ such groups, where the group $i \in \{1, \ldots, I\}$ consists of the samples $n \in \mathcal{N}_i = \{n_i + 1, \ldots, n_{i+1}\}$ (with $n_1 = 0, n_{I+1} = N$), i.e., for each group $i$, we have $z_n = z'_n, \forall n', n' \in \mathcal{N}_i$. With an abuse of notation, let us denote the mapping of the $i^{th}$ group with $z_i$. We have the following result.

**Lemma 2.** If $C_N(\cdot)$ is a minimizer for **Definition 2**, then

$$z_i = z_{i+1} + 1 \quad \text{if} \quad \tilde{y}_i \geq \tilde{y}_{i+1},$$

where $\tilde{y}_i$ is the weighted average of the target variables $y_n$ of group $i$, i.e.,

$$\tilde{y}_i = \frac{\sum_{n=n_i+1}^{n_{i+1}} \alpha_n y_n}{\sum_{n=n_i+1}^{n_{i+1}} \alpha_n}.$$  

Consequently, we have

$$z_i < z_{i+1} + 1 \quad \text{if} \quad \tilde{y}_i < \tilde{y}_{i+1},$$

i.e., the $i^{th}$ groups mapping $z_i$ is distinct (consequently strictly less) than $z_{i+1}$ only if $\tilde{y}_i$ is strictly less than $\tilde{y}_{i+1}$.

**Proof.** The proof follows similar arguments with the proof of **Lemma 1**.  

B. Achieving Optimal Quantized Monotone Mapping

**Proposition 1.** We have the following quantization result
\[
\arg \min_{z \in \mathcal{Q}} \sum_{n=n_1+1}^{n_N+1} \alpha_n (z - y_n)^2 = \arg \min_{z \in \mathcal{Q}} \left| z - \frac{\sum_{n=n_1+1}^{n_N+1} \alpha_n y_n}{\sum_{n=n_1+1}^{n_N+1} \alpha_n} \right|
\]
**Proof.** Let \( \tilde{y}_i \) be the weighted average. We have
\[
\sum_{n=n_1+1}^{n_N+1} \alpha_n (z - y_n)^2 = \sum_{n=n_1+1}^{n_N+1} \alpha_n (z - \tilde{y}_i + \tilde{y}_i - y_n)^2
\]
\[= \sum_{n=n_1+1}^{n_N+1} \alpha_n \left[ (z - \tilde{y}_i)^2 + (\tilde{y}_i - y_n)^2 \right].\]
Hence,
\[
\arg \min_{z \in \mathcal{Q}} \sum_{n=n_1+1}^{n_N+1} \alpha_n (z - y_n)^2 = \arg \min_{z \in \mathcal{Q}} \sum_{n=n_1+1}^{n_N+1} \alpha_n (z - \tilde{y}_i)^2
\]
\[= \arg \min_{z \in \mathcal{Q}} |z - \tilde{y}_i|^2 \]
which concludes the proof.

**Proposition 2.** We have
\[
\arg \min_{z \in \mathcal{Q}} |z - \tilde{y}_i| \leq \arg \min_{z \in \mathcal{Q}} |z - \tilde{y}_j| \quad \text{if} \quad \tilde{y}_i \leq \tilde{y}_j.
\]
**Proof.** The proof comes from the triangle inequality.

**Theorem 1.** There exists a minimizer \( C_N(\cdot) \) for Definition 2 with \( I \) groups such that (with an abuse of notation on \( C_N(\cdot) \))
\[
C_N(\{x_n\}_{n=n_1+1}^{n_N+1}) = \arg \min_{z \in \mathcal{Q}} |z - \tilde{y}_i|, \quad \tilde{y}_i \leq \tilde{y}_{i+1}, \forall i.
\]
where \( \tilde{y}_i \) are the weighted group averages.

**Proof.** The proof comes from [Lemma 2] [Proposition 1] and [Proposition 2].

C. Linear Time Batch Algorithm

We have shown that an optimal quantized monotone mapping can be derived from an optimal monotone mapping. Hence, we can utilize the following linear time algorithm.

1) Let us have \( N \) samples with the scores \( \{x_n\}_{n=1}^{N} \), where \( x_n \) are ordered.
2) Create a group for each sample, i.e., \( n \in \{1, \ldots, N\} \).
   Set the initial minimizers \( \tilde{y}_n = y_n \) and the group weights \( \alpha_n = \alpha_n \) for all \( n \in \{1, \ldots, N\} \).
3) Let \( \mathcal{I} \) be the set of all indices \( i \) such that the consecutive group pair \( (i, i+1) \) have \( \tilde{y}_i \geq \tilde{y}_{i+1} \)
4) IF \( \mathcal{I} \) is not empty, join the groups \( i \) and \( i+1 \) for all \( i \in \mathcal{I} \).
   IF there is consecutive grouping, join all of them;
   e.g., if the pairs \( \{i, i+1\}, \{i+1, i+2\} \) are joined, we join \( \{i, i+1, i+2\} \).
   When combining the groups \( \{i, i+1, \ldots, i+j\} \), we have the new group minimizer \( \tilde{y} = \sum_{i'=i}^{i+j} \alpha_{i'} \tilde{y}_{i'} / \sum_{i'=i}^{i+j} \alpha_{i'} \) and weight \( \alpha = \sum_{i'=i}^{i+j} \alpha_{i'} \).
   Re-index the groups accordingly and return to Step 3.
5) ELSE (if \( \mathcal{I} \) is empty), project all the group minimizers \( \tilde{y}_i \) on to the set \( \mathcal{Q} \) to create an optimal quantized mapping.

**Remark 1.** It takes \( O(1) \) time to update the weighted means and the weight sums per group involved in the joining. Thus, this approach can find an optimal transform in linear space and linear time.

IV. SEQUENTIALIZING ISOTONIC L2 REGRESSION

We first consider the easier version of the sequential problem. Let the samples come in an ordered fashion such that \( x_{n+1} \geq x_n \), where the goal is to sequentially update an optimal mapping \( C_N(\cdot) \) with the new sample \( x_{N+1} \).

An optimal mapping for the first \( N \) samples are given by
\[
\sum_{n=1}^{N} \alpha_n (C_N(x_n) - y_n)^2 = \min_{C(\cdot) \in \mathcal{Q}} \sum_{n=1}^{N} \alpha_n (C(x_n) - y_n)^2,
\]
To create \( C_{N+1}(\cdot) \) using \( x_{N+1} \), we can utilize the standard batch algorithm in [Section III-C], since the joining order of the groups does not matter. The sequential algorithm is as follows:

1) We have \( N = 1, x_1 \), the minimizer \( \tilde{y}_1 = y_1 \), the weight \( \tilde{\alpha}_1 = \alpha_1 \) and the mapping \( \tilde{z}_1 = \arg \min_{z \in \mathcal{Q}} |z - \tilde{y}_1| \).
2) Let the number of groups be \( I = 1 \). Receive \( x_{N+1} \); set \( \tilde{y}_1 = y_{N+1}, \tilde{\alpha}_I = \alpha_{N+1}, \tilde{z}_I = \arg \min_{z \in \mathcal{Q}} |z - \tilde{y}_1| \).
3) WHILE the group with the largest index (denoted as \( I \)) needs to be joined with its immediate preceding group, join them. Set \( \tilde{y} = (\tilde{\alpha}_{I-1} \tilde{y}_{I-1} + \tilde{\alpha}_I \tilde{y}_I) / (\tilde{\alpha}_{I-1} + \tilde{\alpha}_I) \), \( \tilde{z} = \arg \min_{z \in \mathcal{Q}} |z - \tilde{y}| \) and \( \tilde{\alpha} = \tilde{\alpha}_{I-1} + \tilde{\alpha}_I \). Re-index the new group with the preceding group’s index \( I - 1 \).
4) Set \( N \leftarrow N + 1 \) and return to Step 2.

For a given time \( T \), let \( I_T \) be the number of groups in the mapping. Since with each joining the number of groups decreases by 1, we will have exactly \( T - I_T \) joining in total. Since each joining takes \( O(1) \) time, the worst-case time complexity is \( O(T) \).

Note that while the cumulative time complexity of this approach is \( O(T) \), the per sample time complexity by itself is also \( O(T) \) since the decrease in the number of steps is only bounded by the number of groups. Hence, we can find an optimal mapping in linear space and linear amortized time. To limit the per sample time complexity, we can utilize the quantized structure of the mapping.

**Lemma 3.** If two consecutive groups have the same quantized mapping, they will have so in the future.

**Proof.** Let the groups \( i \) and \( i + 1 \) have the group minimizers \( \tilde{y}_i, \tilde{y}_{i+1} \) together with their quantizations \( \tilde{z}_i \) and \( \tilde{z}_{i+1} \). For the quantized mappings to change, their minimizers should change. For \( \tilde{y}_i \) to change, \( \tilde{y}_{i+1} \) should decrease. If the change in \( \tilde{y}_{i+1} \) is not significant enough, \( \tilde{z}_{i+1} \) will not change. However, if it is so, \( \tilde{y}_{i+1} \) will decrease beyond \( \tilde{y}_i \) and the \( \{i, i+1\} \)th group will need to be combined with the \( i^{th} \) group. Hence, their minimizers and their respective quantized mapping will be the same. Thus, if the quantized mappings of two consecutive groups are the same, it will always be so.

**Theorem 2.** We can combine a consecutive group pair \( \{i, i+1\} \) only not when \( \tilde{y}_i \geq \tilde{y}_{i+1} \) but also \( \tilde{z}_i = \tilde{z}_{i+1} \).

**Proof.** The proof comes from [Lemma 3].

By utilizing [Theorem 2], we can also consider the consecutive groups with the same quantized mappings to be joined. This will substantially decrease the per time complexity, since the number of groups will be limited by the quantization. If the mapped set \( \mathcal{Q} \) is bounded it will be constant per time, otherwise the complexity will be pseudo-linear since it will depend on the observed samples so far.
V. A Log-Linear Time Truly Sequential Algorithm

To create the mappings \( C_N(\cdot) \) when the sample scores arrive out of order, we propose a recursive algorithm that iteratively merges sets. Our algorithm has a sequential implementation with logarithmic time complexity per sample. We first provide a batch optimization version to increase comprehension.

At the first stage (i.e., the bottom level), for \( n \in \{1, \ldots, N\} \), we have the mapping set \( A_1^n = \{Y_1^n, Z_1^n\} \) where \( Y_1^n = \{y_n\} \) are the target variables, \( Z_1^n = \{z_n\} \) are the optimal quantized mappings and the indices \( n \) imply by the ordinality of the scores \( X = \{x_n\}_{n=1}^N \). Starting from the bottom level \( k = 1 \) (initial stage), at every level \( k \), we create the sets \( A^{k+1}_i \) by merging the adjacent relevant sets at the \( k^{th} \) level. Whenever the sets \( A_i^{k+1} = \{Y_i^{k+1}, Z_i^{k+1}\} \) are merged together, we have the new target variable set \( Y_i^{k+1} = Y_i^k \cup Y_i^{k+1} \) and the new quantized mappings \( Z_i^{k+1} \), which can be calculated in time linear with the cardinality \( |Y_i^{k+1}| \). If there is no other set that can be merged with \( A_i^k \), it is moved up a level, i.e., \( A_i^{k+1} = A_i^k \). Note that \( i \) may not be twice of \( j \) or \( j \), and the variables at their respective levels. Since the combinations are done via a binary tree, the total time complexity is \( O(N\log N) \) and the space complexity is linear \( O(N) \).

When the samples sequentially arrive in an unordered manner, we can run this merging algorithm from the stretch with each new sample to create an optimal quantized mapping. However, such an implementation would be inefficient and take linear time complexity per sample. To this end, we modify this binary merging structure to create an efficient sequential algorithm. First of all, to efficiently combine the mappings of different sets, we utilize auxiliary variables. We modify the set \( A_i^k \) associated with the index \( i \) at the level \( k \) as the following:

\[
A_i^k = \{X_i^k, Y_i^k, W_i^k\}
\]

where \( |X_i^k| = |Y_i^k| = |W_i^k| = I_{k,i} \) is the number of groups in the mapping of \((k, i)\); \( X_i^k \) are the smallest scores in each group, \( Y_i^k \) are the largest scores in each group, \( W_i^k \) are the group minimizers and \( W_i^k \) are the total group weights.

1. At any point in the algorithm, whenever a new sample comes, we only update the necessary intermediate sets.
2. Schedule a new set \( A_m \) to move up as itself to the next level if it does not fall between a pair of already-merged sets at a level \( k \).
3. If a new set \( A_m \) (middle) falls between a pair of already-merged sets \( A_l \) (left) and \( A_r \) (right) at a level \( k \), combine \( A_l \) with \( A_m \) instead of \( A_r \) (left-biased), i.e., the preceding merger of \( A_l \) with \( A_r \) is updated with the new merger of \( A_l \) with \( A_m \). Schedule \( A_r \) to move up as itself to the next level.
4. If a scheduled-to-move-up set is adjacent to an already-moved-up set at any level \( k \), update the moved-up set with their merger at the next level. If not, move up scheduled-to-move-up set as itself to the next level \( k+1 \).
5. Whenever a set in the recursive structure is updated (a new or updated merger) at a level \( k \), update the subsequent mergers at the next level \( k + 1 \).

A. Complexity analysis

**Lemma 4.** The depth of the recursion tree is logarithmic in the number of samples \( N \), i.e., \( D = O(\log N) \) (where \( k \leq D \), \( \forall k \)), and has size \( O(N) \).

**Proof.** The structure of the sequential algorithm forces any two adjacent not-combined sets at any level to be combined at the next level. Let the number of sets that exists at the level \( k \) be \( N_k \), then the number of sets at the level \( k + 1 \) is bounded as \( N_k+1 \leq \frac{2}{\xi}(N_k+2) \) and concludes the proof.

**Lemma 5.** The number of sets that are updated with each sample is logarithmic in the number of samples \( N \), i.e., \( O(\log N) \).

**Proof.** An updated set at the level \( k + 1 \) will result from a preceding connected set update at the level \( k \). For any new sample, we will have at most two updates at each level and the updates in different levels are connected with chains. Since the depth of the recursion is logarithmic, the number of updates are logarithmic in \( N \), which concludes the proof.

**Lemma 6.** Let us have two adjacent sets \( A^0 \) and \( A^1 \), where \( A^0 = \{X_0^1, Y_0^1, W_0^1\} \) and \( A^1 = \{X_1^1, Y_1^1, W_1^1\} \) with the corresponding number of groups \( I_0 \) and \( I_1 \). The combination of \( A^0 \) and \( A^1 \) denoted by \( A = \{X_0, Y_1, W\} \) with the corresponding number of groups I can be calculated in \( O(I_1 + I_0 - I) \) time complexity.

**Proof.** Because of the monotone mapping structure \( I \leq I_0 + I_1 \), By utilizing the iterative pairwise group combination in the algorithm of Section IV with the group minimizers and the group weights, we can decrease the number of groups in the combination by 1 in \( O(1) \) time. Note that we not only combine the groups when adjacent group minimizers violate monotonicity but also when the quantized mappings are the same as well. Hence, the total time complexity will be \( O(I_1 + I_0 - I) \).}

**Theorem 3.** The sequential algorithm updates the optimal mapping in \( O(K\log N) \) time per sample for some \( K \), which is the upper bound of the number of groups in the sets.

**Proof.** With a new sample, the number of updates totals to \( O(\log N) \) from Lemma 4 and Lemma 5. Since each update takes a time that is linear with the number of groups from Lemma 6, the mapping is updated in \( O(K\log N) \) time, where \( K \) is an upper bound on the number of groups.

**Remark 2.** In the traditional isotonic L2 regression (which is done on the set of reals), the number of groups will be bounded by \( N \); hence, the per sample time complexity would be linear. On the other hand, when the quantization set \( Q \) is bounded with fixed cardinality (e.g., percentage), the per sample time complexity will be logarithmic. If the quantization set is not bounded, then the per sample complexity will be pseudo-linear (dependent on the range of target variables \( y_n \)).

Henceforth, we have shown that for isotonic L2 regression problems, it is possible to achieve efficient optimal sequential algorithms, especially when the mapping is quantized.
