Linear Time Isotonic and Unimodal Regression in the L_1 and L_∞ Norms

Victor Boyarshinov
Rm 207 Lally Building, CS, RPI,
110 8th Street, Troy, NY 12180.
boyarv@cs.rpi.edu

Malik Magdon-Ismail
Rm 207 Lally Building, CS, RPI,
110 8th Street, Troy, NY 12180.
magdon@cs.rpi.edu

Abstract

We consider L_1-isotonic regression and L_∞ isotonic and unimodal regression. For L_1-isotonic regression, we present a linear time algorithm when the number of outputs are bounded. We extend the algorithm to construct an approximate isotonic regression in linear time when the output range is bounded. We present linear time algorithms for L_∞ isotonic and unimodal regression.

1 Introduction

Isotonic regression in the L_p norm, $p > 0$, is defined as follows. Let $\mathbf{x} = [x_1, x_2, \ldots, x_n], x_i \in \mathbb{R}$, be given. The task is to construct a corresponding sequence $\mathbf{w} = [w_1 \leq w_2 \leq \ldots \leq w_n]$ so that $\mathcal{E}_p(\mathbf{w})$ is minimized, where

$$\mathcal{E}_p(\mathbf{w}) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n} |x_i - w_i|^p & 1 \leq p < \infty, \\ \max_i |x_i - w_i| & p = \infty. \end{cases}$$

The regression is *unimodal* if $w_1 \leq w_2 \leq \cdots \leq w_i \geq w_{i+1} \geq \cdots \geq w_n$, for some i; x_i is denoted a *crossover* point. The prefix-isotonic regression problem is to construct the isotonic regression for all prefixes of $\mathbf{x}$. We study the cases $p = 1$ and $p = \infty$. The case $p = 1$ is sometimes denoted isotonic median regression. We will refer to $\mathcal{E}_1(\mathbf{w})$ or $\mathcal{E}_\infty(\mathbf{w})$ as the error of the regression when the context is clear. The efficiency of an algorithm is measured in terms of n.

In this paper, we give new algorithms (with proofs of correctness and run times) for output sensitive isotonic median regression, isotonic L_∞ regression, and unimodal L_∞ regression. More specifically,

(i) Suppose that $x_i \in \mathcal{X}$ where $|\mathcal{X}| = K$. We give an L_1-isotonic regression algorithm with running time in $O(n \log K)$, linear in n. In the worst case, $K = n$ and we have an $O(n \log n)$ algorithm.

(ii) Let $x_i \in [a, b]$ $\forall i$. For any $\epsilon > 0$, we give an $O(n \log(\frac{b-a}{\epsilon}))$ algorithm to construct an approximate L_1-isotonic regression with error at most ϵ above the optimal error.

(iii) We give linear time algorithms for L_∞ prefix-isotonic and unimodal regression.

Note that in (i), we do not need to assume that K (or $\mathcal{X}$) is known, because, using appropriate data structures [4], $\mathcal{X}$ can be determined and sorted in $O(n \log K)$ resulting in no change to the asymptotic running time. In (ii), a and b can be determined in linear time, which also will not affect the asymtotic running time of the algorithm.

Applications of isotonic regression can be found in [12, 13]. Isotonic and unimodal regression are both examples of nonparametric shape constrained regression. Such regressions are useful when prior knowledge about the shape but not the parametric form of a function are known. The importance of isotonic regression stems from the fact that it is often the case in statistical estimation or learning that one wishes to estimate a function that is known to be monotonically increasing (say), even though the data will not necessarily exhibit this behavior, on account of noise, [7, 14]. Examples include the probability of heart attack as a function of cholesterol level, [7]; the "credit worthiness" as a function of income, [14]. To illustrate, suppose that we would like to determine cholesterol level thresholds at which a heart attack becomes more prevalent, and we have a sequence of patients with cholesterol levels $c_1 < c_2 < \ldots < c_n$. Associated to each patient i, let x_i be the number of heart attacks they had within the following year, $x_i = 0, 1, 2, \ldots, K$ for some small value of K. The isotonic regression determines thresholds for the cholesterol levels that identify different severities for heart attack risk.

The outline of the remainder of the paper is as follows. First we summarize the previous work in this area. Then, we present some results on L_1 isotonic regression that lead to the final $O(n \log K)$ algorithm. Lastly, we cover L_∞ isotonic regression, from which the L_∞ unimodal regression naturally follows. We give all algorithms in their respective sections.

1.1 Previous Work

L_2 isotonic regression can be performed efficiently in linear time using some variant of a Pooling Adjacent Violators (PAV) algorithm [1, 12, 13]. For L_1 isotonic regression, algorithms in the efficiency class $O(n \log n)$ are known. Some approaches to isotonic regression are given in [3, 10, 11, 13].

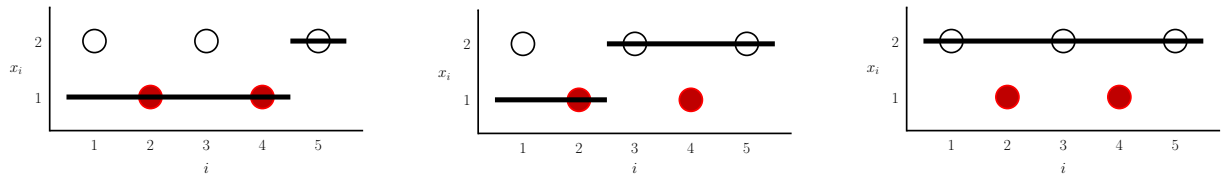
The L_1 and L_2 prefix-isotonic regression problems have been solved optimally in [15]. For L_2, the runtime is $O(n)$, which is clearly optimal, and for L_1 it is $O(n \log n)$, which, by a reduction from sorting, is optimal [15]. While $O(n \log n)$ is optimal for L_1 prefix-isotonic regression, it is not known whether the apparently simpler isotonic regression problem can be performed faster than $O(n \log n)$. We take a first step in this direction by obtaining a linear bound in terms of the size of the output (K).

Unimodal regression has been studied extensively in [15], where the author gives a linear time algorithm for the L_2 case, and an $O(n \log n)$ algorithm for the L_1 cases. This result significantly improved upon prior results which were exponential and quadratic algorithms [5, 6, 8, 9].

A general PAV type algorithm, [15], relies on the ability to efficiently update a suitably defined "mean". Such an algorithm is easily applicable to the L_1 and L_2 cases, however, for $p > 2$, the "L_p-mean" is not conveniently updated. The case $p = \infty$ is not considered in [15]. The statistical properties of L_∞ isotonic regression have been considered before[2, 13, 16, 17]. Our algorithms for L_∞ prefix-isotonic and unimodal regression are simple, not relying on advanced data structures, and linear time, hence optimal.

2 L_1-Isotonic Regression

We use the notation $[i, j]$ to refer to the interval of integers $\{i, i+1, \ldots, j\}$, and $\mathbf{x}[i, j]$ to represent the sequence $[x_i, \ldots, x_j]$. In this section, the isotonic regression will always refer to L_1-optimal isotonic regression. Without loss of generality, we can represent the isotonic regression by a collection of monotonically increasing *level sets*, or intervals to each of which is associated a value or *level*: $\mathcal{C} = \{I_\alpha, h_\alpha\}_{\alpha=1}^K$. Each I_α is an interval of the form $I_\alpha = [i_\alpha, j_\alpha]$, and each h_α is the level corresponding to I_α. We assume that $i_1 = 1$, $j_K = n$, $i_{\alpha+1} = j_\alpha + 1$ and $h_\alpha < h_{\alpha+1}$ for $1 \leq \alpha < K$. The isotonic regression that is induced by $\mathcal{C}$ is given by assigning $w_i = h_\alpha$ for all $i \in I_\alpha$. We define the error for $\mathcal{C}$, $\mathcal{E}_1(\mathcal{C})$, as the error $\mathcal{E}_1(\mathbf{w})$ of the corresponding induced isotonic regression $\mathbf{w}$. The example below illustrates all this notation for the sequence $\mathbf{x} = [2, 1, 2, 1, 2]$.



$\mathcal{C} = \{([1,4],1),([5,5],2)\}$
$\mathbf{w} = [1,1,1,1,2]$
$\mathcal{E}_1(\mathcal{C}) = 2$

$\mathcal{C} = \{([1,2],1),([3,5],2)\}$
$\mathbf{w} = [1,1,2,2,2]$
$\mathcal{E}_1(\mathcal{C}) = 2$

$\mathcal{C} = \{([1,5],2)\}$
$\mathbf{w} = [2,2,2,2,2]$
$\mathcal{E}_1(\mathcal{C}) = 2$

Note that the isotonic regression is not unique. To remove this ambiguity, we will only consider the isotonic regression in which the sum of the w_i is minimized (the leftmost regression in the figure). We define the *weight* of the isotonic regression by $W(\mathcal{C}) = \sum_i w_i$, where $\{w_i\}$ is the isotonic regression induced by $\mathcal{C}$. Thus if $\mathcal{C}$ is an isotonic regression, and $\mathcal{C}'$ is any other monotonically increasing collection of level sets, then $\mathcal{E}_1(\mathcal{C}) \leq \mathcal{E}_1(\mathcal{C}')$, and if $\mathcal{E}_1(\mathcal{C}) = \mathcal{E}_1(\mathcal{C}')$, then $W(\mathcal{C}) < W(\mathcal{C}')$ (we will show later that the minimum weight isotonic regression is indeed unique). Throughout, we will refer to the unique minimum weight isotonic regression by $\mathcal{C} = \{I_\alpha = [i_\alpha, j_\alpha], h_\alpha\}_{\alpha=1}^K$, and in general, we will use I_α to refer both to the interval $[i_\alpha, j_\alpha]$, as well as to the set of points $\{x_{i_\alpha}, \ldots, x_{j_\alpha}\}$.

We define the median of a level set $I = [i, j]$, $M(I)$, to be the median of the points $\{x_i, x_{i+1}, \ldots, x_j\}$, where the median is defined in the usual way:

$$M(y_1 \leq y_2 \ldots \leq y_m) = y_{\lfloor \frac{m+1}{2} \rfloor}$$

Note that $M(I) = x_k$ for some $k \in [i, j]$. Further, note that if $M(S_1) \leq M(S_2)$ for any S_1, S_2, then $M(S_1) \leq M(S_1 \cup S_2) \leq M(S_2)$. It is also well known that the median is a minimizer of the L_1 error. Since we require the weight of the isotonic regression to be minimum, we conclude that the level of each level set has to be the median of the set:

Proposition 2.1 *$h_\alpha = M(I_\alpha)$ for all $\alpha \in [1, K]$.*

Proof: Suppose that $h_\alpha < M(I_\alpha)$ for some α. This means that there are strictly more points in I_α above h_α than below. By raising h_α we can decrease the error, contradicting the optimality of the isotonic regression. Suppose that $h_\alpha > M(I_\alpha)$ for some α. In this case, by the definition of the median, there are at least as many points below h_α as there are above. In this case, we guarantee not to increase the error by lowering h_α, and at the same time decrease the sum of the w_i contradicting the minimality of $W(\mathcal{C})$. ■

In particular, $h_\alpha = x_k$ for some $k \in [i_\alpha, j_\alpha]$, i.e., every level is one of the x_i's. Note that since, $h_\alpha < h_{\alpha+1}$, we immediately have that the sequence of medians must be increasing.

Corollary 2.2 *$M(I_\alpha) < M(I_\beta)$ for $1 \leq \alpha < \beta \leq K$.*

The next proposition is one of the crucial properties that we will use. It essentially states that the isotonic regression for a set of points is the union of the isotonic regressions for two disjoint subsets of the points. Consider *any* level set I_α in the isotonic regression and define the left and right subsets of the points with respect to this level set by $S_l = \{x_1, \ldots, x_{i_\alpha - 1}\}$ and $S_r = \{x_{i_\alpha}, \ldots, x_n\}$. We define the left and right isotonic regressions $\mathcal{C}_l$ and $\mathcal{C}_r$ as the isotonic regressions for the respective left and right subsets. Then $\mathcal{C} = \mathcal{C}_l \cup \mathcal{C}_r$. We will need the following lemma to prove the proposition,

Lemma 2.3 *For any α, with $I_\alpha = [i_\alpha, j_\alpha] \in \mathcal{C}$,*

$$\begin{array}{ll} (i) & M(\{x_{i_\alpha}, \ldots, x_j\}) \geq M(I_\alpha), \text{ for all } j \geq i_\alpha. \\ (ii) & M(\{x_i, \ldots, x_{j_\alpha}\}) \leq M(I_\alpha), \text{ for all } i \leq j_\alpha. \end{array}$$

Proof: *(i)* Let I_α be the last level set for which there exists a $j \geq i_\alpha$, with $M(\{x_{i_\alpha}, \ldots, x_j\}) < M(I_\alpha)$. Suppose $j > j_\alpha$. Then, $M(x_{i_{\alpha+1}}, \ldots, x_j) < M(I_\alpha) < M(I_{\alpha+1})$ and so I_α is not the last level set with this property. Thus, $j < j_\alpha$. Decompose (I_α, h_α) into two level sets: $(I_1 = \{x_{i_\alpha}, \ldots, x_j\}, \max(h_{\alpha-1}, M(I_1)))$ and $(I_2 = \{x_{j+1}, \ldots, x_{j_\alpha}\}, h_\alpha)$. The decomposition guarantees not to increase the error, while lowering the weight of the regression, contradicting the fact that $\mathcal{C}$ has minimum weight among optimal isotonic regressions.

(ii) Let I_α be the first level set for which there exists an $i \leq j_\alpha$, with $M(\{x_i, \ldots, x_{j_\alpha}\}) > M(I_\alpha)$. Suppose $i < i_\alpha$. Then, $M(x_i, \ldots, x_{j_{\alpha-1}}) > M(I_\alpha) > M(I_{\alpha-1})$ and so I_α is not

the first level set with this property. Thus, $i > i_\alpha$. Decompose (I_α, h_α) into two level sets: $(I_1 = \{x_{i_\alpha}, \ldots, x_{i-1}\}, h_\alpha)$ and $(I_2 = \{x_i, \ldots, x_{j_\alpha}\}, \min(h_{\alpha+1}, M(I_2)))$. The decomposition strictly decreases the error, contradicting the fact that $\mathcal{C}$ has minimum error.

■

Proposition 2.4 $\mathcal{C} = \mathcal{C}_l \cup \mathcal{C}_r$

Note that the proposition is valid for any level set I_α that is used to construct S_l, S_r.

Proof: Let $\mathcal{C}' = \mathcal{C}_l \cup C_r = \{I'_\beta, h'_\beta\}_{\beta=1}^{K'}$. Since $h'_\beta = M(I'_\beta)$, it will suffice to show that $I'_\alpha = I_\alpha$ for all $\alpha \in [1, K]$. Suppose to the contrary and let α^* be the first level set for which $I_{\alpha^*} \neq I'_{\alpha^*}$. Further, suppose without loss of generality that $|I_{\alpha^*}| > |I'_{\alpha^*}|$ (a similar argument holds for $|I_{\alpha^*}| < |I'_{\alpha^*}|$). Therefore,

$$I_{\alpha^*} = I'_{\alpha^*} \cup I'_{\alpha^*+1} \cup \cdots \cup I'_{\alpha^*+L} \cup P,$$

where P is a prefix of I'_{α^*+L+1}. Note that $I_{\alpha^*}, \ldots, I'_{\alpha^*+L+1}$ are either all in $\mathcal{C}_l$ or all in $\mathcal{C}_r$. Without loss of generality, assume they are all in $\mathcal{C}_l$. We know that $h'_{\alpha^*+i} = M(I'_{\alpha^*+i})$ for $i \in [0, L+1]$ and by construction, $h'_{\alpha^*+i} < h'_{\alpha^*+i+1}$ for $i \in [0, L]$. From Lemma 2.3, we know that $M(P) \geq M(I'_{\alpha^*+L+1})$ (since $\mathcal{C}_l$ is the isotonic regression for S_l). By Lemma 2.3, we also have that $M(I_{\alpha^*}) \geq M(I'_{\alpha^*})$, and similarly from the optimality of $\mathcal{C}$, we have that $M(I'_{\alpha^*}) \geq M(I_{\alpha^*})$, hence that $M(I'_{\alpha^*}) = M(I_{\alpha^*})$. Therefore, we have that

$$M(I_{\alpha^*}) = M(I'_{\alpha^*}) < M(I'_{\alpha^*+1}) < \cdots < M(I'_{\alpha^*+L}) < M(I'_{\alpha^*+L+1}) \leq M(P).$$

Since P is a suffix of I_{α^*}, by the optimality of $\mathcal{C}$ and Lemma 2.3, we have that $M(P) \leq M(I_{\alpha^*})$ which is the desired contradiction. ■

An immediate consequence of this proposition is that if $\mathcal{C}_l$ and $\mathcal{C}_r$ are unique, then so is $\mathcal{C}$. An induction argument then gives the following result,

Corollary 2.5 *The isotonic regression is unique.*

Suppose we are given a constant γ, we would like to find the first level set whose height is at least γ. In particular, we would like to find the first point of this level set. We call this point a pivot point for γ. More specifically, let $\mathcal{C}$ be the isotonic regression, and let α be such that $h_\alpha \geq \gamma$ and if $\alpha > 1$, then $h_{\alpha-1} < \gamma$. We would like to find x_{i_α}. Note that if all the levels are $< \gamma$, then x_{i_α} does not exist, in which case we can default to $i_\alpha = n + 1$. We know from Lemma 2.3 that it is necessary for x_{i_α} to satisfy two conditions:

(i) for every sequence S begining at x_{i_α}, $M(S) \geq h_\alpha \geq \gamma$;

(ii) for every sequence S' ending at $x_{i_\alpha - 1}$, $M(S') \leq h_{\alpha-1} < \gamma$.

The content of the next proposition is that these conditions are also sufficient.

Theorem 2.6 *Let $\mathcal{C}$ be the isotonic regression. Given γ, let I_α be the first level set with $h_\alpha \geq \gamma$. Then, x_i is the first point in I_α (i.e., $x_i = x_{i_\alpha}$) if and only if for any sequence S begining at x_i and any sequence S' ending at x_{i-1}, $M(S') < \gamma \leq M(S)$.*

Proof: It only remains to prove that if $M(S') < \gamma \leq M(S)$ for any two sequences as described, then $i = i_\alpha$. We know that i must belong to one of the level sets, $i \in I_\beta$ for some β with $1 \leq \beta \leq K$. We need to show three things: (i) $h_\beta \geq \gamma$; (ii) $i = i_\beta$; (iii) $h_{\beta-1} < \gamma$.

(i) Suppose that $h_\beta < \gamma$. Then, consider $S = \{x_i, \ldots, x_{j_\beta}\}$. By Lemma 2.3, $M(S) \leq h_\beta < \gamma$. By construction of x_i, $M(S) \geq \gamma$, a contradiction.

(ii) Suppose that i is not the first point in I_β. Then consider $S' = \{x_{i_\beta}, \ldots, x_{i-1}\}$. By Lemma 2.3, $M(S') \geq h_\beta \geq \gamma$ (by (i)). By construction of x_i, $M(S') < \gamma$, a contradiction.

(iii) Suppose that $h_{\beta-1} \geq \gamma$. Consider $S' = \{x_{i_{\beta-1}}, \ldots, x_{i-1}\}$. From (ii), this is exactly $I_{\beta-1}$. By construction of x_i, $M(S') = M(I_{\beta-1}) = h_{\beta-1} < \gamma$, a contradiction.

■

Thus to find the first point of the first level set with height at least a given γ, we only need to search for an x_i that satisfies the conditions of Theorem 2.6. The remainder of this section is devoted to developing a linear time algorithm to find this point. This algorithm will be the basis of our isotonic regression algorithms that we discuss in the next section.

Define the following three quantities for any interval $[i, j]$.

$$\begin{array}{rl} N^+(i,j): & \text{the number of points} \geq \gamma \text{ in the set } S_{[i,j]} = \{x_i, \ldots, x_j\}. \\ N^-(i,j): & \text{the number of points} < \gamma \text{ in the set } S_{[i,j]} = \{x_i, \ldots, x_j\}. \\ \Delta_r(i,j): & \min\limits_{t \in [i,j]} (N^+(i,t) - N^-(i,t)). \\ \Delta_l(i,j): & \max\limits_{t \in [i,j]} (N^+(t,j) - N^-(t,j)). \end{array}$$

Note that the median of the set $S_{[i,j]}$ is $\geq \gamma$ if and only if $N^+(i,j) - N^-(i,j) > 0$. From this observation, we get the following lemma.

Lemma 2.7 *x_k satisfies the conditions of Theorem 2.6 if and only if one of the following hold:*

(i) $k = 1$ and $\Delta_r(k, n) > 0$;

(ii) $k > 1$, $\Delta_r(k, n) > 0$ and $\Delta_l(1, k-1) \leq 0$.

If no such x_k exists, then the levels of all the level sets are $< \gamma$.

We show how to find such an x_i in *linear* time. Start two pointers $p_l = 0$ and $p_r = n + 1$. The initial conditions of the algorithm are:

$$\begin{array}{ll} N^+(p_r, n) = 0; & N^-(p_r, n) = 0, \\ N^+(1, p_l) = 0; & N^-(1, p_l) = 0. \end{array}$$

Let $\mathbf{x}_l = \mathbf{x}[1, p_l]$, $\mathbf{x}_r = \mathbf{x}[p_r, n]$, and $S = \mathbf{x}[p_l + 1, p_r - 1]$. Initially, $\mathbf{x}_r = \mathbf{x}_l = \{\}$, and $S = \mathbf{x}$. If $M(S) \geq \gamma$, then we know that x_{p_r} is not our solution, so we decrement p_r by 1 and update $\mathbf{x}_l, \mathbf{x}_r, S$. On the other hand, if $M(S) < \gamma$, then x_{p_l+1} is not our solution, so we increment p_l by 1 and update $\mathbf{x}_l, \mathbf{x}_r, S$. We continue this process of decreasing p_r or increasing p_l until $p_r = p_l + 1$. We now prove that this algorithm correctly computes the pivot point. The nature of the algorithm is to move p_r (resp. p_l) until $M(S)$ switches from $\geq \gamma$ (resp. $< \gamma$) to $< \gamma$ (resp. $\geq \gamma$). Denote a phase in the algorithm as the period when one of the pointers begins to move and then stops.

Lemma 2.8 *The following invariants are maintained at the end of every phase.*

i. The median of every prefix of $\mathbf{x}_r$ *is* $\geq \gamma$.

ii. The median of every suffix of $\mathbf{x}_l$ *is* $< \gamma$.

Proof: We prove the claim by induction on the phase number. Initially the invariants hold by default since $\mathbf{x}_r$ and $\mathbf{x}_l$ are empty. Suppose the invariants hold up to some phase, and consider the next phase, i.e., $p_l + 1 < p_r$.

Suppose that $p_l \to p_l'$ and $\mathbf{x}_l \to \mathbf{x}_l'$ in this phase. By construction, $M(\mathbf{x}[k, p_r - 1]) < \gamma$ for $p_l + 1 \leq k \leq p_l'$. Since p_l stopped moving, there are two cases. (i) $p_l' = p_r - 1$, in which case the median of every suffix of $\mathbf{x}[p_l + 1, p_l']$ is $< \gamma$. (ii) $p_l' < p_r - 1$, in which case $M(\mathbf{x}[p_l' + 1, p_r - 1]) \geq \gamma$. But since $M(\mathbf{x}[k, p_r - 1]) < \gamma$ for $p_l + 1 \leq k \leq p_l'$, it follows that $M(\mathbf{x}[k, p_l']) < \gamma$, or once again, the median of every suffix of $\mathbf{x}[p_l + 1, p_l']$ is $< \gamma$. Every suffix of $\mathbf{x}_l'$ is either a suffix of $\mathbf{x}[p_l + 1, p_l']$ or the union of $\mathbf{x}[p_l + 1, p_l']$ with a suffix of $\mathbf{x}_l$. Since $M(S_1) < \gamma$ and $M(S_2) < \gamma$ implies $M(S_1 \cup S_2) < \gamma$ for any S_1, S_2, invariant (ii) now follows, i.e., the median of every suffix of $\mathbf{x}_l'$ is $< \gamma$. Since p_r did not move in this phase, invariant (i) was unchanged.

Similarily, suppose instead that $p_r \to p_r'$ and $\mathbf{x}_r \to \mathbf{x}_r'$ in this phase. This means that $M(\mathbf{x}[p_l + 1, k]) \geq \gamma$ for $p_r' \leq k \leq p_r - 1$. Once again, there are two cases, $p_r' = p_l + 1$ and $p_r' > p_l + 1$. In both cases it follows using similar arguments that the median of every prefix of $\mathbf{x}[p_r', p_r - 1]$ is $\geq \gamma$. Invariant (i) follows from the facts that any prefix of $\mathbf{x}_r'$ is the union of prefixes of $\mathbf{x}[p_r', p_r - 1]$ and $\mathbf{x}_r$, and $M(S_1) \geq \gamma,\ M(S_2) \geq \gamma \implies M(S_1 \cup S_2) \geq \gamma$. Since p_l did not move in this phase, invariant (ii) was unchanged. ∎

Thus when the algorithm concludes, $\Delta_r(p_r, n) > 0$ and $\Delta_l(p_1, l) \leq 0$ and we have the pivot point. The efficiency of the algorithm hinges on being able to determine if $M(S)$ is larger or smaller than γ. Since $M(\mathbf{x}[i, j]) \geq \gamma$ if and only if $N^+(i, j) - N^-(i, j) > 0$, we need to maintain $N^{\pm}(p_l + 1, p_r - 1)$. The following update rules allow us to do this efficiently. Suppose we have computed $N^{\pm}(i, j)$ for $1 \leq i < j \leq n$

$$\begin{array}{lll} N^+(i+1,j) = N^+(i,j) - 1; & N^-(i+1,j) = N^-(i,j) & \text{if } x_i \geq \gamma. \\ N^+(i+1,j) = N^+(i,j); & N^-(i+1,j) = N^-(i,j) - 1 & \text{if } x_i < \gamma. \end{array}$$

$$\begin{array}{lll} N^+(i,j-1) = N^+(i,j) - 1; & N^-(i,j-1) = N^-(i,j) & \text{if } x_j \geq \gamma. \\ N^+(i,j-1) = N^+(i,j); & N^-(i,j-1) = N^-(i,j) - 1 & \text{if } x_j < \gamma. \end{array}$$

Algorithm 1 Algorithm to compute a pivot point.

1: //**Input:** $\mathbf{x} = \{x_i | i \in [1, n]\}$ and $\gamma \in \mathbb{R}$.
2: //**Output:** i such that x_i is the pivot point for γ.
3: Set $p_l = 0$, $p_r = n + 1$ and using a single scan compute $N^{\pm}(p_l + 1, p_r - 1)$;
4: **while** $p_l + 1 \neq p_r$ **do**
5: **if** $N^{+}(p_l + 1, p_r - 1) - N^{-}(p_l + 1, p_r - 1) > 0$ **then**
6: $p_r \leftarrow p_r - 1$, and update $N^{\pm}(p_l + 1, p_r - 1)$;
7: **else**
8: $p_l \leftarrow p_l + 1$, and update $N^{\pm}(p_l + 1, p_r - 1)$;
9: **end if**
10: **end while**
11: **return** p_r;{$p_r = n + 1$ if all levels are $< \gamma$.}

The entire algorithm is summarised in Algorithm 1.

We define an operation as a comparison, a floating point operation or an assignment. Step 3 can be computed in $3n$ operations. An update (steps 5–8) takes 6 operations (2 comparisons and 4 operations to update $N^{\pm}$), and n updates need to be made. We thus have the following theorem.

Theorem 2.9 *Given $\mathbf{x} = \{x_i | i \in [1, n]\}$ and $\gamma \in \mathbb{R}$, the pivot point for γ can be found using at most Cn operations, where $C \approx 9$.*

Summary. The pivot point x_i for any value γ can be found in linear time. The sequence $\mathbf{x}$ can then be partitioned into two disjoint subsets, $\mathbf{x}_l = \mathbf{x}[1, i-1]$ and $\mathbf{x}_r = \mathbf{x}[i, n]$. The isotonic regression $\mathcal{C}_l$ of $\mathbf{x}_l$ will have level sets all of whose levels are $< \gamma$, and the isotonic regression $\mathcal{C}_r$ of $\mathbf{x}_r$ will have level sets all of whose levels are $\geq \gamma$. Further, the isotonic regression $\mathcal{C}$ of $\mathbf{x}$ is given by $\mathcal{C} = \mathcal{C}_l \cup \mathcal{C}_r$. This result already has applications. Suppose we would simply determine a threshold x where the response function exceeds a given value, γ. This can be accomplished by finding the pivot point for γ in linear time.

2.1 L_1-Isotonic Regression: Algorithms

The importance of Proposition 2.4 and Theorem 2.9 from the algorithmic point of view can be summarised as follows. Suppose we have the input $\mathbf{x}$ for which the isotonic regression can only have levels in the set $\{m_1 < m_2 < \cdots < m_K\}$ – for example, this would be the case if x_i can only take values in this set. Let p be the index of the pivot point for $\gamma = m_i$, $i \in [1, K]$. This pivot point, which can be found in linear time, partitions $\mathbf{x}$ into $\mathbf{x}_l = \mathbf{x}[1, p-1]$ and $\mathbf{x}_r = \mathbf{x}[p, n]$ (one of these may be empty). By Proposition 2.4, it then suffices to recursively compute the isotonic regressions for $\mathbf{x}_l$ and $\mathbf{x}_r$. Further, by construction of p, all the levels in $\mathbf{x}_l$ will be $< \gamma = m_i$, and all the levels in $\mathbf{x}_r$ will be $\geq \gamma$. We obtain an efficient algorithm

Algorithm 2 Algorithm to perform the full isotonic regression.

1: // **Wrapper to call the recursive function.**
2: //**Input:** $\mathbf{x} = \{x_i | i \in [1, n]\}$ and $\mathbf{m} = \{m_1 < m_2 < \cdots < m_K\}$.
3: //**Output:** Isotonic regression, $\mathcal{C} = \{(I_\alpha, h_\alpha)\}$
4: Call ISOTONIC($\mathbf{x}, \mathbf{m}, [1, n], [1, K]$);

1: **ISOTONIC**($\mathbf{x}, \mathbf{m}, [i, j], [k, l]$)
2: //**Output:** Isotonic regression $\mathcal{C} = \{(I_\alpha, h_\alpha)\}$ for $\mathbf{x}[i, j]$, given all levels are in $\mathbf{m}[k, l]$.
3: **if** $j < i$ **then**
4: **return** $\{\}$;
5: **else if** $k = l$ **then**
6: **return** $\{([i, j], \mathbf{m}[k])\}$
7: **else**
8: Let $q = k + 1 + \lfloor \frac{l-k}{2} \rfloor$; {$q$ is 1+the median of $[k, l]$}
9: Let p=index of pivot point for $\mathbf{x}[i, j]$ with $\gamma = \mathbf{m}[q]$;
10: $\mathcal{C}_l$ =ISOTONIC($\mathbf{x}, \mathbf{m}, [i, p-1], [k, q-1]$); $\mathcal{C}_r$ =ISOTONIC($\mathbf{x}, \mathbf{m}, [p, j], [q, l]$);
11: **return** $\mathcal{C}_l \cup \mathcal{C}_r$;
12: **end if**

by choosing γ to be the median of the available levels each time in the recursion. The full algorithm is given in Algorithm 2.

The correctness of this algorithm follows from the results in the previous section, specifically Proposition 2.4. What remains is to analyse the run time. It is enough to analyse the runtime of ISOTONIC($\mathbf{x}, \mathbf{m}, [i, j], [k, l]$). Let $T(n, K)$ be the worst case runtime when $|[i, j]| = n$ and $|[k, l]| = K$. Then in the worst case, the algorithm will call itself on a left set of size δ with $\lceil K/2 \rceil$ levels and on a right set of size $n - \delta$ with $\lfloor K/2 \rfloor$ levels, for some $0 \leq \delta \leq n$. As already discussed, the pivot step to perform this partition takes at most Cn operations (step 9), so we have the following recursion for $T(n, K)$:

$$T(n, K) \leq \max_{\delta \in [0,n]} \left(T(\delta, \lceil \tfrac{K}{2} \rceil) + T(n - \delta, \lfloor \tfrac{K}{2} \rfloor) \right) + Cn.$$

For $K = 2^l$, a straight forward induction shows that $T(n, K) \leq Cn \log K$. By monotonicity, $T(n, K) \leq T(n, 2^{\lceil \log K \rceil})$, which gives $T(n, K) \leq Cn \lceil \log K \rceil$, yielding the following theorem.

Theorem 2.10 *The isotonic regression for n points with K possible levels can be obtained in $O(n \log K)$ time.*

If the K levels are not known ahead of time, they can be determined and sorted using standard data structures, such as a balanced binary seach tree in $O(n \log K)$ time, [4]. This does not affect the asymptotic running time. In the worst case, $K = n$ and our algorithm is no worse than existing algorithms. However, there can be significant improvement in the efficiency when K is fixed and small.

Approximate isotonic regression. The algorithm that we have given can be run with any set of levels supplied – the pivot point is defined for any γ. It is not required that the true isotonic regression levels all be from this set in order to run the algorithm. Of course, if the true levels are not from the set of levels supplied to the algorithm, then the result cannot be the true isotonic regression. If the levels chosen are close to the true levels, then the approximate isotonic regression should be close to the true one.

In particular, suppose that $a \leq x_i \leq b$ for all $i \in [1, n]$. Consider the levels $m_i = a + i\epsilon$, where $\epsilon = (b-a)/K$ and $i \in [0, K]$. Suppose that $[i_\alpha, j_\alpha], h_\alpha$ is a (non-empty) level set output by the algorithm, $h_\alpha = a + i_\alpha \epsilon$. Then x_{i_α} is a pivot point, for which all the levels of the *true* isotonic regression to the right are $\geq h_\alpha$. Further, all the levels to the left of the next level set that is output are $< h_\alpha + \epsilon$. Therefore, the error of a point from its corresponding level output by the algorithm differs from its error with respect to the true isotonic regression level by at most ϵ. Thus, the additional error contributed by every point is at most ϵ, so the average error contribution per point is at most ϵ larger than the average error contribution per point for the optimal regression. Further, the runtime is $O(n \log K) = O(n \log((b-a)/\epsilon))$, establishing the following theorem as a corollary of Theorem 2.10.

Theorem 2.11 *Suppose that $a \leq x_i \leq b$ for $i \in [1, n]$ and let $\mathbf{w}$ be the isotonic regression. Then, an approximate isotonic regression $\mathbf{w}'$ can be computed in $O(n \log((b-a)/\epsilon))$ time with $\mathcal{E}_1(\mathbf{w}') - \mathcal{E}_1(\mathbf{w}) \leq \epsilon$.*

3 L_∞-Prefix-Isotonic Regression

In this section, we will refer to the L_∞-optimal isotonic regression more simply as the isotonic regression (which is not necessarily unique). For any sequence of points $\mathbf{x} = [x_1, x_2, ..., x_n]$, define a *Maximally Violating Pair (MVP)* to be a pair of points that maximally violates the monotonicity requirement, i.e., an MVP is a pair (x_l, x_r) with $l < r$, $x_l > x_r$, and $\forall i < j$, $x_l - x_r \geq x_i - x_j$. If $x_i \leq x_j$ for all $i < j$, then no such pair exists. If $\mathbf{x}$ has an MVP (x_l, x_r), we define the *distortion* of $\mathbf{x}$, $D(\mathbf{x})$, to be $(x_l - x_r)$, and $D(\mathbf{x}) = 0$ if $\mathbf{x}$ does not have an MVP. Note that by definition of an MVP, $x_i \leq x_l$ for all $i < r$ and $x_j \geq x_r$ for all $j > l$.

Let $\mathcal{C}$ be an isotonic regression for $\mathbf{x}$ and let (x_l, x_r) be an MVP. Let w_l, w_r be the levels corresponding to x_l, x_r respectively. Then, either $w_l \leq (x_l + x_r)/2$ or $w_r \geq w_l > (x_l + x_r)/2$, so we conclude that $\mathcal{E}_\infty(\mathcal{C})$ cannot be less that $D(\mathbf{x})/2$. The next proposition shows that this lower bound is achievable.

Proposition 3.1 *Let $\mathcal{C}$ be an isotonic regression for $\mathbf{x}$. Then $\mathcal{E}_\infty(\mathcal{C}) = D(\mathbf{x})/2$. Further, if (x_l, x_r) is an MVP, then $w_l = w_r = (x_l + x_r)/2$.*

Proof: If $D(\mathbf{x}) = 0$, then $\mathbf{x}$ is a monotonically nondecreasing sequence. $w_i = x_i$ is the optimal regression with $\mathcal{E}_\infty = 0$. Suppose that $D(\mathbf{x}) > 0$. We will construct (by induction) an isotonic regression with error $D(\mathbf{x})/2$. It then follows immediately that $w_l \geq (x_l + x_r)/2 \geq w_r$, and by monotonicity, $w_r \geq w_l$ from which we get $w_l = w_r = (x_l + x_r)/2$.

The induction basis is when $\mathbf{x} = \{\}$, $\mathbf{x} = [x_1]$ or $\mathbf{x} = [x_1, x_2]$, in which cases the claim is obvious. Suppose that an optimal regression exists with error $D(\mathbf{x})/2$ whenever $|\mathbf{x}| \leq N$, and consider any sequence $\mathbf{x}$ with $|\mathbf{x}| = N+1$ and $D(\mathbf{x}) > 0$. Let (x_l, x_r) be an MVP, and define the left and right sequences: $\mathbf{x}_l = [x_1, x_2, \ldots, x_{l-1}]$; and $\mathbf{x}_r = [x_{r+1}, x_{r+2}, \ldots, x_{N+1}]$. Note that $D(\mathbf{x}_l) \leq D(\mathbf{x})$ and $D(\mathbf{x}_r) \leq D(\mathbf{x})$. Let $\mathcal{C}_l$ and $\mathcal{C}_r$ be the isotonic regressions for $\mathbf{x}_l$ and $\mathbf{x}_r$ respectively. Since the left and right sequences are strictly shorter than $\mathbf{x}$, by the induction hypothesis, we have that $\mathcal{E}_\infty(\mathcal{C}_l) = D(\mathbf{x}_l)/2 \leq D(\mathbf{x})/2$ and $\mathcal{E}_\infty(\mathcal{C}_r) = D(\mathbf{x}_r)/2 \leq D(\mathbf{x})/2$.

We now show how to construct the isotonic regression for $\mathbf{x}$ with error $D(\mathbf{x})/2$ from $\mathcal{C}_l$, $\mathcal{C}_r$ and one additional level set $\mathcal{C}^* = \{(I = [l, r], h = (x_l + x_r)/2)\}$. Consider all level sets in $\mathcal{C}_l$ with level $\geq h$. Reduce *all* these levels to h, and call this new isotonic regression $\mathcal{C}'_l$. We claim that $\mathcal{E}_\infty(\mathcal{C}'_l) \leq D(\mathbf{x})/2$. We only need to consider the level sets whose levels were altered. Let x be any point in such a level set with height $h' \geq h$. $x \leq x_l$ by definition of the MVP (x_l, x_r). $x \geq x_r$, because if $x < x_r$, then $D(\mathbf{x}_l)/2 \geq h' - x > h - x_r = D(\mathbf{x})/2 \geq D(\mathbf{x}_l)/2$, which is a contradiction. Thus $x_r \leq x \leq x_l$ and so the error for any such point is at most $D(\mathbf{x})/2$ for the regression $\mathcal{C}'_l$. The error for all other points has remained unchanged and was originally at most $\mathcal{E}_\infty(\mathcal{C}_l) = D(\mathbf{x}_l)/2 \leq D(\mathbf{x})/2$, so we conclude that $\mathcal{E}_\infty(C'_l) \leq D(\mathbf{x})/2$. Similarily, consider all level sets of $\mathcal{C}_r$ with level $\leq h$. Increase *all* these levels to h and call this new isotonic regression $\mathcal{C}'_r$. Once again any point x in any level set with a level change must satisfy $x_r \leq x \leq x_l$ and so we conclude that $\mathcal{E}_\infty(\mathcal{C}'_r) \leq D(\mathbf{x})/2$.

Consider the regression $\mathcal{C}' = \mathcal{C}'_l \cup \mathcal{C}^* \cup \mathcal{C}'_r$. $\mathcal{E}_\infty(\mathcal{C}') = \max\{\mathcal{E}_\infty(\mathcal{C}'_l), \mathcal{E}_\infty(\mathcal{C}^*), \mathcal{E}_\infty(\mathcal{C}'_r)\} = D(\mathbf{x})/2$. The isotonic regression $\mathcal{C}$ is constructed from $\mathcal{C}'$ by taking the union of all level sets with the height h (these must be consecutive level sets), which does not alter the error. ■

Proposition 3.1 immediately yields a recursive algorithm to compute the isotonic regression. Unfortunately, this recursive algorithm would have a run time that is quadratic in n. We now show how to construct this regression from left to right, using a single pass. This will lead to a linear time algorithm for the prefix-isotonic regression problem. Let $\mathbf{x}_i = \mathbf{x}[1, i]$. Let $\mathcal{C}_i$ be an isotonic regression for $\mathbf{x}_i$. The prefix-isotonic regression is given by $\{\mathcal{C}_i\}_{i=1}^n$. Note that $\mathcal{E}_\infty(\mathcal{C}_{i+1}) \geq \mathcal{E}_\infty(\mathcal{C}_i)$ since $D(\mathbf{x}_{i+1}) \geq D(\mathbf{x}_i)$. We will construct $\mathcal{C}_{i+1}$ from $\mathcal{C}_i$.

Let $\mathcal{C}_i = \{I_\alpha = [i_\alpha, j_\alpha], h_\alpha\}_{\alpha=1}^K$. Let $\inf_\alpha = \min_{k \in I_\alpha} x_k$, and $\sup_\alpha = \max_{k \in I_\alpha} x_k$. Define the distortion of level set I_α, $D(I_\alpha)$ as the distortion of the sequence $\mathbf{x}[i_\alpha, j_\alpha]$. The $\mathcal{C}_i$ that we construct will all satisfy the following properties:

P1: $\forall \alpha \in [1, K]$, $h_\alpha = \frac{1}{2}(\sup_\alpha + \inf_\alpha)$.

P2: $\forall \alpha \in [1, K]$, $D(I_\alpha) = \sup_\alpha - \inf_\alpha$.

P3: $\forall \alpha \in [2, K]$, $h_{\alpha-1} < h_\alpha$.

Property $P3$ is just a restatement of the monotonicity condition. From property $P2$ it follows that for any $i \in I_\alpha$, $|x_i - h_\alpha| \leq D(I_\alpha)/2$. Since $D(I_\alpha) \leq D(\mathbf{x})$, it follows from Proposition 3.1 that any regression that has properties $P2$ and $P3$ is necessarily optimal. Therefore, properties $P1$-$P3$ are sufficient conditions for an isotonic regression. Suppose that $\mathcal{C}_i$ has been constructed, satisfying $P1$-$P3$. Now consider adding the point x_{i+1}. Let $I_{K+1} = \{i+1\}$, $h_{K+1} = x_{i+1}$. Note that $D(I_{K+1}) = 0$, and by construction, I_{K+1} satisfies $P1$ and $P2$.

Lemma 3.2 *If $h_{K+1} > h_K$, let $\mathcal{C}_{i+1} = \mathcal{C}_i \cup \{(I_{K+1}, h_{K+1})\}$. Then $\mathcal{C}_{i+1}$ satisfies $P1$-$P3$.*

If $h_{K+1} \leq h_K$, then to get $\mathcal{C}_{i+1}$, we merge I_{K+1} with I_K. We need to ensure that properties $P1$ and $P2$ continue to hold. We will prove this in general for any two consecutive level sets. Suppose that (I_k, h_k) and (I_{k+1}, h_{k+1}) both satisfy properties $P1$ and $P2$, and suppose that $h_{k+1} \leq h_k$. Define the new level set I'_k by

$$I'_k = I_k \cup I_{k+1} \qquad \inf'_k = \min(\inf_k, \inf_{k+1}) \qquad \sup'_k = \max(\sup_k, \sup_k)$$
$$h'_k = \tfrac{1}{2}(\inf'_k + \sup'_k)$$

Lemma 3.3 *I'_k satisfies properties $P1$ and $P2$.*

Proof: By construction, $P1$ is satisfied. We show that $D(I'_k) = \sup'_k - \inf'_k$, from which $P2$ follows.

Suppose that $\inf_{k+1} \leq \inf_k$. Thus, $\inf'_k = \inf_{k+1}$. Since the first maximum in I_{k+1} occurs before the last minimum in I_{k+1} (as I_{k+1} satisfies $P2$), and the maximum in I_k occurs before any point in I_{k+1}, it follows that the first maximum in I'_k occurs before its last minimum, thus I'_k satisfies $P2$.

Suppose, on the other hand, that $\inf_{k+1} > \inf_k$. Thus, $\inf'_k = \inf_k$. Since $h_{k+1} \leq h_k$, we have that $\sup_{k+1} + \inf_{k+1} \leq \sup_k + \inf_k \implies \sup_{k+1} < \sup_k$, and so $\sup'_k = \sup_k$. Thus, the first maximum in I'_k is the first maximum in I_k and the last minimum in I'_k is the last minimum in I_k. Since I_k satisfies $P2$ then so does I'_k. ∎

The idea of the algorithm should now be clear. The addition of a new point creates a new level set satisfying $P1$ and $P2$. If this new level set also satisfies $P3$, then we are done, and have constructed the isotonic regression for the sequence augmented by this one point. If not, then we merge the last two level sets, maintaining $P1$ and $P2$, and not altering any of the other level sets. We continue to merge until $P3$ is satisfied for the last level set, which must eventually happen. At this point we have a regression that satisfies $P1$-$P3$ and so it is the isotonic regression for the augmented sequence.

Note that I_K is the right most level set of $\mathcal{C}_i$, i.e., $I_K = [i_K, i]$. This rightmost level set is the union of i with some number (possibly zero) of the level sets (from right to left) of $\mathcal{C}_{i-1}$. The remaining level sets of $\mathcal{C}_i$ will be the level sets of $\mathcal{C}_{i-1}$ that remain after the merging. In fact, the remaining level sets will be exactly the level sets of $\mathcal{C}_{i_K-1}$, where it is understood that $\mathcal{C}_{i_K-1} = \{\}$ if $i_K = 1$.

Proposition 3.4 *$\mathcal{C}_i$=$\mathcal{C}_{i_K-1} \cup \{I_K, h_K\}$.*

Proof: If $i = 1$, there is nothing to prove. Assume that $i > 1$ and that the claim holds for all $\mathcal{C}_j$ with $j < i$. Let $\mathcal{C}_i = \{I_\alpha = [i_\alpha, j_\alpha], h_\alpha\}_{\alpha=1}^K$. By construction, C_{i-1} is given by

$$\mathcal{C}_{i-1} = \{(I_1, h_1), \ldots, (I_{K-1}, h_{K-1}), (S_1, h'_1), \ldots, (S_M, h'_M)\}, \qquad (*)$$

where M is possibly zero, and $I_K = \cup_i S_i \cup \{i\}$. Let $S_i = [\alpha_i, \beta_i]$, where $\alpha_1 = i_K$ and $\beta_M = i - 1$. By the induction hypothesis,

$$\begin{aligned}
\mathcal{C}_{i-1} &= \mathcal{C}_{\alpha_M - 1} \cup \{S_M, h'_M\}, \\
\mathcal{C}_{\alpha_M - 1} &= \mathcal{C}_{\alpha_{M-1} - 1} \cup \{S_{M-1}, h'_{M-1}\}, \\
\mathcal{C}_{\alpha_{M-1} - 1} &= \mathcal{C}_{\alpha_{M-2} - 1} \cup \{S_{M-2}, h'_{M-2}\}, \\
&\vdots \\
\mathcal{C}_{\alpha_2 - 1} &= \mathcal{C}_{\alpha_1 - 1} \cup \{S_1, h'_1\}.
\end{aligned}$$

Combining these equalities and using the fact that $\alpha_1 = i_K$, we get that

$$\mathcal{C}_{i-1} = \mathcal{C}_{i_K - 1} \cup_i \{S_i, h'_i\}.$$

using (*), we identify that $\mathcal{C}_{i_K - 1} = \{(I_1, h_1), \ldots, (I_{K-1}, h_{K-1})\}$, concluding the proof. ■

3.1 L_∞-Prefix-Isotonic Regression: Algorithms

Here, we will give the linear time algorithm for L_∞-prefix-isotonic regression that follows from the results of the previous section, along with the analysis of its run time. Our algorithm will process points from left to right. After processing the new point x_i, we will have constructed the isotonic regression $\mathcal{C}_i$ as discussed in the previous section by merging the rightmost two intervals until $P1$-$P3$ are satisfied.

By Proposition 3.4, to reconstruct $\mathcal{C}_i$, we only need to know l_i, the index of the first point of its rightmost level set, the level, h_i, of this rightmost level set, and how to construct $C_{l_i - 1}$. This can be recursively achieved by only storing the parameters l_i and h_i, for every i. The algorithms are given in Algorithm 3. The correctness of this algorithm follows from the results of the previous section, specifically Lemmas 3.2, 3.3. Further, $\mathcal{E}_\infty(\mathcal{C}_i)$ is stored in $D[i]$. By Proposition 3.4, the output of the algorithm stores all the necessary information to extract $\mathcal{C}_m$ as shown in the recursive algorithm Reconstruct.

What remains is to analyse the computational complexity of the algorithms. First consider the prefix-isotonic regression. Lines 7,8,13 constitute 8 operations, thus contributing about $8n$ operations to the total run time. The merging while loop, lines 9-12, uses 6 operations. The maximum number of intervals is n. Each time a merge occurs, this maximum drops by 1. Since this maximum is bounded below by 1, this means that there are at most $n - 1$ merges, so the *total* time spent merging is about $6n$ operations, and the condition of the while loop is checked at most $2n$ times, so the runtime of this algorithm is bounded by Cn where $C \approx 14$. There are at most n level sets at any time, and each level set needs to store 5 numbers, $i_\alpha, j_\alpha, \inf_\alpha, \sup_\alpha, h_\alpha$. The additional space for L, H, D is $3n$, for a total memory requirement bounded by $C'n$, where $C' \approx 8$.

It is not hard to analyse the recursion for Reconstruct, and a straightforward induction shows that the runtime is $O(m)$. We therefore have the following result.

Algorithm 3 Algorithms for L_∞ prefix-isotonic regression.

// **Algorithm to perform L_∞-Prefix-Isotonic Regression.**
// **Input:** $\mathbf{x} = \{x_i | i \in [1, n]\}$.
// **Output:** L, H, D. $\{L[i] = l_i$, $H[i]$=level of $[l_i, i]$ in $\mathcal{C}_i$, $D[i]$=distortion of $\mathbf{x}_i\}$
$I_1 = [1, 1]$, $\inf_1 = x_1$, $\sup_1 = x_1$, $h_1 = x_1$, $K = 1$; {Initialization}
$L[1] = 1$, $H[1] = h_1$, $D[1] = 0$; {Initialization of outputs}
for $i = 2$ to n **do**
  $K \leftarrow K + 1$
  $I_K = [i, i]$, $\inf_K = x_i$, $\sup_K = x_i$, $h_K = x_i$, $D[i] = D[i-1]$;
  while $h_K \le h_{K-1}$ and $1 < K$ **do**
    $I_{K-1} \leftarrow I_{K-1} \cup I_K$; $\inf_{K-1} \leftarrow \min(\inf_{K-1}, \inf_K)$; $\sup_{K-1} \leftarrow \max(\sup_{K-1}, \sup_K)$;
    $K \leftarrow K - 1$; $h_K = \frac{1}{2}(\inf_K + \sup_K)$; $D[i] = \max(D[i], \sup_K - \inf_K)$;
  end while
  $L[i]$=left endpoint of I_K; $H[i] = h_K$;
end for

Reconstruct(m)
// **Output** $\mathcal{C}_m$, the isotonig regression for $\mathbf{x}_m$, assuming L, H are global.
if $m = 0$ **then**
  return $\{\}$;
end if
return Reconstruct$(L[m] - 1) \cup \{[L[m], m], H[m]\}$;

Theorem 3.5 *The L_∞ prefix-isotonic regression (L, H, D) can be found using at most Cn operations ($C \approx 14$) using additional space that is at most $C'n$ ($C' \approx 8$). Given L, H and $m > 0$, the L_∞ isotonic regression $\mathcal{C}_m$ for the prefix $\mathbf{x}[1, m]$ can be constructed using $O(m)$ operations.*

4 L_∞ Unimodal Regression

As pointed out in [15], a prefix-isotonic regression can easily be modified to yield the optimal unimodal regression. The next proposition shows that the crossover point in the L_∞ unimodal regression can always be chosen at a maximum in the sequence (any maximum). Thus, a simpler algorithm that follows directly from the prefix-isotonic regression is to first find a maximum in $\mathbf{x}$ (linear time). Now perform isotonic regression on the sequence to the left of the maximum and the reversal of the sequence to the right. More specifically, suppose that the maximum is x_m. Now consider the sequences $\mathbf{x}_l = \mathbf{x}[1, m]$, $\mathbf{x}_r = \mathbf{x}[m, n]$, and let $\mathbf{x}_r^R$ be the reversal of $\mathbf{x}_r$. Let $\mathcal{C}_l$ and $\mathcal{C}_r^R$ be the isotonic regressions for $\mathbf{x}_l$ and $\mathbf{x}_r^R$ respectively. Then the union, $\mathcal{C}_l \cup \mathcal{C}_r$ (where $\mathcal{C}_r$ is the reversal of $\mathcal{C}_r^R$) is the unimodal regression, with the merging of the last level set of $\mathcal{C}_l$ and the first level set of $\mathcal{C}_r$, as they will have the same level,

equal to the maximum. All that remains is to prove that the crossover point can always be chosen at a maximum.

Proposition 4.1 *The crossover point in the unimodal regression of* $\mathbf{x}$ *can always be chosen to be a maximum (any maximum) of* $\mathbf{x}$*.*

Proof: Let $\mathcal{C}$ be the unimodal regression, and let x_i be the crossover point, so

$$w_1 \leq w_2 \leq \cdots \leq w_i \geq w_{i+1} \geq \cdots \geq w_n.$$

Let $\mathbf{x}_l = \mathbf{x}[1,i]$, $\mathbf{x}_r = \mathbf{x}[i,n]$. Since $\mathbf{w}[1,i]$ is an isotonic regression for $\mathbf{x}_l$ and $\mathbf{w}^R[1,n-i+1]$ is an isotonic regression for $\mathbf{x}_r^R$, the error of the regression is $\mathcal{E}_\infty(\mathcal{C}) \geq \frac{1}{2}\max(D(\mathbf{x}_l), D(\mathbf{x}_r^R))$. Let x_m be any maximum not equal to x_i (if x_i is a unique maximum, then we are done, otherwise x_m exists). Without loss of generality, since a unimodal regression for $\mathbf{x}^R$ is $\mathcal{C}^R$, we can suppose that $m > i$. Let $\mathbf{x}_1 = \mathbf{x}[1,m]$, let $\mathbf{x}_2 = \mathbf{x}[m,n]$, and let $\mathbf{x}_c = \mathbf{x}[i,m]$. For the unimodal regression constructed from the two isotonic regressions on $\mathbf{x}_1$ and $\mathbf{x}_2^R$, x_m will be a crossover point. We show that the error of this regression cannot be more than the error of $\mathcal{C}$. The error of this regression is given by $\max(D(\mathbf{x}_1), D(\mathbf{x}_2^R))$. Since x_m is a maximum, $D(\mathbf{x}_r^R) = \max(D(\mathbf{x}_c^R), D(\mathbf{x}_2^R))$, so $\mathcal{E}_\infty(\mathcal{C}) = \max(D(\mathbf{x}_l), D(\mathbf{x}_c^R), D(\mathbf{x}_2^R))$. $D(\mathbf{x}_1)$ is given by

$$\begin{aligned} D(\mathbf{x}_1) = &= \max_{1\leq k\leq m}\{\max(\mathbf{x}[1,k]) - x_k\} \\ &= \max\left(\max_{1\leq k\leq i}\{\max(\mathbf{x}[1,k]) - x_k\}, \max_{i\leq k\leq m}\{\max(\mathbf{x}[1,k]) - x_k\}\right) \end{aligned}$$

The first term on the right hand side is $D(\mathbf{x}_l)$. Since x_m is a maximum, the second term is bounded by $\max_{i\leq k\leq m}\{x_m - x_k\} = D(\mathbf{x}_c^R)$. Thus $D(\mathbf{x}_1) \leq \max(D(\mathbf{x}_l), D(\mathbf{x}_c^R))$, and so

$$\max(D(\mathbf{x}_1), D(\mathbf{x}_2^R)) \leq \max(D(\mathbf{x}_l), D(\mathbf{x}_c^R), D(\mathbf{x}_2^R)) = \mathcal{E}_\infty(\mathcal{C}).$$

■

We therefore have the following result,

Theorem 4.2 *The* L_∞ *unimodal regression can be found using* $O(n)$ *operations.*

5 Conclusion

For L_1-isotonic regression we presented an output sensitive algorithm whose running time is linear in n when the number of possible values that the levels of the isotonic regression can take is bounded by K. In the worst case, $K = n$ and the algorithm is no worse than existing algorithms. The open question that remains is whether the median isotonic regression can be computed in linear time, or to prove that it cannot. Our algorithms can be extended without much effort to the case of minimizing a weighted L_1 error. In this case, all the results remain true, with minor modifications, by replacing the standard median with the weighted median.

For L_∞ isotonic and unimodal regression, we have given simple (not requiring sophisticated data structures) linear time algorithms.

References

[1] M. Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid, and E. Silverman. An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics*, 1955.

[2] R. Barlow and V. Ubhaya. Isotonic approximation. In J. S. Rustagi, editor, *Optimizing Methods in Statistics*. Academic Press, New York, 1971.

[3] N. Chakravarti. Isotonic median regression, a linear programming approach. *Math. of Oper. Research*, 14(2):303–308, 1989.

[4] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms*. Mcgraw-Hill, Cambridge, MA, 2nd edition, 2001.

[5] M. Frisén. unimodal regression. *The Statistician*, 35:304–307, 1980.

[6] Z. Geng and N.-Z. Shi. Isotonic regression for umbrella orderings. *Applied Statistics*, 39:397–424, 1990.

[7] M. Magdon-Ismail, J. H.-C. Chen, and Y. S. Abu-Mostafa. The multilevel classification problem and a monotonicity hint. *Intelligent Data Engineering and Learning (IDEAL 02), Third International Conference*, August 2002.

[8] R. A. Mureika, T. R. Turner, and P. C. Wollan. An algorithm for unimodal isotonic regression with application to locating a maximum. Technical Report 92-4, Department of Mathematics and Statistics, University of New Brunswick, 1992.

[9] G. Pan. Subset selection with additional order information. *Biometrics*, 52:1363–1374, 1996.

[10] P. M. Pardalos and G.-L. Xue. Algorithms for a class of isotonic regression problems. *Algorithmica*, 23:211–222, 1999.

[11] P. M. Pardalos, G.-L. Xue, and L. Yong. Efficient computation of an isotonic median regression. *Appl. Math. Lett.*, 8(2):67–70, 1995.

[12] T. Robertson and P. Waltman. On estimating monotone parameters. *Ann. Math. Stat.*, pages 1030–1039, 1968.

[13] T. Robertson, F. T. Wright, and R. L. Dykstra. *Order Restricted Statistical Inference*. Wiley Series in Probability and Statistics. Wiley, new York, 1988.

[14] J. Sill and Y. S. Abu-Mostafa. Monotonicity hints. In M. C. Mozer, M. I. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 9, pages 634–640. Morgan Kaufmann, 1997.

[15] Q. F. Stout. Optimal algorithms for unimodal regression. *Computing Science and Statistics*, 32, 2000.

[16] V. Ubhaya. Isotone optimization I. *Journal of Approximation Theory*, 12:146–159, 1974.

[17] V. Ubhaya. Isotone optimization II. *Journal of Approximation Theory*, 12:315–331, 1974.