A Decomposition of the Isotonic Regression

Ad Feelders

Technical Report UU-CS-2010-009
April 2010

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

ISSN: 0924-3275

Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

A Decomposition of the Isotonic Regression

Ad Feelders

Utrecht University, Department of Information and Computing Sciences,
P.O. Box 80089, 3508TB Utrecht, The Netherlands,
ad@cs.uu.nl

Abstract. The isotonic regression is a useful technique in many statistical estimation problems with order constraints, as well as in the related problem of learning monotone models from data. We show that the computation of the isotonic regression decomposes over the connected components of the violation graph. This allows a straightforward divide-and-conquer strategy for its solution.

1 Introduction

The isotonic regression is a useful technique in many statistical estimation problems with order constraints [6]. It is also applied in machine learning and data mining for learning monotone models from data [1, 5]. We show that the computation of the isotonic regression decomposes over the connected components of the violation graph. Hence, a substantial simplification of its computation is possible.

2 The isotonic regression

In this section we define the isotonic regression, for a detailed description we refer to [6].

Let $X = \{x_1, x_2, \ldots, x_n\}$ be a set of constants and let $\preceq$ be a partial order on X. Any real-valued function f on X is *isotonic* with respect to $\preceq$ if, for any $x, x' \in X$, $x \preceq x'$ implies $f(x) \leq f(x')$. We assume that each element x_i of X is associated with a real number $g(x_i)$, and a positive weight $w(x_i)$. An isotonic function g^* on X now is an *isotonic regression* of g with respect to the weight function w and the partial order $\preceq$ if and only if it minimizes the sum

$$\sum_{i=1}^{n} w(x_i)\ [f(x_i) - g(x_i)]^2 \tag{1}$$

in the class of isotonic functions f on X. Brunk [3], proved that g^* exists and is unique. Hence it makes sense to talk about *the* isotonic regression of g on $(X, \preceq)$ with respect to w.

Next, we introduce some useful concepts to describe the computation of the isotonic regression. The downset $\downarrow_x$ of x is the set $\{x' \in X : x' \preceq x\}$. Likewise, the upset $\uparrow_x$ of x is the set $\{x' \in X : x \preceq x'\}$. Also, for $S \subseteq X$, we define $\downarrow_S = \bigcup_{x \in S} \downarrow_x$. A subset L of X is a *lower set* of X if it contains the downset of all its elements. Likewise, a subset U of X is an *upper set* of X if it contains the upset of all its elements. The *weighted average* of g, with weights w, for a nonempty subset A of X is defined as

$$\mathrm{Av}(A) = \frac{\sum_{x \in A} w(x) g(x)}{\sum_{x \in A} w(x)} \tag{2}$$

The minimum lower sets algorithm [2] given in Algorithm 1 correctly computes the isotonic regression for arbitrary partial orders. This algorithm is not really practical, due to the huge size of $\mathcal{L}$ even for relatively small problems, but we will use it to prove the correctness of our decomposition. The best exact algorithm known computes the isotonic regression by solving at most n maximum flow problems on a transportation network with $n+2$ nodes, and hence has time complexity $O(n^4)$, see [7].

Algorithm 1 MinimumLowerSets(X, $\preceq$, $g(x)$, $w(x)$)

1: $\mathcal{L} \leftarrow$ Collection of all lower sets of X with respect to $\preceq$
2: **repeat**
3:    $B \leftarrow \bigcup\{A \in \mathcal{L} : \mathrm{Av}(A) = \min_{L \in \mathcal{L}} \mathrm{Av}(L)\}$
4:    **for all** $x \in B$ **do**
5:       $g^*(x) \leftarrow \mathrm{Av}(B)$
6:    **end for**
7:    **for all** $L \in \mathcal{L}$ **do**
8:       $L \leftarrow L \setminus B$
9:    **end for**
10:    $X \leftarrow X \setminus B$
11: **until** $X = \varnothing$
12: **return** g^*

The algorithm can be described informally as follows. We start with $(X, \preceq)$ and find a minimum lower set B, that is, a lower set with minimum weighted average. For each element x of B we put $g^*(x) = \mathrm{Av}(B)$. Then the points in B are removed from X, and the order is restricted to the induced suborder on $X \setminus B$. Then we find a minimum lower set for this reduced problem, and so on, until X is exhausted. Hence, the isotonic regression partitions X into a number of blocks, where each block is a lower set of an upper set of X. On each of these blocks, the isotonic regression is constant and equal to the block average.

In Algorithm 1 we take the union of all minimum lower sets in line (3), in order to get the maximal size minimum lower set, but in fact it doesn't matter which minimum lower set we pick. In particular, we can pick a minimum lower set of smallest size. This variant of the minimum lower sets algorithm is given in Algorithm 2: it computes what Dykstra et al. [4] call the maximal partition of X with respect to the isotonic regression. In line (6) of the algorithm we choose an arbitrary minimum lower set of smallest size, in case there is more than one. It is this version of the minimum lower sets algorithm that we actually use in our proof.

Finally, it is convenient to work with the following graph representation. Let the order graph $OG = (V, E)$ be defined as the directed graph with nodes $V = \{1, 2, \ldots, n\}$, where node i corresponds to x_i, and $(i, j) \in E$ if and only if $x_i \preceq x_j$. The violation graph $VG = (V, E')$ has the same set of nodes as OG, with $(i, j) \in E'$ if and only if

$$x_i \preceq x_j \text{ and } g(x_i) > g(x_j). \tag{3}$$

Algorithm 2 MaximalPartition(X, $\preceq$, $g(x)$, $w(x)$)

$\mathcal{L} \leftarrow$ Collection of all lower sets of X with respect to $\preceq$
MaxPart $\leftarrow \varnothing$
repeat
  $B_1 \leftarrow \{A \in \mathcal{L} : \text{Av}(A) = \min_{L \in \mathcal{L}} \text{Av}(L)\}$
  $B_2 \leftarrow \{A \in B_1 : |A| = \min_{L \in B_1} |L|\}$
  $B \leftarrow \text{rnd}(B_2)$
  MaxPart $\leftarrow$ MaxPart $\cup \{B\}$
  for all $x \in B$ **do**
    $g^*(x) \leftarrow \text{Av}(B)$
  end for
  for all $L \in \mathcal{L}$ **do**
    $L \leftarrow L \setminus B$
  end for
  $X \leftarrow X \setminus B$
until $X = \varnothing$
return g^*, MaxPart

Obviously, $E' \subseteq E$, so all lower sets of the OG are also lower sets of the VG, but not the other way around. We denote the downset of a point x in the violation graph as $\downarrow_x^{\text{VG}}$. Note that $\downarrow_x^{\text{VG}} \subseteq \downarrow_x^{\text{OG}}$.

3 Decomposition

Lemma 1. *Let M denote an arbitrary set of incomparable elements of the violation graph. Furthermore, let $A = \downarrow_M^{\text{VG}}$, and let $S = \{x \in M : g(x) \leq Av(A)\}$. Then*

$$Av(\downarrow_S^{\text{OG}}) \leq Av(A)$$

Proof: Note that S can not be empty, otherwise all elements of A would have above average g-values. We have

$$\text{Av}(\downarrow_S^{\text{VG}}) \leq \text{Av}(A),$$

because for all $x \in (\downarrow_M^{\text{VG}} \setminus \downarrow_S^{\text{VG}})$: $g(x) > \text{Av}(A)$. Furthermore,

$$\text{Av}(\downarrow_S^{\text{OG}}) \leq \text{Av}(A),$$

because for all $x \in (\downarrow_S^{\text{OG}} \setminus \downarrow_S^{\text{VG}})$: $g(x) \leq \text{Av}(A)$; otherwise x would be in conflict with some element of S, and hence contained in $\downarrow_S^{\text{VG}}$. □

Next, we state our main result.

Theorem 1. *Let $C_1, C_2, \ldots, C_k$ be the connected components of the violation graph on X[1]. Then:*

$$g_X^* = (g|_{C_1})^* \cup (g|_{C_2})^* \cup \ldots \cup (g|_{C_k})^*,$$

that is, the isotonic regression decomposes over the connected components of the violation graph.

[1] In a slight abuse of notation, C refers to a connected component, as well as the corresponding nodes of the violation graph and the corresponding elements of X.

Proof: The proof uses Algorithm 2. In line (6) a minimum lower set of smallest possible size is chosen. We show that such a lower set cannot contain elements from different connected components of the violation graph.

Let L be a minimum lower set of smallest possible size. Suppose the violation graph on L decomposes into connected components $C_1, \ldots, C_m$. From the properties of the weighted average, it follows that either:

1. There is a component C_i, $1 \leq i \leq m$, with lower average than L.
2. All components have the same average as L.

Starting with the first case, let M denote the maximal elements of the violation graph on C_i. Since L is a lower set, we have $\downarrow^{\text{OG}}_M \subseteq L$, and $C_i = \downarrow^{\text{VG}}_M$. Then, according to lemma 1:

$$\text{Av}(\downarrow^{\text{OG}}_S) \leq \text{Av}(C_i),$$

where $S = \{x \in M : g(x) \leq \text{Av}(C_i)\}$. Hence, there is a lower set $L' = \downarrow^{\text{OG}}_S$ with lower average than L. This contradicts the assumption that L is a minimum lower set.

In the second case we apply lemma 1 to all connected components C_j, $1 \leq j \leq m$, and their maximal elements M_j. At least one of the resulting lower sets must have smaller size than L. This contradicts the assumption that L is a minimum lower set with smallest possible size.

Hence, the violation graph on L must be connected, and therefore cannot contain elements of different connected components of the violation graph on X.□

Since the connected components of a graph can be computed in linear time, this decomposition may give a substantial reduction of the computation time for the isotonic regression.

We call a point $x \in X$ monotone if it satisfies the following two conditions:

1. for all $x' \in \downarrow^{\text{OG}}_x$: $g(x') \leq g(x)$, and
2. for all $x' \in \uparrow^{\text{OG}}_x$: $g(x') \geq g(x)$,

that is, x does not violate monotonicity with any other point.

Corrolary 1 *For every monotone point x, we have $g^*(x) = g(x)$.*

Proof: *A monotone point is obviously a connected component in the violation graph.*

4 Examples

Consider the order graph on $X = \{x_1, x_2, x_3, x_4\}$ given in the left part of figure 1.

Table 1 shows the computation of the isotonic regression with the minimum lower sets algorithm as given in Algorithm 1. Table 2 shows its computation with Algorithm 2. The minimum lower set chosen in each iteration is shown in bold. From table 2, we learn that the maximal partition with respect to the isotonic regression is $\{\{x_1, x_3\}, \{x_2\}, \{x_4\}\}$.

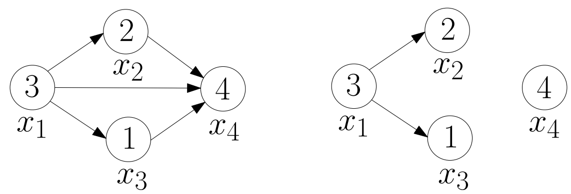

Fig. 1. Order graph on $X = \{x_1, x_2, x_3, x_4\}$ with values of g shown inside the nodes (left). The violation graph is given at the right. For ease of computation, weights are all set to one.

Table 1. Computation of the isotonic regression with Algorithm 1.

Iteration 1		Iteration 2	
Lower Set	Av	Lower Set	Av
x_1	3	–	–
x_1x_2	2.5	–	–
x_1x_3	2	–	–
$\mathbf{x_1x_2x_3}$	2	–	–
$x_1x_2x_3x_4$	2.5	$\mathbf{x_4}$	4
$g^*(x_1) = g^*(x_2) = g^*(x_3) = 2$		$g^*(x_4) = 4$	

Table 2. Computation of the isotonic regression with Algorithm 2.

Iteration 1		Iteration 2		Iteration 3	
Lower Set	Av	Lower Set	Av	Lower Set	Av
x_1	3	–	–	–	–
x_1x_2	2.5	$\mathbf{x_2}$	2	–	–
$\mathbf{x_1x_3}$	2	–	–	–	–
$x_1x_2x_3$	2	$\mathbf{x_2}$	2	–	–
$x_1x_2x_3x_4$	2.5	x_2x_4	3	$\mathbf{x_4}$	4
$g^*(x_1) = g^*(x_3) = 2$		$g^*(x_2) = 2$		$g^*(x_4) = 4$	

From the violation graph in the right part of figure 1, we learn that the isotonic regression can be computed for the connected components $C_1 = \{x_1, x_2, x_3\}$ and $C_2 = \{x_4\}$ separately. Since x_4 is a monotone point (and hence a connected component on its own), we can conclude immediately that $g^*(x_4) = g(x_4) = 4$.

As a second example, consider the order graph on $X = \{x_1, \ldots, x_7\}$ depicted in figure 2 (to avoid clutter, the transitive reduction is shown).

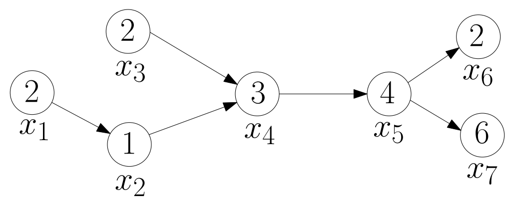

Fig. 2. Transitive reduction of the order graph on $X = \{x_1, \ldots, x_7\}$ with values of g shown inside the nodes.

The corresponding violation graph is shown in figure 3. The partitioning of X corresponding to the connected components of the violation graph is $\{\{x_1, x_2\}, \{x_3\}, \{x_4, x_5, x_6\}, \{x_7\}\}$. Hence, the computation of the isotonic regression can be decomposed accordingly. Figure 4 shows the corresponding induced subgraphs of the order graph given in figure 2.

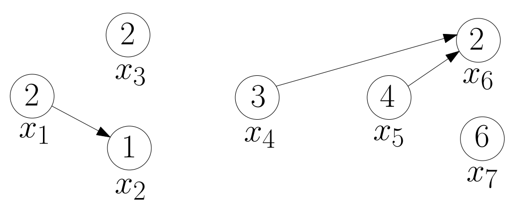

Fig. 3. Violation graph corresponding to the order graph shown in figure 2.

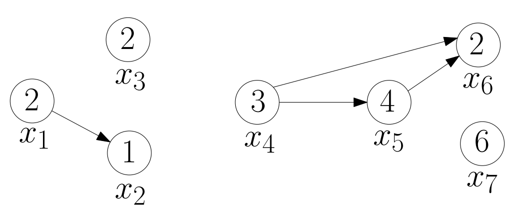

Fig. 4. Induced subgraphs of the order graph corresponding to the decomposition into connected components of the violation graph. The isotonic regression can be computed on each component separately.

5 Conclusion

We have shown how the isotonic regression decomposes over the connected components of the graph that represents the violations of the monotonicity constraint. This allows for a straightforward divide-and-conquer strategy to compute the isotonic regression. Since the best exact algorithm known has time complexity $O(n^4)$, whereas the connected components of the violation graph can be computed in linear time, this may result in a substantial reduction of computation time.

References

1. N. Barile and A. Feelders. Nonparametric monotone classification with MOCA. In F. Giannotti, editor, *Proceedings of the Eighth IEEE International Conference on Data Mining (ICDM 2008)*, pages 731–736. IEEE Computer Society, 2008.
2. H.D. Brunk. Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics*, 26:607–616, 1955.
3. H.D. Brunk. Conditional expectation given a σ-lattice and applications. *Annals of Mathematical Statistics*, 36:1339–1350, 1965.
4. R. Dykstra, J. Hewett, and T. Robertson. Nonparametric, isotonic discriminant procedures. *Biometrika*, 86(2):429–438, 1999.
5. W. Kotlowski and R. Slowinski. Rule learning with monotonicity constraints. In L. Bottou and M. Littman, editors, *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 537–544, 2009.
6. T. Robertson, F. Wright, and R.L. Dykstra. *Order Restricted Statistical Inference*. Wiley, 1988.
7. J. Spouge, H. Wan, and W.J. Wilbur. Least squares isotonic regression in two dimensions. *Journal Of Optimization Theory And Applications*, 117(3):585–605, 2003.