A Dual Active-Set Algorithm for Regularized Slope-Constrained Monotonic Regression

O. Burdakov[1,*], O. Sysoev[2]

In many problems, it is necessary to take into account monotonic relations. Monotonic (isotonic) Regression (MR) is often involved in solving such problems. The MR solutions are of a step-shaped form with a typical sharp change of values between adjacent steps. This, in some applications, is regarded as a disadvantage. We recently introduced a Smoothed MR (SMR) problem which is obtained from the MR by adding a regularization penalty term. The SMR is aimed at smoothing the aforementioned sharp change. Moreover, its solution has a far less pronounced step-structure, if at all available. The purpose of this paper is to further improve the SMR solution by getting rid of such a structure. This is achieved by introducing a lowed bound on the slope in the SMR. We call it Smoothed Slope-Constrained MR (SSCMR) problem. It is shown here how to reduce it to the SMR which is a convex quadratic optimization problem. The Smoothed Pool Adjacent Violators (SPAV) algorithm developed in our recent publications for solving the SMR problem is adapted here to solving the SSCMR problem. This algorithm belongs to the class of dual active-set algorithms. Although the complexity of the SPAV algorithm is $O(n^2)$, its running time is growing in our computational experiments almost linearly with n. We present numerical results which illustrate the predictive performance quality of our approach. They also show that the SSCMR solution is free of the undesirable features of the MR and SMR solutions.

1. Introduction

A wide range of important applied problems is related to solving *monotonic regression* (MR) problem (see, e.g., [2, 8]). The applied MR problems are often associated with the monotonic data fitting, where it is assumed that there is an unknown monotonic response function $\chi(t)$ of an explanatory variable t. In this paper, we restrict our attention to the univariate case and suppose that $\chi(t)$ is monotonically increasing, i.e.,

$$\chi(t') < \chi(t''), \ \ \forall t' < t''.$$

In practice, there are available observed values of the explanatory variable $t_1 < t_2 < \cdots < t_n$ and the corresponding sequence of observed response values

---

* Corresponding Author.
[1] Department of Mathematics, Linköping University, Linköping, Sweden, Email: oleg.burdakov@liu.se.
[2] Department of Computer and Information Science, Linköping University, Linköping, Sweden, Email: oleg.sysoev@liu.se.

$$a_i = \chi(t_i) + \epsilon_i, \quad i = 1, 2, \dots, n, \tag{1}$$

where ϵ_i is an observation error. These errors result in loss, for some indices i, in the expected monotonicity $a_i \leq a_{i+1}$. The MR problem is aimed at restoring the lost monotonicity by finding a least-change correction to the observed values. This can be formally stated as the quadratic optimization problem:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} w_i (x_i - a_i)^2 \quad \text{s.t.} \quad x_1 \leq x_2 \leq \cdots \leq x_n, \tag{2}$$

where $w \in \mathbb{R}^n_{++}$ is a vector of weights.

The most efficient algorithm for solving (2) is a, so-called, *pool adjacent violators* (PAV) algorithm [1, 5, 7]. It belongs to the class of dual active set algorithms [3, 6]. The efficiency and popularity of the PAV algorithm result from its linear computational complexity, $O(n)$. This feature is especially important in case of large data sets.

The solution to (2) resembles a step function, and on some intervals of t, the finite-difference approximation of derivatives may have too large values, like in Fig. 1(a) and Fig. 2(a). This poses a limitation of using the MR in some applications. To prevent from such too large values, a smoothing approach was proposed in [4, 9]. It is based on introducing the penalty term

$$\mu \sum_{i=1}^{n-1} \left(\frac{x_{i+1} - x_i}{t_{i+1} - t_i} \right)^2$$

in (2). Given a data set, the penalty parameter μ can be chosen, as suggested in [9], by making use of Bayesian modeling and a cross-validation technique. Following [4, 9], we denote

$$\mu_i = \frac{\mu}{(t_{i+1} - t_i)^2} \tag{3}$$

and formulate the resulting *smoothed monotonic regression* (SMR) problem as

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} w_i (x_i - a_i)^2 + \sum_{i=1}^{n-1} \mu_i (x_i - x_{i+1})^2 \quad \text{s.t.} \quad x_i \leq x_{i+1}, \ \forall i \in [1, n-1]. \tag{4}$$

Here and later, the segment of indices $\{i, i+1, \dots, j-1, j\}$ is denoted by $[i, j]$. The results of smoothing can be observed in Fig. 1(b) and Fig. 2(b).

In [4, 9], we developed an efficient dual active-set algorithm for solving the SMR problem. It was called *smoothed pool adjacent violators* (SPAV) algorithm. Although its complexity was proved to be $O(n^2)$, its running time was growing in our computational experiments almost linearly with n.

The solution to the SMR problem looks smoother than the one for the MR problem. However, it still inherits some step-shaped structure of the MR solution, although with definitely shorter segments of piece-wise constant values. In the next section, we present a problem whose solution is smooth

and it is free of the step-shaped structure. This improvement is obtained owing to introducing a lower bound for the slope.

2. Smoothed Slope-constrained MR Problem

Consider an extra requirement to the fitted response which assumes that its slope is bounded below by a scalar $\Delta \geq 0$. The value of Δ is supposed to be chosen for a given data set using, for instance, a cross-validation technique as described in Section 4. The restriction on the slope can be written as

$$\frac{x_{i+1} - x_i}{t_{i+1} - t_i} \geq \Delta. \tag{5}$$

In this case, the value of the smoothing penalty term should be zero when (5) is satisfied as equality. Taking this into account, we can formulate our *smoothed slope-constrained monotonic regression* (SSCMR) problem as the following strictly convex quadratic optimization problem:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} w_i(x_i - a_i)^2 + \sum_{i=1}^{n-1} \mu_i(x_{i+1} - x_i - \delta_i)^2 \quad \text{s.t.} \quad x_i + \delta_i \leq x_{i+1}, \ \ \forall i \in [1, n-1], \tag{6}$$

where μ_i is defined by (3), and

$$\delta_i = \Delta(t_{i+1} - t_i). \tag{7}$$

The solution to this problem, x^*, exists and is unique. Note that problems (2) and (4) are special cases of (6) for $\Delta = \mu = 0$ and $\Delta = 0$, respectively.

Denote

$$x_i' = x_i + \Delta(t_n - t_i) \quad \text{and} \quad a_i' = a_i + \Delta(t_n - t_i), \ \ \forall i \in [1, n]. \tag{8}$$

Then, as it can be easily verified, (6) is equivalent to the smoothed monotonic regression problem

$$\min_{x' \in \mathbb{R}^n} \sum_{i=1}^{n} w_i(x_i' - a_i')^2 + \sum_{i=1}^{n-1} \mu_i(x_i' - x_{i+1}')^2 \quad \text{s.t.} \quad x_i' \leq x_{i+1}', \ \ \forall i \in [1, n-1]. \tag{9}$$

In the next section, we adapt the SPAV algorithm to solving the SSCMR problem.

3. A Dual Active-set Algorithm and Its Properties

We shall refer to $x_i + \delta_i \leq x_{i+1}$ and $x_i' \leq x_{i+1}'$ in (6) and (9), respectively, as constraint $i \in [1, n-1]$. Obviously, each of them is active if, and only if, the other one is active. According to [4, 9], each iteration of the SPAV algorithm is related to choosing an *active set* $S \subseteq [1, n-1]$ and solving the corresponding subproblem

$$\min_{x' \in \mathbb{R}^n} \sum_{i=1}^{n} w_i(x'_i - a'_i)^2 + \sum_{i=1}^{n-1} \mu_i(x'_i - x'_{i+1})^2 \quad \text{s.t.} \quad x'_i = x'_{i+1}, \ \forall i \in S. \tag{10}$$

Let $x'(S)$ denote the unique solution to this problem. It is obtained by solving a system of linear equations which originates from the optimality conditions.

Given an active set S, there exist sets of consecutive indices of the form $[\ell, r] \subseteq [1, n]$ such that $[\ell, r-1] \subseteq S$, $\ell - 1 \notin S$ and $r \notin S$. These sets are referred to as *blocks*. Note that a block may be a singleton when $\ell = r$. The total number of blocks, denoted here by m, is equal to $n - |S|$. The block partitioning (segmentation) of $[1, n]$ induced by S can be represented as

$$[1, n] = [\ell_1, r_1], [\ell_2, r_2], \dots, [\ell_m, r_m],$$

where $\ell_1 = 1$, $r_m = n$ and $r_i + 1 = \ell_{i+1}$ for all $i \in [1, m-1]$.

Each block i is characterized by its common value,

$$y_i = x'_{\ell_i} = x'_{\ell_i + 1} = \cdots = x'_{r_i},$$

its common weight,

$$\bar{w}_i = w_{\ell_i} + w_{\ell_i + 1} + \cdots + w_{r_i},$$

and its weighted average observed value,

$$\bar{a}_i = \frac{1}{\bar{w}_i} \sum_{j=\ell_i}^{r_i} w_j a'_j.$$

Denoting $\bar{\mu}_i = \mu_{r_i}$, we can write the subproblem (10) in the notation introduced above as

$$\min_{y \in \mathbb{R}^m} c + \sum_{j=1}^{m} \bar{w}_j(y_j - \bar{a}_j)^2 + \sum_{j=1}^{m-1} \bar{\mu}_j(y_j - y_{j+1})^2, \tag{11}$$

where the scalar c does not depend on y. The optimality conditions for (11) are given by the system of linear equations

$$\begin{cases} \bar{w}_1(y_1 - \bar{a}_1) + \bar{\mu}_1(y_1 - y_2) = 0, \\ \dots \\ \bar{w}_j(y_j - \bar{a}_j) + \bar{\mu}_{j-1}(y_j - y_{j-1}) + \bar{\mu}_j(y_j - y_{j+1}) = 0, \\ \dots \\ \bar{w}_m(y_m - \bar{a}_m) + \bar{\mu}_{m-1}(y_m - y_{m-1}). \end{cases} \tag{12}$$

Its solution, denoted by $y(S)$, is unique because the objective function in (11) is strictly convex. Given $y(S)$, the corresponding values of the components of x are computed as follows:

$$x_j = y_i(S) - \Delta(t_n - t_j), \quad \forall j \in [\ell_i, r_i], i \in [1, m]. \tag{13}$$

The algorithm, as described in [4, 9], starts with any active set such that $S \subseteq S^*$, where S^* is the active set for the optimal solutions in (6) and (9). The simplest of the valid choices is $S = \emptyset$. In each iteration, it solves the tridiagonal system of linear equations (12), and then it extends the set S by additionally making active the constraints in (9) for which the strict monotonicity $y_i(S) < y_{i+1}(S)$ is violated. This, like in the PAV algorithm, assumes merging the corresponding adjacent blocks, which is reduced to updating the coefficients that define the linear system (12). The corresponding number of arithmetic operations is proportional to the number of new active constraints. In contrast to the conventional active set algorithms, SPAV may enlarge the active set with more than one element at once. It operates with the block common values y_i, whereas the values of x_j are computed only at its terminal stage. The outlined algorithm can be formally expressed as follows.

Algorithm 1.
input $a \in \mathbb{R}^n, w \in \mathbb{R}^n_{++}, \mu, \Delta \in \mathbb{R}^1_+, S \subseteq S^*$.
Compute $a', \bar{a}, \bar{w}$ and $\bar{\mu}$.
Find $y(S)$ that solves (12).
while $y(S)$ is not strictly monotone **do**
    Set $S \leftarrow S \cup \{r_i : y_i(S) \geq y_{i+1}(S)\}$.
    Update $\bar{a}, \bar{w}$ and $\bar{\mu}$.
    Find $y(S)$ that solves (12).
end while
return $x(S)$ computed by formula (13).

This algorithm inherits the main properties of the SPAV algorithm. Algorithm 1 can be viewed as a dual active-set algorithm, because it aims at attaining primal feasibility while maintaining dual feasibility and complementary slackness. This does not require manipulation of the dual variables. The number of iterations, in the worst case, is n, and the cost of each iteration is proportional to m. Thus, the complexity of Algorithm 1 is $O(n^2)$. Since the actual number of its iterations, as it was observed in [4, 9] by the numerical experiments for the SPAV algorithm, is usually a very small fraction of n, the computational time of running Algorithm 1 grows in practice in proportion to n.

4.   Numerical Experiments

The presented results of numerical experiments here are aimed at illustrating the predictive performance quality of our approach. We considered the following two monotonic response functions

$$\chi_1(t) = t^3 \quad \text{and} \quad \chi_2(t) = 1.2t + \sin(t).$$

For each of these functions, a data set of $n = 500$ observations were generated using formula (1). The observed values of explanatory variables t_i were uniformly distributed in the intervals $[0.5, 2]$ and $[0, 15]$ for χ_1 and χ_2, respectively. In the both cases, the observation error ϵ_i was normally distributed with zero mean, while the standard deviations were 3 and 1, respectively.

All components of the vector of weights w were ones. The parameters μ_i in the SMR and SSCMR problems were calculated by formula (3). The parameters δ_i in the SSCMR problem were calculated by formula (7).

In the SMR problem, the value of μ was produced, for each data instance, by the cross-validation-based technique especially designed in [9] for this problem. This gave us $\mu = 0.0142$ and $\mu = 0.0320$ for the 1st and 2nd data sets, respectively.

To choose the parameters μ and Δ in the SSCMR problem, we made use of the reduction of this problem to the SMR problem (9) formulated for the data set in which a_i' is defined by (8). This allowed us to apply, for any fixed value of Δ, the aforementioned cross-validation technique for finding the value of $\mu(\Delta)$ that provided the best, over μ, cross-validation score. In the case of the 1st data set, the search over Δ for the best cross-validation score provided by $\mu(\Delta)$ resulted in $\Delta = 2.7$ and $\mu = 0.0104$. For the 2nd data set, we similarly obtained $\Delta = 0.6$ and $\mu = 0.0049$. We compared the results obtained for Δ chosen as described above with the results obtained for Δ equal to the minimal slope of the underlying response functions, which were $\Delta = 0.125$ and $\Delta = 1.2$ for χ_1 and χ_2, respectively.
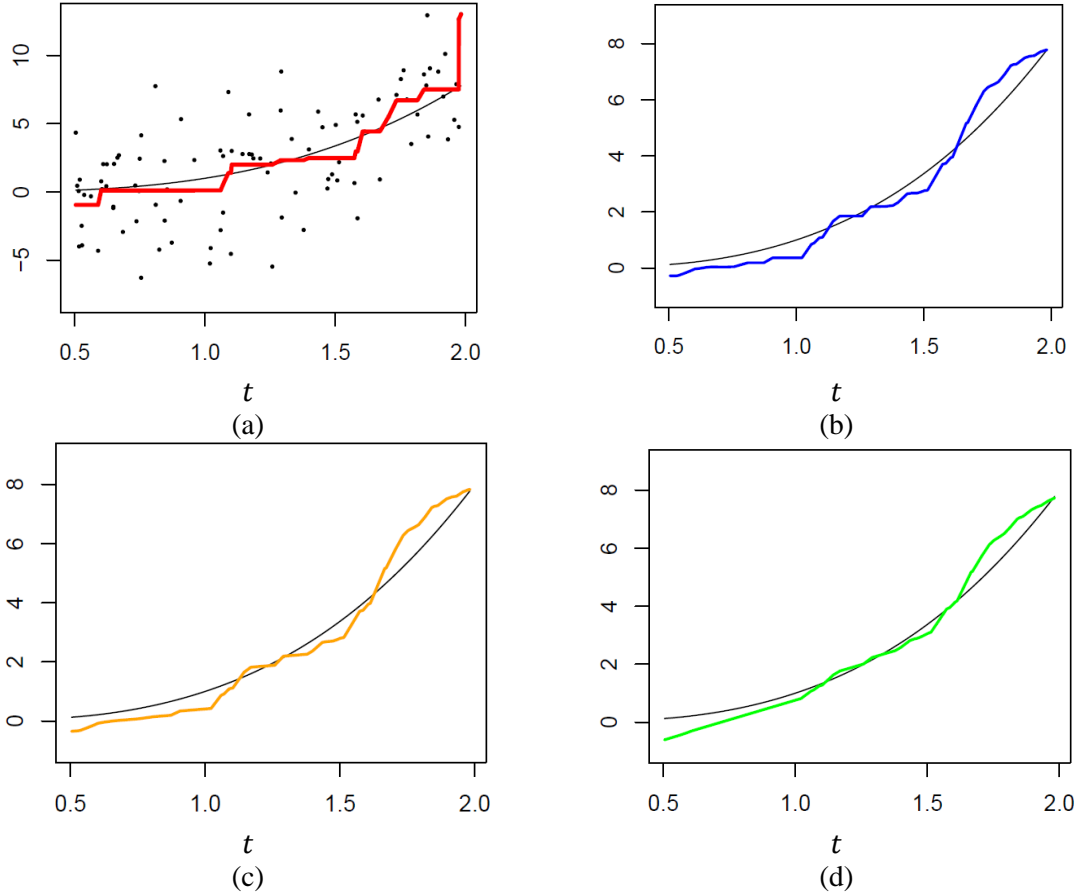

Figure 1. Results of regression on a data set (black dots) of the size $n = 500$ generated for $\chi_1(t)$ (black line). Solutions: (a) MR (red line), (b) SMR (blue line), (c) SSCMR for Δ equal to the minimal slope (yellow line), (d) SSCMR for the cross-validation-based choice of Δ (green line).

The two data sets were processed by the PAV algorithm, the SPAV algorithm and Algorithm 1. The monotonic interpolation of the fitted values was made by making use of the prediction model proposed in [9]. The results are presented in Fig. 1 and Fig. 2 for χ_1 and χ_2, respectively. These figures show that the SMR solution is smoother as compared to the MR solution, although the graph of the former still has some flat segments (shorter than in the case of the MR solution). The SSCMR solution provides a

further smoothing, and it is free of flat segments. For the two alternative choices of Δ, the SSCMR with the cross-validation-based choice produces curves which look, in general, closer to the curves of the corresponding monotonic response functions, with a possible exception of the boundaries.
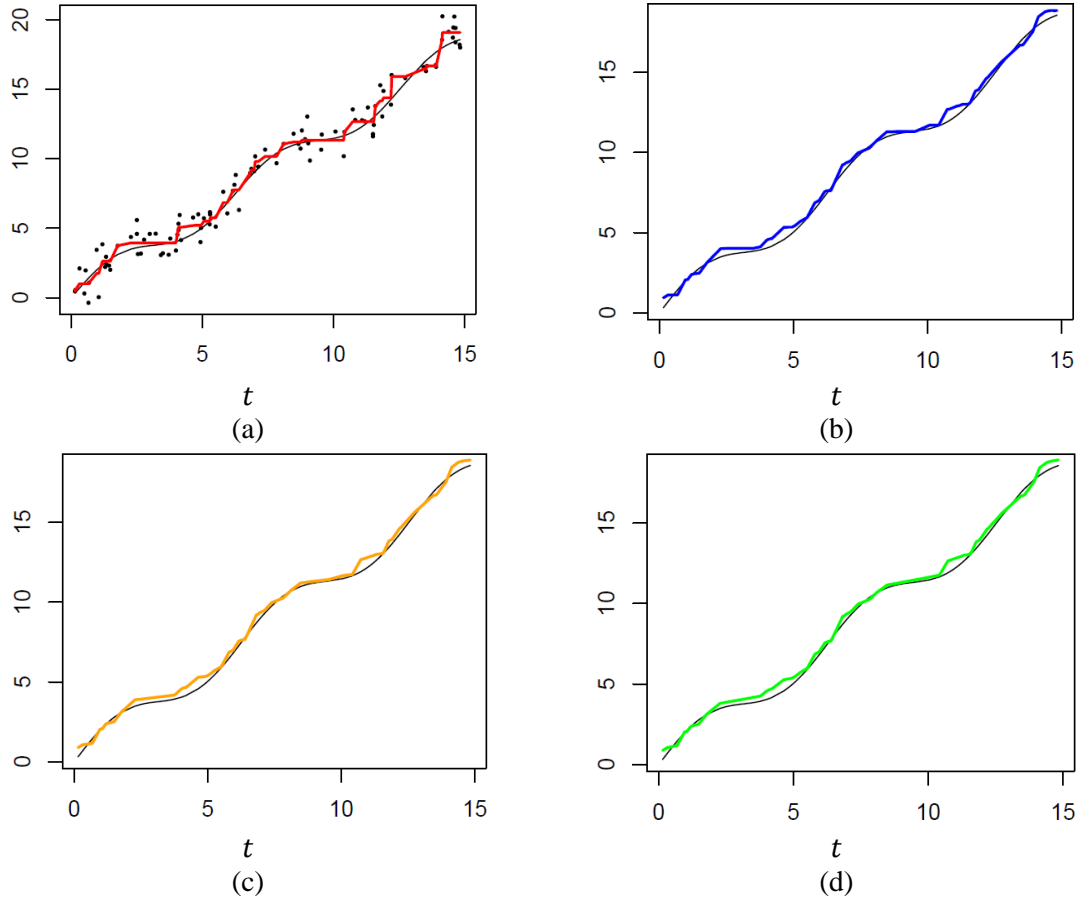

Figure 1. Results of regression on a data set (black dots) of the size $n = 500$ generated for $\chi_2(t)$ (black line). Solutions: (a) MR (red line), (b) SMR (blue line), (c) SSCMR for Δ equal to the minimal slope (yellow line), (d) SSCMR for the cross-validation-based choice of Δ (green line).

5.   Conclusions

Compared to the results of [4, 9], here we made a further progress in getting rid of some undesirable features of the MR solution. The improved fitted response was obtained in a numerically efficient way by adapting the SPAV algorithm to solving the SSCMR problem.

We focused on the SSCMR problem associated with a complete (linear) order of observations. It admits a natural extension to the case of partial order. We plan to develop algorithms for solving the extended SSCMR problem.

Acknowledgments

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. *The Annals of Mathematical Statistics*, 26, 641-647.

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

[3]   Best, M.J. and Chakravarti, N. (1990), Active set algorithms for isotonic regression: a unifying framework. *Mathematical Programming*, 47, 425-439.

[4]   Burdakov, O. and Sysoev, O. (2017), A dual active-set algorithm for regularized monotonic regression. *Journal of Optimization Theory and Applications*, DOI: 10.1007/s10957-017-1060-0.

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

[6]   de Leeuw, J., Hornik, K. and Mair, P. (2009), Isotone optimization in R: pool-adjacent-violators algorithm (PAVA) and active set methods. *Journal of Statistical Software*, 32(5), 1-24.

[7]   Miles, R.E. (1959), The complete amalgamation into blocks, by weighted means, of a finite set of real numbers. *Biometrika*, 46, 317-327.

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

[9]   Sysoev, O. and Burdakov, O. (2018), A smoothed monotonic regression via L2 regularization. To appear in *Knowledge and Information Systems*.