Vestn. Samar. Gos. Tekhn. Univ., Ser. Fiz.-Mat. Nauki
[J. Samara State Tech. Univ., Ser. Phys. Math. Sci.], 2019, vol. 23, no. 1, pp. 113–130
ISSN: 2310-7081 (online), 1991-8615 (print) https://doi.org/10.14498/vsgtu1673

Mathematical Modeling, Numerical Methods and Software Complexes

MSC: 90C20, 65K05, 62G08

A dual active set algorithm for optimal sparse convex regression

A. A. Gudkov, S. V. Mironov, S. P. Sidorov, S. V. Tyshkevich

N. G. Chernyshevsky Saratov State University (National Research University),
83, Astrakhanskaya st., Saratov, 410012, Russian Federation.

Abstract

The shape-constrained problems in statistics have attracted much attention in recent decades. One of them is the task of finding the best fitting monotone regression. The problem of constructing monotone regression (also called isotonic regression) is to find best fitted non-decreasing vector to a given vector. Convex regression is the extension of monotone regression to the case of 2-monotonicity (or convexity). Both isotone and convex regression have applications in many fields, including the non-parametric mathematical statistics and the empirical data smoothing. The paper proposes an iterative algorithm for constructing a sparse convex regression, i.e. for finding a convex vector $z \in \mathbb{R}^n$ with the lowest square error of approximation to a given vector $y \in \mathbb{R}^n$ (not necessarily convex). The problem can be rewritten in the form of a convex programming problem with linear constraints. Using the Karush–Kuhn–Tucker optimality conditions it is proved that optimal points should lie on a piecewise linear function. It is proved that the proposed dual active-set algorithm for convex regression has polynomial complexity and obtains the optimal solution (the Karush–Kuhn–Tucker conditions are fulfilled).

Keywords: dual active set algorithm, pool-adjacent-violators algorithm, isotonic regression, monotone regression, convex regression.

Research Article

Please cite this article in press as:
Gudkov A. A., Mironov S. V., Sidorov S. P., Tyshkevich S. V. A dual active set algorithm for optimal sparse convex regression, *Vestn. Samar. Gos. Tekhn. Univ., Ser. Fiz.-Mat. Nauki* [J. Samara State Tech. Univ., Ser. Phys. Math. Sci.], 2019, vol. 23, no. 1, pp. 113–130. doi: 10.14498/vsgtu1673.

Authors' Details:

Alexander A. Gudkov https://orcid.org/0000-0002-6472-4855
Laboratory Assistant; Risk Institute; e-mail: alex-good96@mail.ru

Sergei V. Mironov https://orcid.org/0000-0003-3699-5006
Cand. Phys. & Math. Sci.; Associate Professor; Dept. of Mathematic Cybernetics and Computer Science; e-mail: MironovSV@info.sgu.ru

Received: 22nd January, 2019 / Revised: 2nd March, 2019 / Accepted: 4th March, 2019 / First online: 5th March, 2019

Introduction. For a given vector $z = (z_1, \ldots, z_n)^\top$ from $\mathbb{R}^n$, $n \in \mathbb{N}$, the finite difference operator of order 2 is defined as follows

$$\Delta^2 z_i = \Delta^1 z_{i+1} - \Delta^1 z_i = z_{i+2} - 2z_{i+1} + z_i,\ 1 \leqslant i \leqslant n-2,$$

where $\Delta^1 z_i = z_{i+1} - z_i$ is the finite difference operator of order 1 at z_i.

A vector $z = (z_1, \ldots, z_n)^\top \in \mathbb{R}^n$ is said to be convex if and only if $\Delta^2 z_i \geqslant 0$ for each $1 \leqslant i \leqslant n-2$. A vector $z = (z_1, \ldots, z_n)^\top \in \mathbb{R}^n$ is called 1-monotone (or monotone) if $z_{i+1} - z_i \geqslant 0$, $i = 1, \ldots, n-1$, and convex vectors also may be called 2-monotone vectors.

The shape-constrained problems in statistics (the task of finding the best fitting monotone regression is one of them) have attracted much attention in recent decades [1–6]. One of most examined problems is the problem of constructing monotone regression (also called isotonic regression) which is to find best fitted non-decreasing vector to a given vector. One can find the detailed review on isotone regression in the work of Robertson and Dykstra [7]. The papers of Barlow and Brunk [8], Dykstra [9], Best and Chakravarti [10], Best [11] consider the problem of finding monotone regression in quadratic and convex programming frameworks. Using mathematical programming approach, the works [12–14] have recently provided some new results on the topic. The papers [15, 16] extend the problem to particular orders defined by the variables of a multiple regression. The recent paper [1] proposes and analyzes a dual active-set algorithm for regularized monotonic regression.

k-monotone regression is the extension of monotone regression to the general case of k-monotonicity [17]. Both isotone and k-monotone regression have applications in many fields, including the non-parametric mathematical statistics [3, 18], the empirical data smoothing [19–21], the shape-preserving dynamic programming [22], the shape-preserving approximation [23–25].

Denote Δ_2^n the set of all vectors from $\mathbb{R}^n$, which are convex. The task of constructing convex regression is to obtain a vector $z \in \mathbb{R}^n$ with the lowest square error of approximation to the given vector $y \in \mathbb{R}^n$ (not necessarily convex) under condition of convexity of z:

$$(z-y)^\top (z-y) = \sum_{i=1}^{n} (z_i - y_i)^2 \to \min_{z \in \Delta_2^n}. \qquad (1)$$

In this paper using the ideas of [26] we develop a new algorithm (the dual active set algorithm) for finding sparse convex regression. We prove that the dual

Sergei P. Sidorov https://orcid.org/0000-0003-4047-8239
Dr. Phys. & Math. Sci.; Head of Department; Dept. of Functions and Approximations Theory; e-mail: sidorovsp@yahoo.com

Sergei V. Tyshkevich https://orcid.org/0000-0001-5417-6165
Cand. Phys. & Math. Sci.; Associate Professor; Dept. of Functions and Approximations Theory; e-mail: tyszkiewicz@yandex.ru

active set algorithm proposed in this paper is optimal, i.e. the solutions obtaining by the algorithm are optimal.

1. The dual active set algorithm for convex regression.

Preliminary Analysis. The problem (1) can be rewritten in the form of a convex programming problem with linear constraints as follows

$$F(z) = \frac{1}{2} z^\top z - y^\top z \to \min, \tag{2}$$

where minimum is taken over all $z \in \mathbb{R}^n$ such that

$$g_i(z) := -(z_{i+2} - 2z_i + z_i) \leqslant 0, \quad 1 \leqslant i \leqslant n-2. \tag{3}$$

The problem (2)–(3) is a quadratic programming problem and is strictly convex, and therefore it has a unique solution.

Let $\hat{z}$ be a (unique) global solution of (2)–(3), then there is a Lagrange multiplier $\mu = (\mu_1, \ldots, \mu'_{n-2})^\top \in \mathbb{R}^{n-2}$ such that

$$\nabla F(z) + \sum_{i=1}^{n-2} \mu_i \nabla g_i(z) = 0, \tag{4}$$

$$g_i(z) \leqslant 0, \quad 1 \leqslant i \leqslant n-2, \tag{5}$$

$$\mu_i \geqslant 0, \quad 1 \leqslant i \leqslant n-2, \tag{6}$$

$$\mu_i g_i(z) = 0, \quad 1 \leqslant i \leqslant n-2, \tag{7}$$

where ∇g_i denotes the gradient of g_i. The equations (4)–(7) are the well-known Karush–Kuhn–Tucker optimality conditions. It follows from (4) that

$$\frac{\partial}{\partial z_i}\left[\frac{1}{2}\sum_{i=1}^{n}(z_i - y_i)^2 + \sum_{i=1}^{n-2}\mu_i(-z_{i+2} + 2z_{i+1} - z_i)\right] = 0, \quad 1 \leqslant i \leqslant n-2,$$

i.e.

$$\begin{cases} z_1 - y_1 - \mu_1 = 0 \\ z_2 - y_2 - \mu_2 + 2\mu_1 = 0 \\ z_3 - y_3 - \mu_3 + 2\mu_2 - \mu_1 = 0 \\ \ldots \\ z_j - y_j - \mu_j + 2\mu_{j-1} - \mu_{j-2} = 0 \\ \ldots \\ z_{n-2} - y_{n-2} - \mu_{n-2} + 2\mu_{n-3} - \mu_{n-4} = 0 \\ z_{n-1} - y_{n-1} + 2\mu_{n-2} - \mu_{n-3} = 0 \\ z_n - y_n - \mu_{n-2} = 0 \end{cases} \tag{8}$$

One can ask the natural question: is it possible to reduce the problem of constructing convex regression for points $y = (y_1, y_2, \ldots, y_n)$ to the problem of constructing monotone regression for points $\Delta y = (y_2 - y_1, y_3 - y_2, \ldots, y_n - y_{n-1})$? The answer to this question is negative. First of all, the Karush–Kuhn–Tucker optimality conditions for these problems are not identical. Secondly, we give a simple example showing that it is not possible to lower the order of monotonicity.

If we take $y = (1,0,1,0,1)$ then $\Delta y = (-1,1,-1,1)$. Optimal monotone regression for Δy is $\Delta z = \left(-\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, 1\right)$. The convex regression recovered from Δz is $\left(1, \frac{2}{3}, \frac{1}{3}, 0, 1\right)$. The square error of approximation to $y = (1,0,1,0,1)$ is equal to $0 + \left(\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + 0 + 0 = \frac{8}{9}$.

On the other hand, if we take the convex sequence $z = \left(1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1\right)$, then the error of approximation to the given vector $y = (1,0,1,0,1)$ is equal to $0 + \left(\frac{1}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + 0 = \frac{5}{9}$, which is less than the error for the convex regression recovered from monotone regression for Δy.

If one sums up all equalities in (8) then

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} y_i. \tag{9}$$

If we multiply s-th equality in (8) by s and sum up the resulting equalities, we obtain

$$\sum_{i=1}^{n} i z_i = \sum_{i=1}^{n} i y_i. \tag{10}$$

Thus, equalities (9)–(10) are necessary conditions for optimal solution.

Preliminary analysis of (4)–(7) shows that the first order differences of the optimal solution $\hat{z}$ should be sparse, i.e. the sequence $\Delta^1 \hat{z}_i$, $1 \leqslant i \leqslant n-2$, should have many zeroes. It follows from the Karush–Kuhn–Tucker conditions that optimal points should lie on a piecewise linear function. Note that some of these linear functions may be linear regressions constructed from the points of corresponding blocks. In this case the necessary conditions (9)–(10) are fulfilled.

A dual active-set algorithm for convex regression. In this subsection we propose an algorithm such that

- it is with polynomial complexity, i.e. the number of operations required to complete the algorithm for a given input y from $\mathbb{R}^n$ is $O(n^k)$ for some nonnegative integer k;
- the solution is convex (the reachability analysis will show this);
- the solution is optimal (the Karush–Kuhn–Tucker conditions are fulfilled).

The proposed algorithm uses so called active set. The active set S consists of blocks of the form $[l, r-2] \subset [1, n-2]$, such that $[l, r-2] \subset S$, $l-1 \notin S$, $r-1 \notin S$, and

$$S = [l_1, r_1] \cup [l_2, r_2] \cup \ldots \cup [l_{m-1}, r_{m-1}] \cup [l_m, r_m],$$

where $l_1 \geqslant 1$, $r_m \leqslant n-2$, $r_i + 3 \leqslant l_{i+1}$, $i \in [1, m-1]$, and m is the number of blocks. If $r_i = l_i$ then the i-th block consists of only one point.

Points $z_{r_i}, z_{r_i+1}, \ldots, z_{l_i}, z_{l_i+1}, z_{l_i+2}$ corresponding to the i-th block (plus two points from right) lies on a linear line, for each i.

At each iteration of the algorithm, the active set $S \subset [1, n-2]$ is chosen and the corresponding optimization problem is solved

$$\frac{1}{2} \sum_{i=1}^{n} (z_i - y_i)^2 \to \min, \tag{11}$$

where the minimum is taken over all $z \in \mathbb{R}^n$ satisfying

$$z_i - 2z_{i+1} + z_{i+2} = 0 \quad \forall\, i \in S. \tag{12}$$

Note that there exists a unique solution to the problem (11)–(12). Let us denote the solution by $z(S)$.

THE DUAL ACTIVE-SET ALGORITHM FOR CONVEX REGRESSION

begin
- · Input $y \in \mathbb{R}^n$;
- · Active set $S = \varnothing$;
- · Initial point $z(S) = y$;
- · **while** $z(S) \notin \Delta_2^n$ **do**
  - · set $S \leftarrow S \cup \{r_i : \; z_i(S) - 2z_{i+1}(S) + z_{i+2}(S) < 0\}$;
  - · solve (11)–(12) using points from the active set S;
  - · update $z(S)$;
- · Return $z(S)$;

end

The computational complexity of the dual active set algorithm for convex regression is $O(n^3)$. It follows from two remarks:
- at each iteration of the algorithm, the active set S is attaching, at least, one index from $[1 : n-2]$, which means that the number of the while loop iterations can not be greater that $n-2$.
- the computational complexity of solving the problem (11)–(12) is $O(n^2)$.

2. The convergence analysis of the dual active set algorithm.

Lemmas. The next lemma finds the values of Lagrange multipliers.

LEMMA 1. *Let z be a global solution of (2)–(3). Then the values of the Lagrange multipliers $\mu = (\mu_1, \ldots, \mu_{n-2})^\top \in \mathbb{R}^{n-2}$ defined in (4)–(7) are*

$$\mu_i = \sum_{j=1}^{i} (i - j + 1)(z_j - y_j), \quad 1 \leqslant i \leqslant n-2. \tag{13}$$

Proof. Lemma can be proved by induction. If $i = 1$ then it follows from (8) that

$$z_1 - y_1 = \mu_1. \tag{14}$$

Then from (8) and (14) we obtain

$$z_2 - y_2 = \mu_2 - 2\mu_1 = \mu_2 - 2(z_1 - y_1),$$

and therefore

$$\mu_2 = (z_2 - y_2) + 2(z_1 - y_1).$$

Suppose that (13) is fulfilled for $i > 2$. Then it follows from (8) that

$$z_{i+1} - y_{i+1} = \mu_{i+1} - 2\mu_i + \mu_{i-1},$$

and consequently we have

$$\begin{aligned}\mu_{i+1} &= (z_{i+1} - y_{i+1}) + 2\mu_i - \mu_{i-1} = \\ &= (z_{i+1} - y_{i+1}) + 2\sum_{j=1}^{i}(i-j+1)(z_j - y_j) - \sum_{j=1}^{i-1}(i-j)(z_j - y_j) = \\ &= (z_{i+1} - y_{i+1}) + 2(z_i - y_i) + \sum_{j=1}^{i-1}\big(2(i-j+1) - (i-j)\big)(z_j - y_j) = \\ &= (z_{i+1} - y_{i+1}) + 2(z_i - y_i) + \sum_{j=1}^{i-1}(i-j+2)(z_j - y_j) = \\ &= \sum_{j=1}^{i+1}(i-j+2)(z_j - y_j).\end{aligned}$$

□

LEMMA 2. *Let $1 \in S$, i.e. $\Delta^2 y_1 < 0$, and assume that $2, 3 \notin S$. Let z_1, z_2, z_3 be the values of linear regression constructed for input points $(1, y_1)$, $(2, y_2)$, $(3, y_3)$ (see Fig. 1(a)). Then the values of corresponding Lagrange multipliers defined in (13) are non-negative.*

Proof. We will show that

$$\begin{aligned}\mu_1 &= z_1 - y_1 \geqslant 0, \\ \mu_2 &= 2(z_1 - y_1) + (z_2 - y_2) = 0, \\ \mu_3 &= 3(z_1 - y_1) + 2(z_2 - y_2) + (z_3 - y_3) = 0.\end{aligned}$$

We have $\mu_1 = z_1 - y_1 > 0$, since y_1, y_2, y_3 lie on a concave line. It follows from (9) and (10) that

$$\sum_{i=1}^{3}(z_i - y_i) = 0 \tag{15}$$

and

$$\sum_{i=1}^{3} i(z_i - y_i) = 0. \tag{16}$$

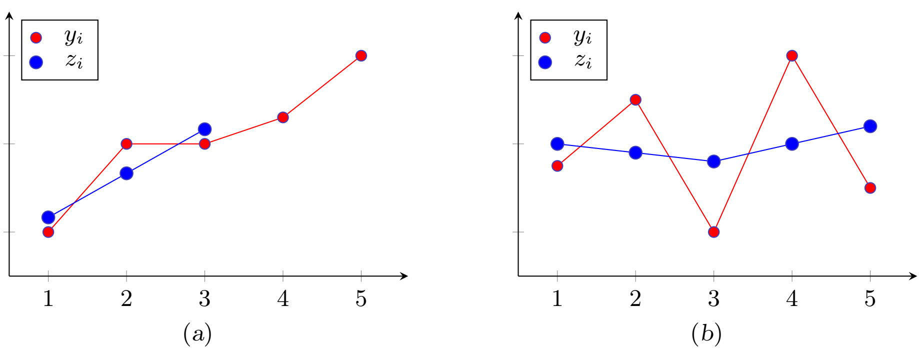

Figure 1. (a) The case $1 \in S$, i.e. $\Delta^2 y_1 < 0$, and $2, 3 \notin S$, z_1, z_2, z_3 are the values of linear regression constructed for input points $(1, y_1)$, $(2, y_2)$, $(3, y_3)$. (b) The case $1, 3 \in S$, i.e. $\Delta^2 y_1 < 0$, $\Delta^2 y_3 < 0$, and $2, 4, 5 \notin S$, z_i's are the solutions to the problem (17)

If we multiply (15) by 3 and subtract (16), we get $2(z_1 - y_1) + (z_2 - y_2) = 0$, and therefore $\mu_2 = 0$.

If we multiply (15) by 4 and subtract (16), we get $3(z_1 - y_1) + 2(z_2 - y_2) + (z_3 - y_3) = 0$, and therefore $\mu_3 = 0$. □

LEMMA 3. *Let y_1, y_2, y_3, y_4, y_5 be such that $1, 3 \in S$, i.e. $\Delta^2 y_1 < 0$, $\Delta^2 y_3 < 0$, and assume that $2, 4, 5 \notin S$ (see Fig. 1(b)). Let z_i be the solutions to the optimization problem*

$$\frac{1}{2}\sum_{i=1}^{5}(z_i - y_i)^2 \to \min, \tag{17}$$

where minimum is taken over all $z = (z_1, \ldots, z_5)^\top$ such that $\Delta^2 z_1 = \Delta^2 z_3 = 0$. Then the values of corresponding Lagrange multipliers $\mu_1, \ldots, \mu_5$ defined in (13) are non-negative.

Proof. Since $\Delta^2 z_1 < 0$ we have $\mu_1 = z_1 - y_1 \geqslant 0$. The Lagrangian for the constrained optimization problem (17) is

$$L(z, \lambda_1, \lambda_2) = \frac{1}{2}\sum_{i=1}^{5}(z_i - y_i)^2 - \lambda_1(z_1 - 2z_2 + z_3) - \lambda_3(z_3 - 2z_4 + z_5),$$

where multipliers λ_1, λ_2 satisfy

$$\frac{\partial L}{\partial z_1} = z_1 - y_1 - \lambda_1 = 0, \qquad \frac{\partial L}{\partial z_2} = z_2 - y_2 + 2\lambda_1 = 0,$$

$$\frac{\partial L}{\partial z_3} = z_3 - y_3 - \lambda_1 - \lambda_2 = 0, \qquad \frac{\partial L}{\partial z_4} = z_4 - y_4 + 2\lambda_2 = 0,$$

$$\frac{\partial L}{\partial z_5} = z_5 - y_5 - \lambda_2 = 0.$$

Then

$$\mu_2 = 2(z_1 - y_1) + (z_2 - y_2) = 2\lambda_1 - 2\lambda_1 = 0,$$

$$\mu_3 = 3(z_1 - y_1) + 2(z_2 - y_2) + (z_3 - y_3) = 3\lambda_1 - 4\lambda_1 + \lambda_1 + \lambda_2 = \lambda_2,$$

$$\mu_4 = 4(z_1 - y_1) + 3(z_2 - y_2) + 2(z_3 - y_3) + (z_4 - y_4) = 4\lambda_1 - 6\lambda_1 + 2\lambda_1 + 2\lambda_2 - 2\lambda_2 = 0,$$

$$\mu_5 = 5(z_1 - y_1) + 4(z_2 - y_2) + \ldots + (z_5 - y_5) = 5\lambda_1 - 8\lambda_1 + 3\lambda_1 + 3\lambda_2 - 4\lambda_2 + \lambda_2 = 0.$$

Since $\Delta^2 z_3 < 0$ we have $\lambda_2 = z_5 - y_5 > 0$, and therefore $\mu_3 > 0$. □

If $z_2 - z_1 \leqslant z_4 - z_3$ (as it is shown in Fig. 1(b)) then $\Delta^2 z_i \geqslant 0$ for all $i \in [1:3]$, and the algorithm constructed a convex solution. If $z_2 - z_1 > z_4 - z_3$ (as it is in Fig. 2(a)), then the point 2 must be added to the active set $S = \{1, 3\}$ at the next iteration of the algorithm and the block $\{1, 2, 3\}$ of the active set S will be formed. The solution to the optimization problem (11)–(12) corresponding to the block $\{1, 2, 3\}$ are points lying on linear regression line constructed for points $(1, z_1)$, ..., $(5, z_5)$ (see Fig. 2(b)). Note that the same regression will be constructed for the initial points $(1, y_1)$, ..., $(5, y_5)$.

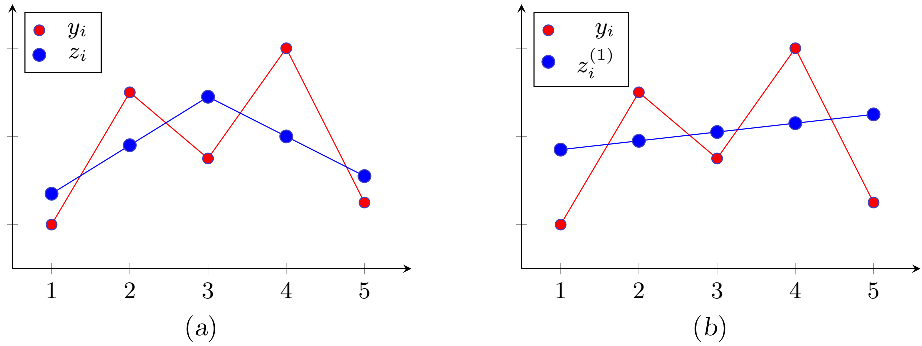

Figure 2. (a) The case $1, 3 \in S$, i.e. $\Delta^2 y_1 < 0$, $\Delta^2 y_3 < 0$, and $2, 4, 5 \notin S$, z_i's are the solutions to the problem (17), $z_2 - z_1 > z_4 - z_3$. (b) The solution to the optimization problem (11)–(12) corresponding to the block $\{1, 2, 3\} \subset S$ are points lying on linear regression line constructed for points $(1, z_1), \ldots, (5, z_5)$

Lemma 4. *Let at an iteration of the algorithm the pairs*

$$Y = \{(1, z_1), \ldots, (k, z_{k+1})\}$$

be such that $[1 : k-2] \subset S$, $\Delta^2 z_i < 0$ *for all* $i \in [1 : k-2]$, *and* $k-1, k, k+1 \notin S$ (*see Fig.* 3). *Let* $z_i^{(0)}$, $i \in [1 : k]$, *be the solution to the optimization problem*

$$\frac{1}{2}\sum_{i=1}^{k}(\zeta_i - z_i)^2 \to \min,$$

where the minimum is taken over all $\zeta \in \mathbb{R}^k$ *such that* $\zeta_i - 2\zeta_{i+1} + \zeta_{i+2} = 0$ *for all* $i \in [1 : k-2]$. *Assume that* $z_{k-1}^{(0)} - 2z_k^{(0)} + z_{k+1} < 0$, *i.e. the point* $k-1$ *should be added to the active set* S *at the next iteration of the algorithm. Let* $z_i^{(1)}$, $i \in [1 : k+1]$, *be the solution to the optimization problem*

$$\frac{1}{2}\sum_{i=1}^{k+1}(\zeta_i - z_i)^2 \to \min,$$

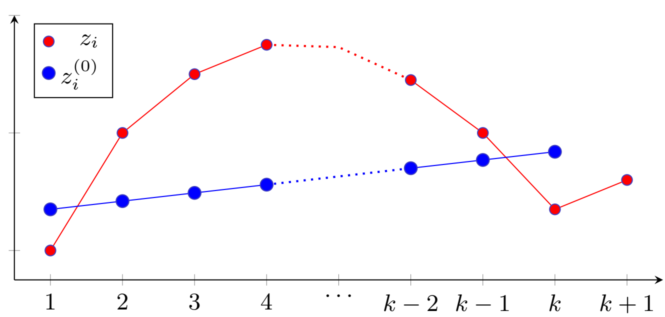

Figure 3. The case $1, 2, \ldots, k-2 \in S$ and $k-1 \notin S$, $z_i^{(0)}$'s are the values of linear regression constructed for input points (i, z_i), $i \in [1 : k]$. After the iteration, the point $k-1$ must be added to the block $[1 : k-2]$ since $z_{k-1}^{(0)} - 2z_k^{(0)} + z_{k+1} < 0$

where the minimum is taken over all $\zeta \in \mathbb{R}^{k+1}$ *satisfying* $\zeta_i - 2\zeta_{i+1} + \zeta_{i+2} = 0$ *for all* $i \in [1:k-1]$. *Then for every* $i = 1, 2, \dots, k$ *we have*

$$\mu_i^{(1)} - \mu_i^{(0)} \geqslant 0,$$

where $\mu_i^{(0)}$, $\mu_i^{(1)}$ *are Lagrange multipliers for* $z^{(0)}$ *and* $z^{(1)}$ *respectively, and* $\mu_k^{(1)} = \mu_{k+1}^{(1)} = 0$.

Proof. Note that $z_i^{(0)}$ are the values of the linear regression which uses input data points $\{(1, z_1), \dots, (k, z_k)\}$, obtained at points $i \in [1:k]$. By the same reason, $z_i^{(1)}$ are the values of linear regression which uses input data points Y points $i \in [1:k+1]$.

It follows from (13) that

$$\mu_i^{(0)} = \sum_{j=1}^{i} (i+1-j)(z_j^{(0)} - z_j), \quad \mu_i^{(1)} = \sum_{j=1}^{i} (i+1-j)(z_j^{(1)} - z_j)$$

for all $i \in [1:k]$, and therefore, we have

$$\mu_i^{(1)} - \mu_i^{(0)} = \sum_{j=1}^{i} (i+1-j)(z_j^{(1)} - z_j^{(0)}). \tag{18}$$

Without the loss of generality we will assume that points $z_i^{(0)}$, $i \in [1:k]$, lie on a straight line passing through the origin and with a slope a, where $-1 < a < 1$, i.e.

$$z_j^{(0)} = aj, \; j = 1, \dots, k. \tag{19}$$

Since $z_{k-1}^{(0)} - 2z_k^{(0)} + z_{k+1} < 0$, there exists $d > 0$ such that $z_{k+1} = a(k+1) - d$. In this case the points $z_i^{(1)}$, $i = 1, \dots, k+1$, lie on a straight line with a slope $a - \frac{6d}{(k+1)(k+2)}$ and intercept $\frac{2d}{k+1}$, i.e.

$$z_j^{(1)} = j\Big(a - \frac{6d}{(k+1)(k+2)}\Big) + \frac{2d}{k+1}, \quad j = 1, \dots, k+1. \tag{20}$$

It follows from (18), (19) and (20) that

$$\mu_i^{(1)} - \mu_i^{(0)} = \sum_{j=1}^{i} (i+1-j)\left(j\Big(a - \frac{6d}{(k+1)(k+2)}\Big) + \frac{2d}{k+1} - aj\right) =$$

$$= \sum_{j=1}^{i} (i+1-j)\Big(-j\frac{6d}{(k+1)(k+2)} + \frac{2d}{k+1}\Big) =$$

$$= -(i+1)\frac{6d}{(k+1)(k+2)} \sum_{j=1}^{i} j + (i+1)\frac{2d}{k+1} \sum_{j=1}^{i} 1 +$$

$$+\frac{6d}{(k+1)(k+2)}\sum_{j=1}^{i} j^2 - \frac{2d}{k+1}\sum_{j=1}^{i} j =$$

$$= -(i+1)\frac{6d}{(k+1)(k+2)}\frac{i(i+1)}{2} + (i+1)\frac{2d}{k+1}i+$$

$$+\frac{6d}{(k+1)(k+2)}\frac{i(i+1)(2i+1)}{6} - \frac{2d}{k+1}\frac{i(i+1)}{2} =$$

$$= \frac{di(i+1)}{(k+1)}\Big(-\frac{3(i+1)}{k+2} + 2 + \frac{2i+1}{k+2} - 1\Big) = \frac{di(i+1)(k-i)}{(k+1)(k+2)} \geqslant 0. \qquad \square$$

REMARK. *It follows from Lemmas 2 and 4 that the values of Lagrange multipliers are increasing when adding a point to an existing block from right hand side.*

LEMMA 5. *Suppose that before an iteration of the algorithm the pairs*

$$Y = \{(1, z_1), \dots, (k, z_k)\}$$

are such that $[2 : k-2] \subset S$, *i.e.* $\Delta^2 z_i = 0$ *for all* $i \in [2 : k-2]$, *and* $k-1, k \notin S$. *If the inequality* $z_1 - 2z_2 + z_3 < 0$ *holds* (*see Fig.* 4), *i.e.* 1 *must be added to the block* $[2 : k-2] \subset S$ *at the iteration, then we obtain* $\mu_i \geqslant 0$ *for every* $i \in [1 : k]$, *where* μ_i, $i \in [1 : k]$, *corresponds to the solutions* $z_i^{(1)}$, $i \in [1 : k]$, *of the optimization problem*

$$\frac{1}{2}\sum_{i=1}^{k}(\zeta_i - z_i)^2 \to \min,$$

where the minimum is taken over all $\zeta \in \mathbb{R}^k$ *such that* $\Delta^2 \zeta_i = 0$ *for all* $i \in [1 : k-2]$.

Proof. If we take $S = \{1\}$, then it follows from Lemma 2 that $\mu_1^{(0)} \geqslant 0$, $\mu_2^{(0)} = \mu_3^{(0)} = 0$. Now let us consider the sequence of the basic operations of adding one point to a block from the right hand side:

– $S = \{1, 2\}$, then we have $\mu_1^{(1)}, \mu_2^{(1)} \geqslant 0$, $\mu_3^{(1)} = \mu_4^{(1)} = 0$ by Remark.

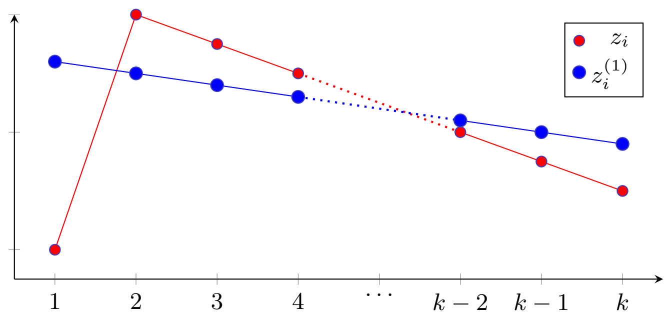

Figure 4. The case $[2 : k-2] \subset S$, i.e. $\Delta^2 z_i = 0$ for all $i \in [2 : k-2]$, and $k-1, k \notin S$. After the iteration, the point $k-1$ must be added to the block $[1 : k-2]$ since $z_1 - 2z_2 + z_3 < 0$. Points $z_i^{(1)}$, $i \in [1 : k]$, are the solutions to the optimization problem (5).

– $S = \{1,2,3\}$, then we have $\mu_1^{(2)}, \mu_2^{(2)}, \mu_3^{(2)} \geqslant 0$, $\mu_4^{(2)} = \mu_5^{(2)} = 0$ (by Remark).
– ...
– $S = \{1,2,\ldots,k-2\}$, then $\mu_1^{(k-3)}, \ldots, \mu_{k-2}^{(k-3)} \geqslant 0$, $\mu_{k-1}^{(k-3)} = \mu_k^{(k-3)} = 0$ (by Remark).

At each step we add a point to an existing block of the active set from the right hand side, and therefore the values of Lagrange multipliers are increasing, i.e. they are remaining non-negative. □

LEMMA 6. *Suppose that before an iteration of the algorithm the pairs*

$$Y_1 = \{(1, z_1), \ldots, (k-3, z_{k-3})\}, \quad Y_2 = \{(k, z_k), \ldots, (s, z_p)\}, \quad 4 \leqslant k \leqslant p,$$

are such that $[1 : k-3] \subset S$, $[k : p-2] \subset S$, $k-2, k-1 \notin S$, *i.e.* $\Delta^2 z_i = 0$ *for all* $i \in [1 : k-3]$, $\Delta^2 z_i = 0$ *for all* $i \in [k : p-2]$, *and* $p-1, p \notin S$. *Assume that one of the following conditions is fulfilled*:

– (*the case* 1) *the inequality* $z_{k-2} - 2z_{k-1} + z_k < 0$ *holds* (*see Fig.* 5), *i.e.* $k-2$ *must be added to* S *at the iteration*;
– (*the case* 2) *the inequality* $z_k - 2z_{k+1} + z_{k+2} < 0$ *holds* (*see Fig.* 6), *i.e.* k *must be added to* S *at the iteration.*

Let $z_i^{(1)}$, $i \in [1 : p]$, *be the solutions to the optimization problem*

$$\frac{1}{2} \sum_{i=1}^{p} (\zeta_i - z_i)^2 \to \min, \tag{21}$$

where the minimum is taken over all $\zeta \in \mathbb{R}^p$ *satisfying* $\zeta_i - 2\zeta_{i+1} + \zeta_{i+2} = 0$ *for all* $i \in [1 : p-2] \setminus \{k-1\}$. *Then for every* $i = 1, 2, \ldots, p$ *we have*

$$\mu_i^{(1)} \geqslant 0,$$

where $\mu_i^{(1)}$ *are the values of Lagrange multipliers corresponding to the solutions* $z_i^{(1)}$, $i \in [1 : p]$.

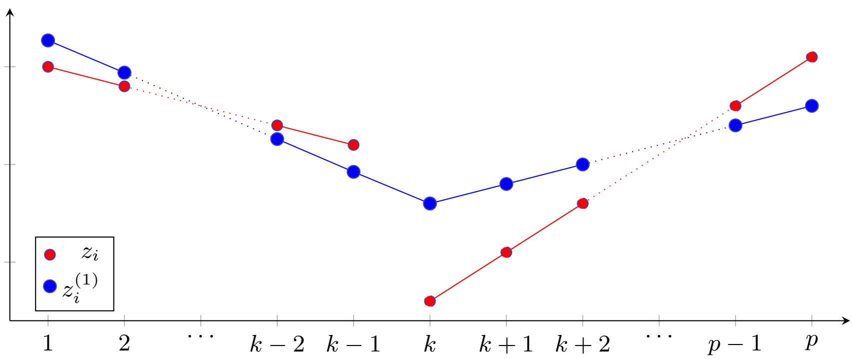

Figure 5. The case 1 of merging two blocks

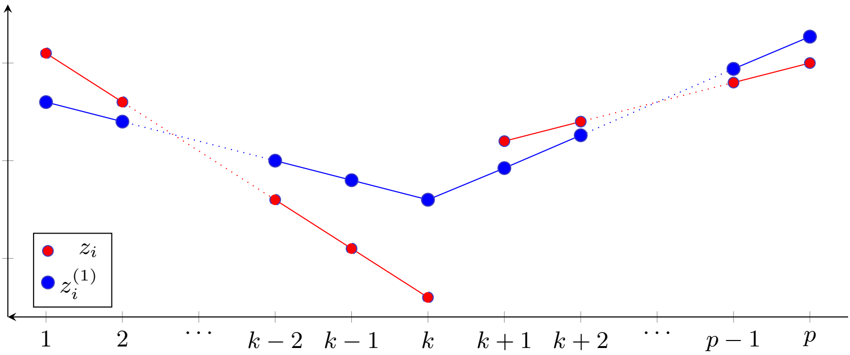

Figure 6. The case 2 of merging two blocks

Proof. The case 1. If $z_i^{(1)}$, $i \in [1 : p]$, are the optimal solutions to the problem (21) then there exists a point z_k^* such that

– $z_{k-2} - 2z_{k-1} + z_k^* < 0$,

– $z_i^{(1)}$, $i \in [1 : k]$, are lying on the regression line constructed by points $(1, z_1), \ldots, (k-1, z_{k-1}), (k, z_k^*)$.

It follows from Lemma 1 that $\mu_1, \mu_2, \ldots, \mu_{k-1} \geqslant 0$, $\mu_{k-1} = 0$.

There exists a point z'_{p+1} such that

– the inequality $z_{p-1} - 2z_p + z'_{p+1} < 0$ is fulfilled,

– points $z_i^{(1)}$, $i \in [k : p]$, are lying on the regression line constructed by points $(k, z_k), (k+1, z_{k+1}), \ldots, (p, z_p), (p+1, z'_{p+1})$.

Then it follows from Lemma 1 that $\mu_k, \mu_{k+1}, \ldots, \mu_p \geqslant 0$, $\mu_{p+1} = 0$.

The case 2. If $z_i^{(1)}$, $i \in [1 : p]$, are the optimal solutions to the problem (21) then there exists a point z_0^* such that

– the inequality $z_0^* - 2z_1 + z_2 < 0$ holds,

– points $z_i^{(1)}$, $i \in [1 : k]$, are lying on the regression line constructed by points $(0, z_0^*), (1, z_1), \ldots, (k-1, z_{k-1}), (k, z_k)$.

It follows from Lemma 5 that $\mu_1, \mu_2, \ldots, \mu_{k-1} \geqslant 0$, $\mu_{k-1} = 0$.

There exists a point z'_k such that

– the inequality $z'_k - 2z_{k+1} + z_{k+2} < 0$ is fulfilled,

– points $z_i^{(1)}$, $i \in [k : p]$, are lying on the regression line constructed by points $(k, z'_k), (k+1, z_{k+1}), \ldots, (p, z_p))$.

Then it follows from Lemma 5 that $\mu_k, \mu_{k+1}, \ldots, \mu_{p-1} \geqslant 0$, $\mu_p = 0$. □

Availability and optimality analysis. When the algorithm is completed, the output vector has the form

$$(\underbrace{z_1, \ldots, z_{r_1}}_{\lambda_1 \text{ times}}, \underbrace{z_{r_1+1}, \ldots, z_{r_2}}_{\lambda_2 \text{ times}}, \ldots, \underbrace{z_{r_{m-1}+1}, \ldots, z_n}_{\lambda_m \text{ times}}),$$

where

– m is the number of segments;

– λ_j is the number of points at jth segment, $\lambda_1 = m_1$, $m_2 = \lambda_1 + \lambda_2$, $\ldots$, $m_{i+1} = \lambda_1 + \lambda_2 + \ldots + \lambda_i$, $n = \lambda_1 + \lambda_2 + \ldots \lambda_m$.

All points z_j of each segment $[r_i+1, \dots, r_{i+1}]$ lie on a straight line. A segment can consists of one point.

The solution obtained as a result of the algorithm is convex, by design of the algorithm. It remains to show that the solution is optimal, i.e. it is necessary to show that the Karush–Kuhn–Tucker conditions are fulfilled.

The algorithm starts with any active set such that $S \subset S^*$, where S^* is the active set corresponding the optimal solution. At he first iteration we can always take $S = \emptyset$ for simplicity.

THEOREM. *For any initial $S \subset S^*$, the Algorithm converges to the optimal solution of the problem (1) in, at most, $n - 1 - |S|$ iterations.*

Proof. At each while loop of the algorithm, the active set S is expanded by attaching at least one index point from $[1, n-2]$, which is previously not belonged to the set S. The Algorithm terminates when $z(S)$ becomes convex. If $S = [1, n-2]$, then the number of blocks is equal to 1 and, therefore, there is no violation of the convexity. If $|S| < n-2$ then the number of iterations must be less than $n - |S|$, where $|S|$ is the number of indices in the initial active set S.

The optimality of the solution follows from Lemmas 2–6.

At each iteration of the algorithm, the points at which the convexity is violated are attached to the active set S. If one of such points i with a negative value of the second order finite difference $\Delta^2 z_i$ is isolated (i.e. $i-2, i-1, i+1, i+2 \notin S$), then the algorithm replaces z_i, z_{i+1}, z_{i+2} with the values of linear regression constructed by the points $(i, z_i), (i+1, z_{i+1}), (i+2, z_{i+2})$. Lemma 2 states that the values of the corresponding Lagrange multipliers will be non-negative.

If the second-order differences are negative at several consecutive $k \geqslant 1$ neighboring points, e.g. $z_i, z_{i+1}, \dots, z_{i+k}$ are such that $\Delta^2 z_j < 0$, $j = i, \dots, i+k$, and $\Delta^2 z_{i-2}, \Delta^2 z_{i-1}, \Delta^2 z_{i+k+1}, \Delta^2 z_{i+k+2} \geqslant 0$, then the algorithm replaces $z_i, z_{i+1}, \dots, z_{i+k+2}$ with the values of a linear regression constructed by the points $(i, z_i), (i+1, z_{i+1}), (i+k+2, z_{i+k+2})$. Lemma 4 states that the values of the corresponding Lagrange multipliers will be non-negative.

When the algorithm works on subsequent iterations, it may turn out that a point with a negative second-order finite difference is not isolated. In this case the point where the violation of convexity takes place will be attached to the one of the neighboring blocks of the active set S as follows:

– Let $r < r_1 < r_2 < \dots < r_s < p-2$ be such that $r+1 < r_1, r_1+1 < r_2, \dots, r_j+1 < r_{j+1}, \dots, r_s+1 < p-2$. Suppose that $[r : p-2] \subset S$ and $r_1, r_2, \dots, r_s \notin S$ at the previous iteration of the algorithm. If $p-1$ must be added to the active set S at the current iteration of the algorithm, i.e. $z_{p-1} - 2z_p + z_{p+1} < 0$ for the current solution z, then it follows from Lemma 1 that the values of the corresponding Lagrange multipliers will increase, and therefore they will remain non-negative.

– Let $r < r_1 < r_2 < \dots < r_s < p-2$ be such that $r+1 < r_1, r_1+1 < r_2, \dots, r_j+1 < r_{j+1}, \dots, r_s+1 < p-2$. Suppose that $[r : p-2] \subset S$ and $r_1, r_2, \dots, r_s \notin S$ at the previous iteration of the algorithm. If $r-1$ must be added to the active set S at the current iteration of the algorithm, i.e. $\Delta^2 z_{r-1} < 0$ for the current solution z, then it follows from Lemma 5 that the values of the corresponding Lagrange multipliers will increase, and therefore they will remain non-negative.

– Let $r < r_1 < r_2 < \dots < r_s < p-2$ be such that $r+1 < r_1, r_1+1 <$

$r_2, \ldots, r_j + 1 < r_{j+1}, \ldots, r_s + 1 < p - 2$. Suppose that $[r : p-2] \subset S$ and $r_1, r_2, \ldots, r_s \notin S$ at the previous iteration of the algorithm. If p must be added to the active set S at the current iteration of the algorithm, i.e. $z_p - 2z_{p+1} + z_{p+2} < 0$ for the current solution z, then it follows from Lemma 6 that the values of the corresponding Lagrange multipliers will increase, and therefore they will remain non-negative.

– Let $r < r_1 < r_2 < \ldots < r_s < p - 2$ be such that $r + 1 < r_1, r_1 + 1 < r_2, \ldots, r_j + 1 < r_{j+1}, \ldots, r_s + 1 < p - 2$. Suppose that $[r : p-2] \subset S$ and $r_1, r_2, \ldots, r_s \notin S$ at the previous iteration of the algorithm. If $r - 2$ must be added to the active set S at the current iteration of the algorithm, i.e. $\Delta^2 z_{r-2} < 0$ for the current solution z, then it follows from Lemma 6 that the values of the corresponding Lagrange multipliers will increase, and therefore they will remain non-negative.

The remaining cases of mergers for two or more adjacent blocks and isolated points can be represented as sequences of these basic cases. □

Conclusion. In this paper using the ideas of [26] we develop a new algorithm (the dual active set algorithm) for finding sparse convex regression. At each iteration of the algorithm, it first determines the active set and then solve a standard least-squares subproblem on the active set with small size, which exhibits a local superlinear convergence. Therefore, the algorithm is very efficient when coupled with a parallel execution. The classical optimization algorithms for nonconvex problems (e.g. coordinate descent or proximal gradient descent) only possess sublinear convergence in general or linear convergence under certain conditions [27]. On the other hand, k-FWA and k-PAVA examined in [17] are not optimal.

Competing interests. We declare that we have no conflicts of interest in the authorship and publication of this article.

Authors' contributions and responsibilities. Each author has participated in the article concept development and in the manuscript writing. The authors are absolutely responsible for submitting the final manuscript in print. Each author has approved the final version of manuscript.

Funding. This work was supported by the Russian Foundation for Basic Research (project no. 18–37–00060).

Acknowledgments. The authors are grateful to the referee for careful reading of the paper and his constructive comments.

References

1. Burdakov O., Sysoev O. A Dual Active-Set Algorithm for Regularized Monotonic Regression, *J. Optim. Theory Appl.*, 2017, vol. 172, no. 3, pp. 929–949. doi: 10.1007/s10957-017-1060-0.
2. Brezger A., Steiner W. J. Monotonic Regression Based on Bayesian P-Splines: An application to estimating price response functions from store-level scanner data, *J. Bus. Econ. Stat.*, 2008, vol. 26, no. 1, pp. 90–104. doi: 10.1198/073500107000000223.
3. Chen Y. *Aspects of Shape-constrained Estimation in Statistics*, PhD Thesis. Cambridge, UK, University of Cambridge, 2013, https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.648300.
4. Balabdaoui F., Rufibach K., Santambrogio F. Least-squares estimation of two-ordered monotone regression curves, *J. Nonparametric Stat.*, 2010, vol. 22, no. 8, pp. 1019–1037. doi: 10.1080/10485250903548729.

5. Hazelton M. L., Turlach B. A. Semiparametric Regression with Shape-Constrained Penalized Splines, *Comput. Stat. Data Anal.*, 2011, vol. 55, no. 10, pp. 2871–2879. doi: 10.1016/j.csda.2011.04.018.
6. Lu M. Spline estimation of generalised monotonic regression, *J. Nonparametric Stat.*, 2014, vol. 27, no. 1, pp. 19–39. doi: 10.1080/10485252.2014.972953.
7. Robertson T., Wright F. T. Dykstra R. L. *Order Restricted Statistical Inference*, Wiley Series in Probability and Mathematical Statistics. Chichester (UK) etc., John Wiley & Sons, 1988, xix+521 pp.
8. Barlow R. E. Brunk H. D. The Isotonic Regression Problem and its Dual, *J. Amer. Stat. Assoc.*, 1972, vol. 67, no. 337, pp. 140–147. doi: 10.1080/01621459.1972.10481216.
9. Dykstra R. L. An Isotonic Regression Algorithm, *J. Stat. Plan. Inf.*, 1981, vol. 5, no. 1, pp. 355–363. doi: 10.1016/0378-3758(81)90036-7.
10. Best M. J., Chakravarti N. Active set algorithms for isotonic regression: A unifying framework, *Mathematical Programming*, 1990, vol. 47, no. 3, pp. 425–439. doi: 10.1007/bf01580873.
11. Best M. J., Chakravarti N., Ubhaya V. A. Minimizing Separable Convex Functions Subject to Simple Chain Constraints, *SIAM J. Optim.*, 2000, vol. 10, no. 3, pp. 658–672. doi: 10.1137/s1052623497314970.
12. Ahuja R. K., Orlin J. B. A Fast Scaling Algorithm for Minimizing Separable Convex Functions Subject to Chain Constraints, *Operations Research*, 2001, vol. 49, no. 1, pp. 784–789. doi: 10.1287/opre.49.5.784.10601.
13. Strömberg U. An Algorithm for Isotonic Regression with Arbitrary Convex Distance Function, *Comput. Stat. Data Anal.*, 1991, vol. 11, no. 2, pp. 205–219. doi: 10.1016/0167-9473(91)90072-a.
14. Hansohm J. Algorithms and Error Estimations for Monotone Regression on Partially Preordered Sets, *J. Multivariate Analysis*, 2007, vol. 98, no. 5, pp. 1043–1050. doi: 10.1016/j.jmva.2006.11.001.
15. Burdakow O., Grimvall A., Hussian M. A Generalised PAV Algorithm for Monotonic Regression in Several Variables, In: *COMPSTAT, Proceedings in Computational Statistics*, 16th Symposium Held in Prague, Czech Republic, vol. 10; ed. J. Antoch. New York, Springer-Verlag, 2004, pp. 761–767, http://users.mai.liu.se/olebu87/PAPERS/COMPSTAT04.pdf.
16. Dykstra R. L., Robertson T. An Algorithm for Isotonic Regression for Two or More Independent Variables, *Ann. Statist.*, 1982, vol. 10, no. 3, pp. 708–716. doi: 10.1214/aos/1176345866.
17. Sidorov S. P., Faizliev A. R., Gudkov A. A., Mironov S. V. Algorithms for Sparse k-Monotone Regression, In: *Integration of Constraint Programming, Artificial Intelligence, and Operations Research*, Lecture Notes in Computer Science, 10848; ed. W.-J. van Hoeve. Cham, Springer International Publishing, 2018, pp. 546–556. doi: 10.1007/978-3-319-93031-2_39.
18. Bach F. R. *Efficient Algorithms for Non-convex Isotonic Regression through Submodular Optimization*, 2017, arXiv: 1707.09157 [cs.LG].
19. Altmann D., Grycko Eu., Hochstättler W., Klützke G. *Monotone smoothing of noisy data*, Technical Report feu-dmo034.15, FernUniversität in Hagen, 2014, https://www.fernuni-hagen.de/mathematik/DMO/pubs/feu-dmo034-14.pdf.
20. Gorinevsky D. M., Kim S.-J., Beard Sh., Boyd S. P., Gordon G. Optimal Estimation of Deterioration From Diagnostic Image Sequence, *IEEE Transactions on Signal Processing*, 2009, vol. 57, no. 3, pp. 1030–1043. doi: 10.1109/tsp.2008.2009896.
21. Hastie T. Tibshirani R. Wainwright M. *Statistical Learning with Sparsity: The Lasso and Generalizations*, Monographs on Statistics and Applied Probability, vol. 143. New York, Chapman and Hall/CRC, 2015, xvi+351 pp. doi: 10.1201/b18401.
22. Cai Y., Judd K. L. Advances in Numerical Dynamic Programming and New Applications, In: *Handbook of Computational Economics*, vol. 3; eds. K. Schmedders, K. L. Judd, Elsevier, 2014, pp. 479–516. doi: 10.1016/b978-0-444-52980-0.00008-6.

23. Boytsov D. I., Sidorov S. P. Linear approximation method preserving k-monotonicity, *Sib. Èlektron. Mat. Izv.*, 2015, vol. 12, pp. 21–27. doi: 10.17377/semi.2015.12.003.
24. Cullinan M. P. Piecewise Convex-Concave Approximation in the Minimax Norm, In: *Abstracts of Conference on Approximation and Optimization: Algorithms, Complexity, and Applications* (June 29–30, 2017, Athens, Greece); eds. I. Demetriou, P. Pardalos. Athens, National and Kapodistrian University of Athens, 2017, pp. 4.
25. Shevaldin V. T. *Approksimatciia lokalnymi splainami* [Local approximation by splines]. Ekaterinburg, Ural Branch of RAS, 2014, 198 pp. (In Russian)
26. Leeuw J. Hornik K. Mair P. Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods, *J. Stat. Softw.*, 2009, vol. 32, no. 5, pp. 1–24. doi: 10.18637/jss.v032.i05.
27. Li G., Pong T. K. Calculus of the Exponent of Kurdyka–Łojasiewicz Inequality and Its Applications to Linear Convergence of First-Order Methods, *Found. Comput. Math.*, 2017, vol. 18, no. 5, pp. 1199–1232, arXiv: 1602.02915 [math.OC]. doi: 10.1007/s10208-017-9366-8.

Вестн. Сам. гос. техн. ун-та. Сер. Физ.-мат. науки. 2019. Т. 23, № 1. С. 113–130
ISSN: 2310-7081 (online), 1991-8615 (print) https://doi.org/10.14498/vsgtu1673

УДК 519.65, 519.25, 519.853

Двойственный алгоритм на основе активного множества для построения оптимальной разреженной выпуклой регрессии

А. А. Гудков, С. В. Миронов, С. П. Сидоров, С. В. Тышкевич

Саратовский государственный университет им. Н. Г. Чернышевского
(национальный исследовательский университет),
Россия, 410012, Саратов, ул. Астраханская, 83.

Аннотация

В последнее время задачи статистики с ограничениями на форму данных привлекают повышенное внимание. Одной из таких задач является задача поиска оптимальной монотонной регрессии. Проблема построения монотонной регрессии (которая также называется изотонной регрессией) состоит в том, чтобы для данного вектора (не обязательно монотонного) найти неубывающий вектор с наименьшей ошибкой приближения к данному. Выпуклая регрессия есть развитие понятия монотонной регрессии для случая 2-монотонности (т.е. выпуклости). Как изотонная, так и выпуклая регрессия находят применение во многих областях, включая непараметрическую математическую статистику и сглаживание эмпирических данных. В данной статье предлагается итерационный алгоритм построения разреженной выпуклой регрессии, т.е. для нахождения выпуклого вектора $z \in \mathbb{R}^n$ с наименьшей квадратичной ошибкой приближения к данному вектору $y \in \mathbb{R}^n$ (не обязательно являющемуся выпуклым). Задача может быть представлена в виде задачи выпуклого программирования с линейными ограничениями. Используя условия оптимальности Каруша—Куна—Таккера, доказано, что оптимальные точки должны лежать на кусочно-линейной функции. Доказано, что предложенный двойственный алгоритм на основе активного

Научная статья

Образец для цитирования

Gudkov A. A., Mironov S. V., Sidorov S. P., Tyshkevich S. V. A dual active set algorithm for optimal sparse convex regression, *Vestn. Samar. Gos. Tekhn. Univ., Ser. Fiz.-Mat. Nauki* [J. Samara State Tech. Univ., Ser. Phys. Math. Sci.], 2019, vol. 23, no. 1, pp. 113–130. doi: 10.14498/vsgtu1673.

Сведения об авторах

Александр Александрович Гудков https://orcid.org/0000-0002-6472-4855
лаборант; институт рисков СГУ; e-mail: alex-good96@mail.ru

Сергей Владимирович Миронов https://orcid.org/0000-0003-3699-5006
кандидат физико-математических наук; доцент; каф. математической кибернетики и компьютерных наук; e-mail: MironovSV@info.sgu.ru

Сергей Петрович Сидоров https://orcid.org/0000-0003-4047-8239
доктор физико-математических наук, доцент; заведующий кафедрой; каф. теории функций и стохастического анализа; e-mail: sidorovsp@yahoo.com

Сергей Викторович Тышкевич https://orcid.org/0000-0001-5417-6165
кандидат физико-математических наук, доцент; доцент; каф. теории функций и стохастического анализа; e-mail: tyszkiewicz@yandex.ru

множества для построения оптимальной разреженной выпуклой регрессии имеет полиномиальную сложность и позволяет найти оптимальное решение (для которого выполнены условия Каруша—Куна—Таккера).

Ключевые слова: двойственный алгоритм, активное множество, изотонная регрессия, монотонная регрессия, выпуклая регрессия.

Получение: 22 января 2019 г. / Исправление: 2 марта 2019 г. /
Принятие: 4 марта 2019 г. / Публикация онлайн: 5 марта 2019 г.

Конкурирующие интересы. Мы заявляем, что у нас нет конфликта интересов в авторстве и публикации этой статьи.

Авторский вклад и ответственность. Все авторы принимали участие в разработке концепции статьи и в написании рукописи. Авторы несут полную ответственность за предоставление окончательной рукописи в печать. Окончательная версия рукописи была одобрена всеми авторами.

Финансирование. Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 18–37–00060).

Благодарность. Авторы благодарны рецензенту за тщательное прочтение статьи и сделанные им конструктивные замечания.