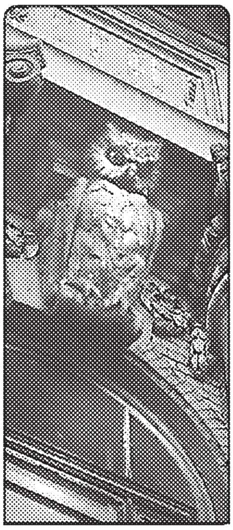

Научный отдел

ИНФОРМАТИКА

Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2022. Т. 22, вып. 2. С. 216–223
Izvestiya of Saratov University. Mathematics. Mechanics. Informatics, 2022, vol. 22, iss. 2, pp. 216–223
https://mmi.sgu.ru
https://doi.org/10.18500/1816-9791-2022-22-2-216-223

Article

Dual active-set algorithm for optimal 3-monotone regression

A. A. Gudkov, S. P. Sidorov✉, K. A. Spiridonov

Saratov State University, 83 Astrakhanskaya St., Saratov 410012, Russia
Alexandr A. Gudkov, alex-good96@mail.ru, AuthorID: 1083215
Sergei P. Sidorov, sidorovsp@sgu.ru, https://orcid.org/0000-0003-4047-8239, AuthorID: 16120
Kirill A. Spiridonov, kir.spiridonov@gmail.com

Abstract. The paper considers a shape-constrained optimization problem of constructing monotone regression which has gained much attention over the recent years. This paper presents the results of constructing the nonlinear regression with 3-monotone constraints. Monotone regression of high orders can be applied in many fields, including non-parametric mathematical statistics and empirical data smoothing. In this paper, an iterative algorithm is proposed for constructing a sparse 3-monotone regression, i.e. for finding a 3-monotone vector with the lowest square error of approximation to a given (not necessarily 3-monotone) vector. The problem can be written as a convex programming problem with linear constraints. It is proved that the proposed dual active-set algorithm has polynomial complexity and obtains the optimal solution.
Keywords: dual algorithm, isotonic regression, monotone regression, k-monotone regression, convex regression
Acknowledgements: This work was supported by the Ministry of science and education of the Russian Federation in the framework of the basic part of the scientific research state task (project FSRR-2020-0006).
For citation: Gudkov A. A., Sidorov S. P., Spiridonov K. A. Dual active-set algorithm for optimal 3-monotone regression. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics*, 2022, vol. 22, iss. 2, pp. 216–223. https://doi.org/10.18500/1816-9791-2022-22-2-216-223

Научная статья
УДК 519.85

Двойственный алгоритм на основе активного множества для построения оптимальной 3-монотонной регрессии

А. А. Гудков, С. П. Сидоров✉, К. А. Спиридонов

Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, Россия, 410012, г. Саратов, ул. Астраханская, д. 83

Гудков Александр Александрович, студент кафедры теории функций и стохастического анализа, alex-good96@ mail.ru, AuthorID: 1083215

Сидоров Сергей Петрович, доктор физико-математических наук заведующий кафедрой теории функций и стохастического анализа, sidorovsp@sgu.ru, https://orcid.org/0000-0003-4047-8239, AuthorID: 16120

Спиридонов Кирилл Александрович, аспирант кафедры теории функций и стохастического анализа, kir.spiridonov@gmail.com

Аннотация. В статье рассматривается задача оптимизации с ограничениями на форму для построения монотонной регрессии, которая в последние годы привлекает большое внимание исследователей. В статье представлены результаты построения нелинейной регрессии с 3-монотонными ограничениями. Монотонная регрессия высоких порядков может применяться во многих областях, включая непараметрическую математическую статистику и сглаживание эмпирических данных. Предлагается итерационный алгоритм для построения разреженной 3-монотонной регрессии, т.е. для нахождения 3-монотонного вектора с наименьшей квадратичной ошибкой приближения к заданному (не обязательно 3-монотонному) вектору. Задачу можно записать как задачу выпуклого программирования с линейными ограничениями. Доказано, что предложенный двойственный алгоритм на основе использования активного множества имеет полиномиальную сложность и дает оптимальное решение.

Ключевые слова: двойственный алгоритм, изотонная регрессия, монотонная регрессия, k-монотонная регрессия, выпуклая регрессия

Благодарности: Работа поддержана Министерством науки и образования Российской Федерации в рамках базовой части государственного задания (проект FSRR-2020-0006).

Для цитирования: *Gudkov A. A., Sidorov S. P., Spiridonov K. A.* Dual active-set algorithm for optimal 3-monotone regression [*Гудков А. А., Сидоров С. П., Спиридонов К. А.* Двойственный алгоритм на основе активного множества для построения оптимальной 3-монотонной регрессии] // Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2022. Т. 22, вып. 2. С. 216–223. https://doi.org/10.18500/1816-9791-2022-22-2-216-223

Introduction

Let $z = (z_1, \dots, z_n)^T \in \mathbb{R}^n$ be the vector of a given function values taken at some points $x = (x_1, \dots, x_n)^T \in \mathbb{R}^n$, $n \in \mathbb{N}$. Denote $\Delta_i = x_{i+1} - x_i$, $i = 1, 2, \dots, n-1$. Then the k-th order finite difference operator Δ^k (for $k \geqslant 1$) is defined recursively as follows:

$$\Delta^k z_i = \frac{1}{\sum\limits_{j=0}^{k-1} \Delta_{i+j}} \left(\Delta^{k-1} z_{i+1} - \Delta^{k-1} z_i\right),$$

where $\Delta^0 z_i = z_i$, $i = 1, \dots, n$.

We will call the vector $z = (z_1, \dots, z_n)^T \in \mathbb{R}^n$ as k-monotone with respect to $x = (x_1, \dots, x_n)^T$, if $\Delta^k z_i \geqslant 0$ for all $i = 1, \dots, n-k$.

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 [1,2]. The most studied has been the problem of constructing monotone (or isotonic) regression, i.e. the task of finding the best fitted non-decreasing vector to a given vector. One can find a detailed review of isotonic regression in the work of Robertson and Dykstra [3, 4].

k-monotone regression is the extension of monotone regression to the general case of k-monotonicity. Both isotonic and k-monotone regression may be applied in many fields, including non-parametric mathematical statistics [1, 5], the empirical data smoothing [6–8], the shape-preserving dynamic programming [9], and the shape-preserving approximation [10, 11]. Moreover, k-monotone sequences and vectors are also used in solving various mathematical problems [12–15].

In this paper, we will use the idea of a dual active-set algorithm that proposes and analyzes regularized monotonic regression in the paper [2]. It should be noted that some algorithms for constructing k-monotone regressions were considered in papers [16, 17].

Denote Δ_3^n the set of all vectors from $\mathbb{R}^n$, which are 3-monotone. The task of constructing 3-monotone 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 3-monotone) under condition $z \in \Delta_3^n$:

$$(z-y)^T(z-y) = \sum_{i=1}^{n} (z_i - y_i)^2 \to \min_{z \in \Delta_3^n(x)} . \tag{1}$$

In this paper we propose a dual active-set algorithm for constructing 3-monotone regression and prove that the algorithm has polynomial complexity and obtains the optimal solution.

1. Preliminary analysis

The problem (1) can be rewritten in the form of a convex programming problem with linear constraints:

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

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

$$\begin{aligned} g_i(z) := -\Bigg(\Delta_{i+1}\Delta_i(\Delta_{i+1}+\Delta_i) z_{i+3} - \Delta_i(\Delta_{i+2}+\Delta_{i+1}) \left(\sum_{j=i}^{i+2} \Delta_j \right) z_{i+2} + \\ + \Delta_{i+2}(\Delta_{i+1}+\Delta_i) \left(\sum_{j=i}^{i+2} \Delta_j \right) z_{i+1} - \Delta_{i+2}\Delta_{i+1}(\Delta_{i+2}+\Delta_{i+1}) z_i \Bigg) \leqslant 0, \end{aligned} \tag{3}$$

for $1 \leqslant i \leqslant n-3$. Problem (2)–(3) is a quadratic programming problem and is strictly convex, therefore there is a unique solution for it.

Let $\hat{z}$ be the global solution of the problem (2)–(3), then there is Lagrange multiplier $\mu = (\mu_1, \dots, \mu_{n-3})^T \in \mathbb{R}^{n-3}$ such that

$$\nabla F(z) + \sum_{i=1}^{n-3} \mu_i \nabla g_i(z) = 0, \tag{4}$$

$$g_i(z) \leqslant 0, \quad 1 \leqslant i \leqslant n-3, \tag{5}$$
$$\mu_i \geqslant 0, \quad 1 \leqslant i \leqslant n-3, \tag{6}$$
$$\mu_i g_i(z) = 0, \quad 1 \leqslant i \leqslant n-3, \tag{7}$$

where ∇g_i is the gradient of the function g_i.

The equations (4)–(7) are the Karush – Kuhn – Tucker conditions. From (4) it follows that

$$\frac{\partial}{\partial z_j}\Bigg[\frac{1}{2}\sum_{i=1}^{n}(z_i-y_i)^2+\sum_{i=1}^{n-3}\mu_i\Big(-\Delta_{i+1}\Delta_i(\Delta_{i+1}+\Delta_i)z_{i+3}+$$
$$+\Delta_i(\Delta_{i+2}+\Delta_{i+1})\left(\sum_{j=i}^{i+2}\Delta_j\right)z_{i+2}-\Delta_{i+2}(\Delta_{i+1}+\Delta_i)\left(\sum_{j=i}^{i+2}\Delta_j\right)z_{i+1}+$$
$$+\Delta_{i+2}\Delta_{i+1}(\Delta_{i+2}+\Delta_{i+1})z_i\Big)\Bigg]=0, \quad 1 \leqslant j \leqslant n-3.$$

2. A dual active-set algorithm for 3-monotone regression

In this subsection, a dual active-set algorithm is proposed. It will be shown that it possesses the following useful properties:

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

The proposed algorithm uses as so-called *active set*. The active set S consists of blocks of the form $[l, r-3] \subset [1, n-3]$, such that $[l, r-3] \subset S, l-1 \notin S, r-2 \notin S$, and

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

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

At each iteration of the algorithm, the active set $S \subset [1, n-3]$ is chosen and the corresponding optimization problem is solved

$$\frac{1}{2}\sum_{i=1}^{n}(z_i-y_i)^2 \to \min, \tag{8}$$

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

$$\Delta_{i+1}\Delta_i(\Delta_{i+1}+\Delta_i)z_{i+3}-\Delta_i(\Delta_{i+2}+\Delta_{i+1})\left(\sum_{j=i}^{i+2}\Delta_j\right)z_{i+2}+$$
$$+\Delta_{i+2}(\Delta_{i+1}+\Delta_i)\left(\sum_{j=i}^{i+2}\Delta_j\right)z_{i+1}-\Delta_{i+2}\Delta_{i+1}(\Delta_{i+2}+\Delta_{i+1})z_i=0, \ \forall i \in S. \tag{9}$$

THE DUAL ACTIVE-SET ALGORITHM FOR 3-MONOTONE REGRESSION

begin
· Input data $y \in \mathbb{R}^n$ · Active set $S = \varnothing$ · Initial approximation $z(S) = y$ **while** $z(S) \notin \Delta_3^n$ **do**
· Change the active set $S \leftarrow S \cup \{i : \ g_i(z(S)) > 0\}$ · We solve the problem (8)–(9) using values from the active set S · Rewriting the vector $z(S)$
end
· Returning the solution $z(S)$
end

The computational complexity of the dual active set algorithm for 3-monotone regression is $O(n^3)$. It follows from two remarks:
– at each iteration of the algorithm, the active set S attaches, at least, one index from $[1, n-3]$, which means that the number of the while loop iterations can not be greater than $n-3$;
– the computational complexity of solving the problem (8)–(9) is $O(n^2)$.

3. The convergence and optimality analysis of the dual active set algorithm

We need the following auxiliary lemmas, the proof of which can be obtained similarly to the proof of the corresponding lemmas in the paper [16].

Lemma 1. *Let z be a global solution to the problem (2)–(3). Then the Lagrange multipliers $\mu = (\mu_1, \dots, \mu_{n-3})^T \in \mathbb{R}^{n-3}$, identified in (4)–(7), are calculated as follows:*

$$\mu_i = -\frac{z_i - y_i}{\Delta_{i+2}\Delta_{i+1}(\Delta_{i+2}+\Delta_{i+1})} + \frac{(\Delta_i + \Delta_{i-1})\left(\sum\limits_{j=i}^{i+2}\Delta_j\right)}{\Delta_{i+2}\Delta_{i+1}(\Delta_{i+2}+\Delta_{i+1})}(\Delta_{i+1}\mu_{i-1} - \Delta_{i-2}\mu_{i-2}) + \\ + \frac{\Delta_{i-2}\Delta_{i-3}(\Delta_{i-2}+\Delta_{i-3})}{\Delta_{i+2}\Delta_{i+1}(\Delta_{i+2}+\Delta_{i+1})}\mu_{i-3}, \qquad 1 \leqslant i \leqslant n-3, \tag{10}$$

and $\mu_i = 0$ $\forall i < 1$.

Lemma 2. *Let $1 \in S$ i.e. $\Delta^2 y_1 < 0$ and suppose that $2, 3, 4 \notin S$. Let z_1, z_2, z_3, z_4 be the values of linear regression, built on pairs of values $(x_1, y_1), (x_2, y_2), (x_3, y_3)$, (x_4, y_4). Then the values of the corresponding Lagrange multipliers (10) will be non-negative.*

Lemma 3. *Let at some iteration of the algorithm the pairs of values $Y = \{(x_1, z_1), \dots, (x_{k+1}, z_{k+1})\}$ such that $[1 : k-3] \subset S$, $\Delta^3 z_i < 0$ for all $i \in [1 : k-3]$, and $k-2, k-1, k, k+1 \notin S$. Let $z_i^{(0)}$, $i \in [1 : k]$ be an optimization problem solution*

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

where the minimum is taken among all values $\zeta \in \mathbb{R}^k$ satisfying $g_i(\zeta) = 0$ for all $i \in [1 : k-3]$, where $g_i(\zeta)$ is defined in (3). Moreover, suppose that $g_{k-2}(z^{(0)}) < 0$ i.e.

x_{k-2} will be added to the active set S at the next iteration of the algorithm. Let $z_i^{(1)}$, $i \in [1 : k+1]$ be an optimization problem solution

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

where the minimum is taken among all values $\zeta \in \mathbb{R}^{k+1}$ satisfying the equality $g_i(\zeta) = 0$ for all $i \in [1 : k-2]$. Then for all $i = 1, 2, \ldots, k-2$ get

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

where $\mu_i^{(1)}$ is Lagrange multiplier for $z^{(1)}$ and $\mu_{k-1}^{(1)} = \mu_k^{(1)} = \mu_{k+1}^{(1)} = 0$.

Theorem 1. *For any initial $S \subset S^*$, the algorithm converges to the optimal solution of the problem (1) in, at most, $n-|S|$ iterations. Where S^* is the active set corresponding to the optimal solution and n is the dimension of the problem.*

Proof. The algorithm is designed in such a way that, at each iteration, the active set S is expanded by attaching at least one index point from $[1, n-3]$. This point should not previously belong to the set S. In the case of $S = [1, n-3]$, the number of blocks is equal to 1 and the input vector is already 3-monotone. Another case is $|S| < n-3$. Then the number of iterations must be less than $n - |S|$ where $|S|$ is the number of indices in the initial active set S.

If the point i has a negative value of the third-order finite difference $\Delta^3 z_i$ and is isolated (i.e. $i-3, i-2, i-1, i+1, i+2, i+3 \notin S$) then the dual active set algorithm replaces $z_i, z_{i+1}, z_{i+2}, z_{i+3}$ with the values of linear regression constructed by the points $(x_i, z_i), (x_{i+1}, z_{i+1}), (x_{i+2}, z_{i+2}), (x_{i+3}, z_{i+3})$. This situation is considered in Lemma 2 in which is proved that the values of the corresponding Lagrange multipliers are non-negative.

Another case we should analyse is the case when the violation of 3-monotonicity occurs at several consecutive $k > 1$ neighboring points, which can be written as follows $\Delta^3 z_j < 0$, $j = i, \ldots, i+k$ and $\Delta^3 z_{i-3}, \Delta^3 z_{i-2}, \Delta^3 z_{i-1} \geqslant 0$, $\Delta^3 z_{i+k+1}, \Delta^3 z_{i+k+2}, \Delta^3 z_{i+k+3} \geqslant 0$. In this case, the algorithm replaces values $z_i, z_{i+1}, \ldots, z_{i+k+3}$ with the values of a linear regression constructed by the points $(x_i, z_i), \ldots, (x_{i+k+3}, z_{i+k+3})$. This situation is considered in Lemma 3 which shows that the values of the corresponding Lagrange multipliers are non-negative.

In the same way, in this theorem, the non-negativity of the Lagrange multipliers can be proved in other cases. □

Conclusion

The paper presents the algorithm for constructing optimal 3-monotone regression based on an active set. This algorithm has already been applied when constructing regression of other orders and with a constant distance between values [16]. At each iteration of the algorithm, it first determines the active set and then solves a standard least-squares subproblem on the active set with a small size, which exhibits a local superlinear convergence. Therefore, the algorithm is very efficient when coupled with parallel execution. The classical optimization algorithms (e.g. coordinate descent or proximal gradient descent) only possess sublinear convergence in general or linear convergence under certain conditions.

References

1. Chen Y. *Aspects of Shape-constrained Estimation in Statistics*. Ph. D. thesis, University of Cambridge, 2013. 143 p.
2. Burdakov O., Sysoev O. A dual active-set algorithm for regularized monotonic regression. *Journal of Optimization Theory and Applications*, 2017, vol. 172, no. 3, pp. 929–949. https://doi.org/10.1007/s10957-017-1060-0
3. Robertson T., Wright F., Dykstra R. *Order Restricted Statistical Inference*. John Wiley & Sons, New York, 1988. 488 p.
4. Dykstra R. An isotonic regression algorithm. *Journal of Statistical Planning and Inference*, 1981, vol. 5, iss. 4, pp. 355–363. https://doi.org/10.1016/0378-3758(81)90036-7
5. Bach F. R. Efficient algorithms for non-convex isotonic regression through submodular optimization. In: S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, R. Garnett, eds. *Advances in Neural Information Processing Systems* 32^{nd}: Annual Conference on Neural Information Processing Systems 2018 (NeurIPS 2018). December 3–8, 2018. Montréal, Canada, 2018, pp. 1–10.
6. Hastie T., Tibshirani R., Wainwright M. *Statistical Learning with Sparsity: The Lasso and Generalizations*. New York, USA, Chapman and Hall/CRC, 2015. 367 p. https://doi.org/10.1201/b18401
7. Altmann D., Grycko E., Hochstättler W., Klützke G. *Monotone smoothing of noisy data*. Diskrete Mathematik und Optimierung. Technical Report feu-dmo034.15. FernUniversität in Hagen, Fakultät für Mathematik und Informatik, 2014. 6 p.
8. Diggle P., Morris S., Morton-Jones T. Case-control isotonic regression for investigation of elevation in risk around a point source. *Statistics in Medicine*, 1999, vol. 18, iss. 13, pp. 1605–1613. https://doi.org/10.1002/(sici)1097-0258(19990715)18:13<1605::aid-sim146>3.0.co;2-v
9. Cai Y., Judd K. L. Chapter 8 – Advances in numerical dynamic programming and new applications. *Handbook of Computational Economics*, 2014, vol. 3, pp. 479–516. https://doi.org/10.1016/b978-0-444-52980-0.00008-6
10. Shevaldin V. T. *Approksimatsiya lokal'nymi splainami* [Approximation by Local Splines]. Ekaterinburg, UMC UPI Publ., 2014. 198 p. (in Russian).
11. Boytsov D. I., Sidorov S. P. Linear approximation method preserving k-monotonicity. *Siberian Electronic Mathematical Reports*, 2015, vol. 12, pp. 21–27.
12. Milovanović I. Z., Milovanović E. I. Some properties of l_p^k-convex sequences. *Bulletin of the International Mathematical Virtual Institute*, 2015, vol. 5, no. 1, pp. 33–36.
13. Niezgoda M. Inequalities for convex sequences and nondecreasing convex functions. *Aequationes Mathematicae*, 2017, vol. 91, no. 1, pp. 1–20. https://doi.org/10.1007/s00010-016-0444-9
14. Latreuch Z., Belaïdi B. New inequalities for convex sequences with applications. *International Journal of Open Problems in Computer Science and Mathematics*, 2012, vol. 5, no. 3, pp. 15–27. https://doi.org/10.12816/0006115
15. Marshall A. W., Olkin I., Arnold B. C. *Inequalities: Theory of Majorization and Its Applications*. New York, USA, Springer, 2011. 909 p. https://doi.org/10.1007/978-0-387-68276-1
16. Gudkov A., Mironov S. V., Sidorov S. P., Tyshkevich S. V. A dual active set algorithm for optimal sparse convex regression. *Journal of Samara State Technical University, Ser. Physical and Mathematical Sciences*, 2019, vol. 23, no. 1, pp. 113–130. https://doi.org/10.14498/vsgtu1673

17. Sidorov S. P., Faizliev A. R., Gudkov A. A., Mironov S. V. Algorithms for sparse k-monotone regression. In: W. J. van Hoeve, ed. *Integration of Constraint Programming, Artificial Intelligence, and Operations Research. CPAIOR 2018*. Lecture Notes in Computer Science, vol. 10848. Springer, Cham, 2018, pp. 546–566. https://doi.org/10.1007/978-3-319-93031-2_39

Поступила в редакцию / Received 03.12.2021
Принята к публикации / Accepted 15.01.2022
Опубликована / Published 31.05.2022