L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

Polynomial algorithms for isotonic regression

Victor Chepoi

Universitatea de stat din Moldova, Chişinău, Moldavia and Université d'Aix Marseille II, France

Daniel Cogneau and Bernard Fichet

Université d'Aix Marseille II, France

Abstract: In this paper we consider the following problem. Let $X = \{x_1, \ldots, x_n\}$ be a set of observation points endowed with a partial order $\prec$, and let $y_1, \ldots, y_n$ be the values of the dependent variable y. We are searching an isotonic function $f : X \to \mathbb{R}$ (i.e. $x_i \prec x_j$ implies that $f(x_i) \leq f(x_j)$) that minimizes the l_p–error

$$D_p(f) = [\sum_{x_i \in X} |y_i - f(x_i)|^p]^{\frac{1}{p}}.$$

We recall some general algorithms for solving this and related regression problems and we present new polynomial algorithms for some versions of the isotonic regression problem.

Key words: L_p-norm, polynomial algorithm, curve fitting, graphs, partial ordering.

AMS subject classification: 62J05.

1 Introduction

The basic isotonic regression problem can be formulated as follows: given values $y_1, \ldots, y_n$ of the dependent variable y, corresponding to values $x_1, \ldots,$ x_n of the independent variable x, which constitute a set X with a partial order $\prec$ (i.e., a reflexive, transitive and antisymmetric binary relation on X), fit to the y_i a best function $y = f(x)$ which is non–decreasing (alias isotonic) with respect to $\prec$. The error norm usually chosen is l_2 : we are seeking an isotonic function f on X that minimizes

$$D_2(f) = [\sum_{i=1}^{n}(y_i - f(x_i))^2]^{\frac{1}{2}}.$$

Algorithms for this problem have received a great deal of attention and a collection of them have been discussed in details in [3, 11, 25]. In the case when $\prec$ is a total order on X all of the algorithms work in linear time $O(n)$ provided $\prec$ is given. In particular we mention the simple Pool–Adjacent–Violators algorithm introduced by Ayer et al. [1] and popularized by Kruskal [18] under the name Up–and–Down Blocks algorithm. It has been extended to rooted trees by Thompson [31] with the Minimum Violator algorithm. The Pool–Adjacent–Violators algorithm is also implicit in van Eeden [9], who extended in [10] the procedures to regressions bounded by two given functions.

The situation becomes much more complex if $\prec$ is a partial order, say $x_1, \ldots, x_n$ are points in the d–dimensional space and $x_i \prec x_j$ if and only if $x_i = (x_i^{(1)}, \ldots, x_i^{(d)})$, $x_j = (x_j^{(1)}, \ldots, x_j^{(d)})$ and $x_i^{(k)} \leq x_j^{(k)}$ for each $1 \leq k \leq d$. Although some algorithms described in [3] are applicable in the general case, their computational complexity is already exponential; see [8]. A convergent numerical algorithm for $X \subset \mathbb{R}^2$ has been proposed in [8]. Geometrically, one can formulate the l_2–regression problem as the computing of the projection of the vector $y = (y_1, \ldots, y_n)$ onto the convex cone K of the isotonic functions on X. Since K is defined by a finite number of constraints, it is polyhedral. Therefore, one can use any algorithm for solving a quadratic optimization problem, especially those established for projecting onto polyhedral cones, as, for example, that presented in [19]. The alternative procedure of Dykstra [7] is also efficient for polyhedral cones. However, several specific algorithms have been developed for regression problems. The reader will find a vast literature on this topic.

Among other criteria, the choice of l_2 as an error norm is due to the connection with special estimates. Usually, the estimates studied in order restricted statistical inference can be expressed by compact "max–min" formulas (of course, being an useful tool in consistency proofs, these formulas are not very appropriate for computing of the estimates). For instance, let $M(A)$ be the mean of a collection A of observations taken from a poset X. Now, a subset L of X is called an *upper layer* if $x_i \in L$ and $x_i \prec x_j$ imply that $x_j \in L$. Then

$$\hat{f}(x_i) = \max_{\{L:\ x_i \in L\}}\ \min_{\{L':\ x_i \notin L'\}}\ M(L - L')$$

is an isotonic function and could be used as an estimate. It is shown in [3] that $\hat{f}$ provides the best l_2–approximation of $y_1, \ldots, y_n$ and can be calculated by a special minimum lower sets algorithm introduced by Brunk et al. [5] (the complexity of the latter is exponential). A deep generalization of this result to all Cauchy mean value functions has been obtained in [28] (a function M defined on the nonempty subsets of X is said to be a *Cauchy mean value function* [27, 28] if $M(A \cup B)$ belongs to the segment $[M(A), M(B)]$ whenever A and B are nonempty and disjoint subsets of X). A description of all linear Cauchy mean value functions has been given in [21]. The result from [28] asserts that if M is such a function and the error measure $D(f)$ verifies three rather natural conditions, then $\hat{f}$ given by the abovementioned "min–max" formula minimizes $D(f)$ subject to the restriction that f is isotonic on X. Moreover, $\hat{f}$ can be computed using a refined version of the minimum lower set algorithm. As is noticed in [28], this result includes all l_p–regression problems ($1 \leq p \leq \infty$) in their most general form as special cases (however, the modal regression problem [29] does not fit in this framework).

Namely, assume that with each element x_i of a poset $(X, \prec)$ is associated a set of numbers $y_{i1}, \ldots, y_{ir_i}$, corresponding, for example, to a sample from the ith distribution. We are looking for an isotonic function f on X (i.e., $x_i \prec x_j$ implies $f(x_i) \leq f(x_j)$) that minimizes

$$D_p(f) = [\sum_{i=1}^{n} \sum_{l=1}^{r_i} (y_{il} - f(x_i))^p]^{\frac{1}{p}}.$$

If $p = 1$ we obtain the isotonic median regression problem that corresponds to the l_1–error norm:

$$D_1(f) = \sum_{i=1}^{n} \sum_{l=1}^{r_i} |y_{il} - f(x_i)|.$$

For $X \subset \mathbb{R}$ this problem have been investigated in [26, 24] (unfortunately, the algorithm presented in [24] contains a serious gap, since its two steps do not cover all possible cases). The minimum lower set algorithm from [28], acting on the collection of upper layers, has an exponential complexity already if this collection is of exponential cardinality. This is the case of rather simple partial orders as rooted trees or series–parallel partial orders. To our knowledge there are no polynomial time algorithms to find a best isotonic l_p–regression function ($p < \infty$) in the general case of a partial order or even in case when $X \subset \mathbb{R}^d$, $d \geq 2$. It will be interesting and important to know *what instances of this problem are NP–complete algorithmical problems.*

2 Isotonic l_p–regression problem for rooted trees

This section is devoted to the isotonic l_p–regression problem ($1 \leq p \leq \infty$) for partial orders whose covering graphs are rooted trees. We propose a simple extension of the maximum (minimum) violator algorithm, originally established for the l_2–criterion. Let us consider a rooted tree, the vertices of which are the elements $x_1, \dots, x_n$ of a set X. The tree is here oriented from the leaves to the root. To every vertex x_i, a sample $y_{i1}, \dots, y_{ir_i}$ is preassigned. We consider the l_p–regression problem defined in the introduction: minimize $D_p(f)$ subject to the constraint that f is isotonic. The y_{il} define a vector of $\mathbb{R}^r$, where $r = \sum_{i=1}^{n} r_i$. Noting $\theta_i = f(x_i)$ and duplicating r_i times the θ_i, we get a current vector $\theta \in \mathbb{R}^r$, and the minimization problem may be written as: minimize $||y - \theta||_p$, with equality and inequality constraints over θ. Such a problem corresponds to the projection onto a closed polyhedral cone, with respect to a norm l_p. That provides the existence of a solution.

For every sample of real numbers $A = \{u_1, \dots, u_k\}$, let us consider the problem: $\min_x [\sum_{j=1}^{k} |x - u_j|^p]^{\frac{1}{p}}$. It is well–known that such a problem admits a solution. This is the mean for $p = 2$, the midrange for $p = \infty$ and a median point for $p = 1$. Every solution lies between $\min u_j$ and $\max u_j$, and due to the convexity, the set of solutions is a closed interval. We denote it by $M(A) = [a, b]$. Moreover, it is easy to see that $[\sum_{j=1}^{k} |x - u_j|^p]^{\frac{1}{p}}$ is strictly decreasing when x varies from $-\infty$ to a, and is strictly increasing when x varies from b to $+\infty$. Again, given two samples A and A', we denote by $A + A'$ their amalgamation. Let $M(A) = [a, b]$, $M(A') = [a', b']$, $M(A + A') = [c, d]$. Then, clearly $M(A + A') = M(A) \cap M(A')$ provided $M(A) \cap M(A') \neq \emptyset$. Moreover, it is easy to prove that M obeys the Cauchy mean condition: if $M(A) \cap M(A') = \emptyset$ with, for instance, $b < a'$, then $b \leq c$ and $d \leq a'$.

Now we describe our algorithm. At a current step, we have a rooted tree with a partition of X, say V, as a set of vertices. To every vertex v_i of V, a sample A_i is assigned, where A_i stands for the amalgamation of the samples $\{y_{jl}, l = 1, \dots, r_j\}$ for all j such that $x_j \in v_i$. Initially, V is the finest partition of X. For every $v_i \in V$, denote $M(A_i) = [a_i, b_i]$. If for every edge (v_i, v_j) of the current tree with $v_i \prec v_j$ we have $a_i < a_j$, then put $\hat{f}(x_k) = a_i$ for every $x_k \in v_i$ and stop. Otherwise, find $v_0 \in V$ with predecessors $v_1, \dots, v_m$ obeying the following conditions

(i) $a_0 \leq a_{max} := \max_{i=1,\dots,m} a_i$;
(ii) for all $i = 1, \dots, m$ if $v_l \prec v_i$ then $a_l < a_0$.

Then aggregate v_0 and all v_i such that $a_i = a_{max}$, and amalgamate the corresponding samples A_0 and A_i. We get a new partition of X with

$v_0 := v_0 \cup \{v_i : a_i = a_{max}\}$ and all other subsets of V, and a new rooted tree, to which we apply the same procedure.

Proposition 1 *$\hat{f}$ obtained by this algorithm minimizes the l_p–criterion function $D_p(f)$ subject to the restriction that f is isotonic. The function $\hat{f}$ can be defined in total $O(nr)$ number of operations.*

Proof: The current step corresponds to the new isotonic l_p–regression problem:

(P) : minimize $[\sum_{v_i \in V} \sum_{y \in A_i} |\theta_i - y|^p]^{\frac{1}{p}}$, with the constraints: $v_i \prec v_j$ implies $\theta_i \leq \theta_j$.

We prove by induction that every solution $\hat{\theta}$ of the reduced problem furnishes a solution $\hat{f}$ of the initial problem by letting $\hat{f}(x_k) = \hat{\theta}_i$ provided $x_k \in v_i$. Clearly, that is true at the initial step. Suppose that at a current step, there is v_0 obeying the conditions of the algorithm.

First, we show that there is a solution $\hat{\theta}$ of (P) verifying $\hat{\theta}_0 = \hat{\theta}_j$ for some $j \in \{1, \ldots, m\}$. Suppose there is a solution θ^* of (P) verifying: $\theta_0^* > \theta_i^*$ for all $i = 1, \ldots, m$. Then $a_i \leq \theta_i^* < \theta_0^* \leq b_0$ for all $i = 1, \ldots, m$. Indeed, if for some i, $\theta_i^* < a_i$, defining θ' by $\theta'_i := \theta_i^* + \delta$ and $\theta' := \theta^*$ otherwise, θ' should satisfy the isotony constraints and should reduce the value of the criterion for $\delta > 0$ sufficiently small. Similarly, if $\theta_0^* > b_0$, defining θ'' by $\theta''_0 := \theta_0^* - \delta$ and $\theta'' := \theta^*$ otherwise, θ'' should satisfy the constraints and should reduce the value of the criterion for $\delta > 0$ sufficiently small, establishing our assertion.

Let θ_j^* realize $\max\{\theta_i^* : i = 1, \ldots, m\}$. Then $a_0 \leq \max a_i \leq \theta_j^* < \theta_0^* \leq b_0$. Define $\hat{\theta}$ by $\hat{\theta}_j := \theta_j^*$ and $\hat{\theta} = \theta^*$ otherwise. Then clearly $\hat{\theta}$ satisfies the isotony constraints and preserves the value of the criterion. Thus, $\hat{\theta}$ is a solution of a problem obtained from (P) by adding another constraint $\theta_0 = \theta_j$, which is clearly equivalent to a reduced problem of (P) by aggregating v_0 and v_j, by amalgamating A_0 and A_j to A'_0 and by joining the predecessors of v_j to $v'_0 := v_0 \cup v_j$. The hypotheses show that $a'_0 \leq a_{max}$, where $M(A'_0) = [a'_0, b'_0]$. By induction and previous assertion one can deduce that there is a solution $\overline{\theta}$ of (P) verifying $\overline{\theta_0} = \overline{\theta_i}$ for every i such that $a_i = a_{max}$. Since every solution of the reduced problem defined by the algorithm is a solution of (P) satisfying those new constraints, the induction is proved. Moreover, at the end of the algorithm, $\hat{\theta}$ defined by $\hat{\theta}_i := a_i$ is clearly a solution of (P). □

The condition required in the algorithm for pooling two adjacent vertices indicates that we must start from leaves and proceed in an up–and–down way. However, when we have a total order, the proof shows that we may aggregate any pair of consecutive vertices violating isotony condition, as in the Pool–Adjacent–Violator algorithm.

3 Isotonic l_∞–regression problem and its variants

In this section we consider the isotonic regression problem with the l_∞–error norm (i.e. the well–known uniform or Chebychev measure of error); for algorithmic approaches to similar approximation problems see [4, 15, 14] and the references there. For this problem we present a strikingly simple optimal estimate which can be computed in time proportional to the size of the covering graph of the poset $(X, \prec)$. If $X \subset \mathbb{R}^d$ and $|X| = n$ the computational complexity to compute this estimate is $O(dn^2)$. The result is due to Ubhaya [32, 33] and was rediscovered by one of the authors of this note. Also we consider the p–isotonic regression problem with the Chebychev norm and show how to reduce it to a graph–theoretical problem. In the particular case $p = 2$ this allows to present a polynomial algorithm.

3.1 Isotonic l_∞–regression problem

Let $X = \{x_1, \ldots, x_n\}$ be a set of observation points endowed with a partial order $\prec$ and let $y_1, \ldots, y_n$ be the corresponding values of the dependent variable y. The simplest (and the most economical way) to present a partial order is to define its *covering graph* $G = (X, E)$: in G two elements x_i and x_j are joined by an arc if $x_i \prec x_j$ and there is no other element x_k so that $x_i \prec x_k \prec x_j$. The goal of the isotonic regression problem with the l_∞–norm is to determine an isotonic function $f : X \to \mathbb{R}$ that minimizes the l_∞–error

$$D_\infty(f) = \max_{x_i \in X} |y_i - f(x_i)|$$

(in [12] the history of using this and other criteria in estimation procedures has been discussed).

For an element $x_i \in X$ consider the order ideals

$$L_\prec(x_i) = \{x_j \in X : x_j \prec x_i\}, L_\succ(x_i) = \{x_j \in X : x_i \prec x_j\}.$$

Let

$$f^*(x_i) = \max\{y_j : x_j \in L_\prec(x_i)\}$$

and

$$f_*(x_i) = \min\{y_j : x_j \in L_\succ(x_i)\}.$$

The reflexivity of $\prec$ implies that $L_\prec(x_i) \cap L_\succ(x_i) = \{x_i\}$. In particular, $f^*(x_i) \geq f_*(x_i)$.

Proposition 2 *The function $\hat{f}(x_i) = \frac{1}{2}(f^*(x_i) + f_*(x_i))$ minimizes $D_\infty(f)$ subject to the restriction that f is isotonic. Given the covering graph G,* *the values of $\hat{f}$ can be computed in total time $O(|E|)$. If $X \subset \mathbb{R}^d$, then the computation of $\hat{f}$ can be performed in $O(dn^2)$ time.*

Proof: Let (x_i, x_j) be an arbitrary edge of G, and assume $x_i \prec x_j$. Since $L_\prec(x_i) \subset L_\prec(x_j)$ and $L_\succ(x_j) \subset L_\succ(x_i)$, we conclude that $f^*(x_j) \geq f^*(x_i)$ and $f_*(x_j) \geq f_*(x_i)$, yielding that $\hat{f}$ is isotonic on X.

Suppose by way of contradiction that $D_\infty(g) < D_\infty(\hat{f})$ for an isotonic function g. Let $\epsilon = D_\infty(\hat{f})$ and consider an element x_i such that $\epsilon = |\hat{f}(x_i) - y_i|$. Let $f^*(x_i) = y_j$ and $f_*(x_i) = y_k$ for elements $x_j \in L_\prec(x_i)$ and $x_k \in L_\succ(x_i)$. Suppose without loss of generality that y_i belongs to the segment $[y_k, \hat{f}(x_i)]$. Since $|g(x_i) - y_i| < \epsilon$ and $g(x_j) \leq g(x_i)$ we immediately obtain that $g(x_j) < \hat{f}(x_i)$. But then $|g(x_j) - y_j| > y_j - \hat{f}(x_i) > \epsilon$, contrary to the choice of g.

The values of f^* can be computed recursively, starting from the minimal elements of X. If we know f^* for all predecessors of x_i, then $f^*(x_i)$ is the maximum among the y_i and $\max\{f^*(x_j) : (x_j, x_i) \in E\}$. Analogously, the values of f_* can be computed recursively starting from the maximal elements of X. If f_* is computed for all successors of x_i, then $f_*(x_i)$ is the minimum among the y_i and $\min\{f_*(x_j) : (x_i, x_j) \in E\}$. Evidently, this can be done in $O(|E|)$ time. If $X \subset \mathbb{R}^d$, then the covering graph of the resulting poset can be computed in $O(dn^2)$ time. □

In Figure 1 we present an example of application of Proposition 2 (the optimal error is $\epsilon^* = 4$).

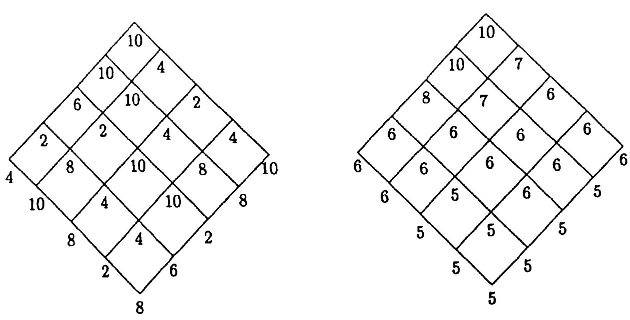

FIGURE 1.

With few efforts one can present an optimal estimate to the general isotonic l_∞–regression problem. Assume as before that with each element x_i of $(X, \prec)$ is associated a set of (distinct) numbers $y_{i1}, \ldots, y_{ir_i}$, and we wish to find an isotonic function f on X that minimizes

$$D_\infty(f) = \max_{x_i \in X} \max_{l=1,\ldots,r_i} |y_{il} - f(x_i)|.$$

For an element $x_i \in X$ set

$$f^*(x_i) = \max_{x_j \in L_{\prec}(x_i)} \max_{l=1,\ldots,r_j} \; y_{jl}$$

and

$$f_*(x_i) = \min_{x_j \in L_{\succ}(x_i)} \min_{l=1,\ldots,r_j} \; y_{jl}.$$

We assert that the function $\hat{f}(x_i) = \frac{1}{2}(f^*(x_i) + f_*(x_i))$ minimizes $D_\infty(f)$ subject to the restriction that f is isotonic. For this we extend the partial order $\prec$ from X to the multiset $\{y_{il} : i = 1,\ldots,n, l = 1,\ldots,r_i\}$: set $y_{il} \prec y_{jt}$ if and only if $x_i \prec x_j$ or $i = j$ and $y_{il} \geq y_{jt}$. Let $\hat{f}(y_{il})$ be the function defined as in Proposition 2. One can easily note that $\hat{f}(y_{il}) = \hat{f}(x_i)$ for all y_{il} $(l = 1,\ldots,r_i)$. From this and Proposition 2 we deduce that $\hat{f}(x_i)$ minimizes $D_\infty(f)$. The values of $\hat{f}$ can be computed in $O(|E| + \sum_{i=1}^n r_i)$ number of operations.

3.2 p–Isotonic l_∞–regression problem

In some recent papers [22, 2, 20] new generalizations of the classical linear regression problem have been given. For example, [20] presents an efficient algorithm for partitioning a planar set $S = \{s_1 = (x_1, y_1), \ldots, s_n = (x_n, y_n)\}$ into two parts S_1 and S_2 such that

$$\sum_{s_i \in S_1} (y_i - f_1(x_i))^2 + \sum_{s_i \in S_2} (y_i - f_2(x_i))^2$$

is minimized, where f_1 and f_2 are the regression lines of the sets S_1 and S_2 (the multidimensional case is treated in [16]). Agarwal and Sharir [2] presented an algorithm with complexity $O(n^2 log^5 n)$ for solving a similar problem, replacing the l_2–criterion function by the l_∞–error function. Namely, they are searching a bipartition of a planar set S such that their maximum width is as small as possible. Recall that the width of a set is the smallest distance between a pair of parallel supporting lines. Equivalently, it is neccesary to find two linear functions f_1 and f_2, such that $\max_{s_j \in S}\{\min_{i=1,2}|y_j - f_i(x_j)|\}$ is minimized (one can formulate this problem for p–partitions as is done in [22]). For isotonic regressions, this leads us to the following general formulation.

As before let $X = \{x_1,\ldots,x_n\}$ be a set with a partial order $\prec$, and let $y_1,\ldots,y_n$ be the corresponding values of the variable y. We wish to find a partition $X_1,\ldots,X_p$ of X and the isotonic functions $f_1,\ldots,f_p$ on $X_1,\ldots,X_p$, respectively, such that the l_∞–error

$$D_\infty(f_1,\ldots,f_p) = \max\{\max_{x_i \in X_1}|y_i - f_1(x_i)|,\ldots,\max_{x_i \in X_p}|y_i - f_p(x_i)|\}$$

is minimized. Below we will show how to reduce this problem to a special graph–theoretic problem and how to solve it efficiently for $p = 2$. Define a symmetric matrix $D = (d_{ij})$, where $d_{ij} = \frac{1}{2}(y_i - y_j)$ if $x_i \prec x_j$ and $y_i \geq y_j$, and $d_{ij} = 0$ otherwise. Let ϵ^* be the minimum of the function $D_\infty(f_1, \ldots, f_p)$. From Proposition 1 we obtain the following result.

Lemma 1 *ϵ^* is an element of the matrix D.*

To find an optimal partition of X with respect to the criterion function $D_\infty(f_1, \ldots, f_p)$ we proceed as follows (the idea is borrowed from the methods of solving center location problems; see for example [17, 30]). We sort the elements of the matrix D in the increasing order and search the obtained list for the minimum value which is feasible in the following sense. A value ϵ is *feasible* if there is a p–partition $X_1, \ldots, X_p$ of X and the isotonic functions $f_1, \ldots, f_p$ on $X_1, \ldots, X_p$, respectively, such that $D_\infty(f_1, \ldots, f_p) \leq \epsilon$. To decide if a value $\epsilon \in D$ is feasible we define a new graph Γ_ϵ. The vertices of Γ_ϵ are the elements of X, and two vertices x_i and x_j are adjacent in Γ_ϵ if and only if either x_i and x_j are incomparable or $x_i \prec x_j$ are comparable and $y_i - y_j \leq 2 \cdot \epsilon$. A *clique* of Γ_ϵ is a subset of pairwise adjacent vertices.

Lemma 2 *ϵ is a feasible value if and only if the vertices of Γ_ϵ can be covered with at most p cliques.*

Proof: First, assume that $D_\infty(f_1, \ldots, f_p) \leq \epsilon$ for isotonic functions $f_1, \ldots, f_p$ defined on classes $X_1, \ldots, X_p$ of a partition of X. We assert that each X_k is a clique of the graph Γ_ϵ. Assume the contrary, i.e. $y_i - y_j > 2 \cdot \epsilon$ for some $x_i, x_j \in X_k$, $x_i \prec x_j$. We can suppose without loss of generality that y_j is the smallest value in $\{y_s : x_s \in L_\succ(x_i)\} \cap X_k$. Additionaly, we can assume that f_k if defined as in Proposition 2. By this result $f_k(x_i) = \frac{1}{2}(y_j + y_t)$, where y_t is the largest value in $\{y_s : x_s \in L_\prec(x_i)\} \cap X_k$. Since $|y_i - f_k(x_i)| \leq \epsilon$ and $f_k(x_i) \leq f_k(x_j)$, from $y_i - y_j > 2 \cdot \epsilon$ one can easily deduce that $|y_j - f_k(x_j)| > \epsilon$, contrary to feasibility of ϵ. Therefore, if ϵ is feasible, then $X_1, \ldots, X_p$ are cliques of the graph Γ_ϵ.

Conversely, let $X_1, \ldots, X_{p'}$ be a covering of the vertices of Γ_ϵ with p' cliques ($p' \leq p$). Let $\hat{f}_k$ be the isotonic function on X_k defined in Proposition 2. Pick an arbitrary element $x_i \in X_k$. Then $\hat{f}_k(x_i) = \frac{1}{2}(y_j + y_t)$ where y_j is the smallest value in $\{y_s : x_s \in L_\succ(x_i)\} \cap X_k$ and y_t is the largest value in $\{y_s : x_s \in L_\prec(x_i)\} \cap X_k$. Since $x_t \prec x_i \prec x_j$ and $x_t, x_j \in X_t$ we deduce that $0 \leq y_j - y_t \leq 2 \cdot \epsilon$. Since $y_i \in [y_t, y_j]$ we immediately obtain that $|y_i - \hat{f}_k(x_i)| \leq \epsilon$. Therefore $D_\infty(\hat{f}_1, \ldots, \hat{f}_{p'}) \leq \epsilon$, i.e. ϵ is feasible. $\square$

The problem of covering of a graph with a given number of cliques is known to be NP–complete [13], however in the particular case $p = 2$ it can be easily solved. Indeed, a graph can be covered with two cliques if and only if its complement is bipartite. To find a bipartition (alias bicolouring) of the complement $\overline{\Gamma}_\epsilon$ of Γ_ϵ one can simply use the breadth–first search; see [13] for details. We can construct directly $\overline{\Gamma}_\epsilon$: two elements $x_i, x_j \in X$ are adjacent in $\overline{\Gamma}_\epsilon$ if and only if $x_i \prec x_j$ and $y_i - y_j > 2\cdot\epsilon$. Therefore, to solve the initial regression problem by Lemma 1 we must find the smallest feasible value in D. We use the binary search in the ordered matrix D. Namely, we start from a median ϵ of this list. We construct the graph Γ_ϵ and check if this graph has a covering with p cliques. If the answer is "yes" we continue the search in the first half of the list (removing the second sublist from further considerations). Otherwise, if the answer is "not", then we remove the first half and continue the search in the sublist of D containing the elements larger that ϵ. In the current list we take a median element as a current ϵ and check if it is feasible. We continue the procedure, until we arrive at a list containing only one element ϵ^*. This is the optimal error for the formulated regression problem, while any covering $X_1, \ldots, X_p$ of Γ_{ϵ^*} with at most p cliques and the isotonic functions $\hat{f}_1, \ldots, \hat{f}_p$ defined on $X_1, \ldots, X_p$ according to Proposition 2, represent the optimal solution. To find it, we must perform $O(logn)$ feasibility tests (namely, $logn^2$ such tests) in the sorted matrix D (to order the elements of D we need $O(n^2 logn)$ operations). The graph Γ_ϵ can be constructed in $O(n^2)$ time. If $p = 2$ within the same time bounds one can decide if $\overline{\Gamma}_\epsilon$ is bipartite. Therefore, the whole complexity of the algorithm for $p = 2$ is $O(n^2 logn)$.

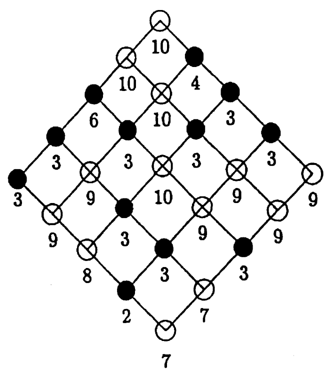

FIGURE 2.

Proposition 3 *For $p = 2$ the optimal bipartition of X with respect to the l_∞-criterion function can be constructed in $O(n^2 logn)$ number of operations.*

In Figure 2 we present an optimal bipartition of the poset from Figure

1. Note that the optimal error is $\epsilon^* = 1$.

3.3 Isotonic l_∞–regression problem with a given number of values

Some papers [4, 15, 14] consider the following approximation problem: given an integer p and $(x_1, y_1), \ldots, (x_n, y_n)$ in $\mathbb{R}^2$ with $x_1 < \ldots < x_n$ find a piecewise–linear function f with at most p links such that $\max_{i=1,\ldots,n}|f(x_i) - y_i|$ is minimized. Efficient algorithms for solving this problem are presented in [15, 14], for motivation see [4, 15]. If instead of piecewise–linear functions we consider stepwise functions with a fixed number of steps we obtain a particular case of the rectilinear center trajectory problem investigated in [6] (for the latter problem [6] presents an algorithm with the complexity $O(n^p)$). In this section we consider the following regression problem: consider the numbers $x_1 < \ldots, x_n$, and, assume that with each x_i is associated a set of (distinct) numbers $y_{i1}, \ldots, y_{ir_i}$. Given an integer p we wish to find an isotonic stepwise function f that minimizes

$$D_\infty(f) = \max_{x_i \in X} \max_{l=1,\ldots,r_i} |y_{il} - f(x_i)|$$

subject to the restriction that f takes at most p distinct values. Let $a_i = \min_{l=1,\ldots,r_i} y_{il}$ and $b_i = \max_{l=1,\ldots,r_i} y_{il}$. The key observation is that, as in the previous section, the optimal error ϵ^* of D_∞ is an element of the matrix $D = (d_{ij})$, where $d_{ij} = |b_i - a_j|$. Therefore we can use a binary search in the ordered list of the elements of D. With a current $\epsilon \in D$ we must answer the following question: "There is an isotonic function f with at most p steps such that $D_\infty(f) \leq \epsilon$?" To perform this test we proceed as follows. For a given x_i denote by S_i the intersection of the segments $[a_i - \epsilon, a_i + \epsilon]$ and $[b_i - \epsilon, b_i + \epsilon]$. We sweep the list $x_1, \ldots, x_n$ from left to right. We need three parameters S, q and $\mathcal{S}$ whose meaning shall became clear immediately. Initially, let $S := S_1$, $q := 0$ and $\mathcal{S} = (-\infty, x_1]$. At point x_i we do the following. Find $S \cap S_i$. If this intersection is nonempty, then set $S := S \cap S_i$, $\mathcal{S} = \mathcal{S} \cup (x_{i-1}, x_i]$ and go to the point x_{i+1}. Otherwise, if $S \cap S_i = \emptyset$, then for all $x \in \mathcal{S}$ define $f(x) := s$, where s is an arbitrary value from the segment S. If $x_i < s$, then stop: the test has a negative answer. Otherwise, set $S := S_i$, $q := q + 1$, $\mathcal{S} = (x_{i-1}, x_i]$ and consider the next point x_{i+1} (of course, if $i = n$ we simply put $f(x_i) = y_i$ and finish the procedure). After n steps we return answer "yes" if $q \leq p$ and the answer "no", otherwise. The complexity of this procedure is $O(n)$. The proof of correctness is straighforward. To find an optimal isotonic function f with at most p values we must perform $O(logn)$ feasibility tests in the ordered matrix D. If we simply sort the matrix D, the total complexity of the algorithm will be $O(n^2 logn)$ (actually, this is the time to sort D, because the complexity of testing is only $O(nlogn)$. We can improve the whole complexity of our algorithm. Instead of constructing and sorting the matrix D, we can use the selection algorithm of [23]. It presents an $O(nlog^2n)$ time algorithm for computing the kth largest element in the set of all simple paths in a tree. One can view the sorted list of numbers $\{a_i, b_i : i = 1, \ldots, n\}$ as a path; therefore, we can apply the algorithm from [23] $O(logn)$ times, leading us to an algorithm with the total complexity $O(nlog^3n)$.

Proposition 4 *Given a total order $x_1 < \ldots < x_n$ and an integer $p > 0$ an isotonic function f minimizing the l_∞–criterion function subject to the restriction that f has at most p distinct values can be constructed in $O(nlogn^3)$ number of operations.*

Most likely, using the parametric search as in [2, 14] one can solve this problem more efficiently. We leave open the question whether a similar problem for all partial orders is NP–complete. Finally note that within the same time bounds we can solve the problem of approximating with a stepwise function with at most p distinct values. Again, the optimal l_∞–error is an element of the matrix D. We can apply the same test, but in case $S \cap S_i = \emptyset$ it is not necessary to check whether $x_i < s$.

References

[1] Ayer, M., Brunk, H.D., Ewing, G.M., Reid, W.T. and Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Ann. Math. Statist.* **26**, 641–647.

[2] Agarwal, P. and Sharir, M., Planar (1994). Geometric location problems. *Algorithmica* **11,** 185–195.

[3] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Statistical Inference Under Order Restrictions.* New York: Wiley.

[4] Bellman, R.E. and Roth R.S. (1969). Curve fitting by segmented straight lines. *J. Am. Statist. Assoc.* **64**, 1079–1084.

[5] Brunk, H.D., Ewing, G.M. and Utz, W.R. (1957). Minimizing integrals in certain classes of monotone functions. *Pacif. J. Math.* **7**, 833–847.

[6] Diaz-Banez, J.M. and Mesa, J.A. (1995). An algorithm for a rectilinear center trajectory. Preprint Universidad de Sevilla.

[7] Dykstra, R.L. (1983). An algorithm for restricted least squares regressions. *J. Am. Statist. Assoc.* **78**, 837–842.

[8] Dykstra, R.L. and Robertson, T. (1982). An algorithm for isotonic regression for two or more independent variables. *Ann. Statist.* **10**, 708–716.

[9] Eeden, C. van (1956). Maximum likelihood estimation of ordered probabilities. *Proc. Kon. Nederl. Akad. Wetensch.* A **59**, 444-455.

[10] Eeden, C. van (1958). Testing and estimating ordered parameters of probability distributions. Doctoral Dissertation, University of Amsterdam.

[11] Eeden, C. van. Estimation in restricted parameter spaces. Some history and some recent developments. Technical report 163, Dept. Stat. University of British Columbia, Vancouver, Canada (to appear in CWI Quarterly).

[12] Farebrother, R.W. (1987). The historical development of the L_1 and L_∞ estimation procedures. In *Statistical Data Analysis Based on the L_1-Norm and Related Methods,* Ed. Y. Dodge, pp. 37–63.

[13] Golumbic M.C. (1980). *Algorithmic Graph Theory and Perfect Graphs.* New York: Academic Press.

[14] Goodrich, M.T. (1995). Efficient piecewise–linear function approximation using the uniform metric. *Discrete Comput. Geom.* **14**, 445–462.

[15] Hakimi, S.L. and Schmeichel, E.F. (1991). Fitting polygonal functions to a set of points in the plane. *CVGIP: Graph. Mod. Image Proc.* **53**, 132–136.

[16] Hayes, A.C. and Larman, D.G. (1985). Least squares procedures for problems involving more that one hyperplane. In *Discrete Geometry and Convexity,* Eds. J.E. Goodman, E. Lutwak, J. Malkevitch and R. Pollack, pp. 335–363.

[17] Kariv, O. and Hakimi, S.L. (1979). An algorithmic approach to network location problems, I,II. *SIAM J. Appl. Math.* **37**, 513–537 and 538–563.

[18] Kruskal, J.B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* **29**, 115–129.

[19] Lawson, C. and Hanson, R.J. (1974). *Solving Least Squares Problems.* Prentice–Hall.

[20] Lenstra, A.K., Lenstra, J.K., Rinnooy Kan A.H.G. and Wansbeek, T.J. (1981). Two lines least squares. *Math. Centrum.* Preprint.

[21] Leurgans, S. (1981). The Cauchy mean value property and linear functions of order statistics. *Ann. Statist.* **9**, 905–908.

[22] Megiddo, N. and Supowit, K.(1984). On the complexity of some common geometric location problems. *SIAM J. Comput.* **13**, 182–196.

[23] Megiddo, N., Tamir, A., Zemel, E. and Chandrasekaran, R. (1981). An $O(nlog^2n)$ algorithm for the kth longest path in a tree with application to location problems. *SIAM J. Comput.* **10**, 328–337.

[24] Menéndez, J.A. and Salvador, B. (1987). An algorithm for isotonic median regression. *Comput. Statist. Data Anal.* **5**, 399–406.

[25] Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order Restricted Statistical Inference.* New York: Wiley

[26] Robertson, T. and Wright, F.T. (1973). Multiple isotonic median regression. *Ann. Statist.* **1,** 422–432.

[27] Robertson, T. and Wright, F.T. (1974). A norm reducing property for isotonized Cauchy mean value functions. *Ann. Statist.* **6**, 1302–1307.

[28] Robertson, T. and Wright, F.T. (1980). Algorithms in order restricted statistical inference and the Cauchy mean value property. *Ann. Statist.* **8**, 645–651.

[29] Sager, T.W. and Thisted, R.A. (1982). Isotonic modal regression. *Ann. Statist.* **10**, 692–707.

[30] Tamir, A. (1988). Improved complexity bounds for center location problems on networks by using dynamic data structures. *SIAM J. Discr. Math.* **1**, 377–396.

[31] Thompson, W.A. Jr. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33**, 273–289.

[32] Ubhaya, V.A. (1974). Isotone optimization, I,II. *J. Approxim. Th.* **12**, 146–159 and 315–331.

[33] Ubhaya, V.A. (1974). Approximation by isotone functions–a min–max form of best approximation. *Notices Amer. Math. Soc.* **21**, A–307.