An $O(n^2)$ Algorithm for Isotonic Regression

Oleg Burdakov, Oleg Sysoev, Anders Grimvall, and Mohamed Hussian

Department of Mathematics, Linköping University, SE-58183 Linköping, Sweden
({olbur,olsys,angri,mohus}@mai.liu.se)

Summary. We consider the problem of minimizing the distance from a given n-dimensional vector to a set defined by constraints of the form $x_i \leq x_j$. Such constraints induce a partial order of the components x_i, which can be illustrated by an acyclic directed graph. This problem is also known as the isotonic regression (IR) problem. IR has important applications in statistics, operations research and signal processing, with most of them characterized by a very large value of n. For such large-scale problems, it is of great practical importance to develop algorithms whose complexity does not rise with n too rapidly. The existing optimization-based algorithms and statistical IR algorithms have either too high computational complexity or too low accuracy of the approximation to the optimal solution they generate. We introduce a new IR algorithm, which can be viewed as a generalization of the Pool-Adjacent-Violator (PAV) algorithm from completely to partially ordered data. Our algorithm combines both low computational complexity $O(n^2)$ and high accuracy. This allows us to obtain sufficiently accurate solutions to IR problems with thousands of observations.

Key words: quadratic programming, large scale optimization, least distance problem, isotonic regression, pool-adjacent-violators algorithm.

1 Introduction

We consider the isotonic regression problem (IR) in the following least distance setting. Given a vector $a \in R^n$, a strictly positive vector of weights $w \in R^n$ and a directed acyclic graph $G(N, E)$ with the set of nodes $N = \{1, 2, \ldots, n\}$, find $x^* \in R^n$ that solves the problem:

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} w_i (x_i - a_i)^2 \\ \text{s.t. } & x_i \leq x_j \quad \forall (i,j) \in E. \end{aligned} \tag{1}$$

Since this is a strictly convex quadratic programming problem, its solution x^* is unique. The optimality conditions and an analysis of the typical block (or cluster) structure of x^* can be found in [BC90,Lee83,PX99]. The monotonicity constraints defined by the acyclic graph $G(N, E)$ imply a partial order of the components x_i, $i = 1, \dots, n$.

A special case of IR problem arises when there is a complete order of the components. This problem, referred to as IRC problem, is defined by a directed path $G(N, E)$ and is formulated as follows:

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} w_i (x_i - a_i)^2 \\ \text{s.t.} \quad & x_1 \leq x_2 \leq \cdots \leq x_n. \end{aligned} \tag{2}$$

IR problem has numerous important applications, for instance in statistics [BBBB72,DR82,Lee83], operations research [MM85,R86] and signal processing [AB98, RB93]. These applications are characterized by a very large value of n. For such large-scale problems, it is of great practical importance to develop algorithms whose complexity does not rise with n too rapidly. The existing optimization-based algorithms [DMT01] and statistical IR algorithms have either high computational complexity or low accuracy of the approximation to the optimal solution they generate. In this connection, let us quote a recent paper by Strand [Str03]:

> Unfortunately, in the case where $m > 1$ and at least some of the explanatory variables are continuous (i.e. typical multiple regression data), *there is no practical algorithm* to determine LSIR ... estimates

where m stands for the number of explanatory variables. The case $m > 1$ in the least squares isotonic regression (LSIR) corresponds to problem (1), while the case $m = 1$ corresponds to (2).

The most widely used method for solving IRC problem (2) is the so-called pool-adjacent-violators (PAV) algorithm [ABERS55, BBBB72, HPW73]. This algorithm is of computational complexity $O(n)$ (see [BC90, PX99]). The PAV algorithm has been extended by Pardalos and Xue [PX99] to the special case of IR problem (1), in which the partial order of the components is presented by a directed tree. Several other special cases, in which the PAV algorithm is applied repeatedly to different subsets of the components, are considered in [BDPR84,DR82,SS97,Str03]. In [BC90], it was shown that some algorithms for problems (1) and (2) may be viewed as active set quadratic programming methods. The minimum lower set algorithm [Br55] is known to be the first algorithm proposed for the general IR problem. It was shown in [BC90] that this algorithm, being applied to IRC problem, is of complexity $O(n^2)$. Perhaps, the most widely used approaches for solving applied IR problems are based on simple averaging techniques [M88, MS94, Str03]. They can be easily implemented and enjoy a relatively low computational burden, but the quality of their approximations to x^* are very case-dependent and can be far from