ИНФОРМАТИКА

УДК 519.6

О СХОДИМОСТИ ЖАДНОГО АЛГОРИТМА ДЛЯ РЕШЕНИЯ ЗАДАЧИ ПОСТРОЕНИЯ МОНОТОННОЙ РЕГРЕССИИ

А. А. Гудков, С. В. Миронов, А. Р. Файзлиев

Гудков Александр Александрович, лаборант Института рисков, Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, 410012, Россия, Саратов, Астраханская, 83, alex-good96@mail.ru

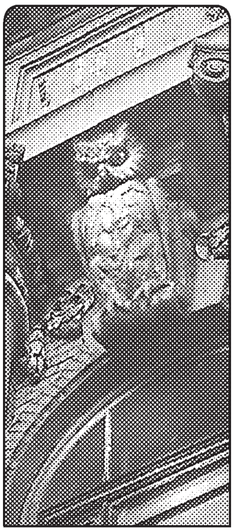
Миронов Сергей Владимирович, кандидат физико-математических наук, заведующий кафедрой математической кибернетики и компьютерных наук, Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, 410012, Россия, Саратов, Астраханская, 83, mironovsv@info.sgu.ru

Файзлиев Алексей Раисович, кандидат экономических наук, Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, 410012, Россия, Саратов, Астраханская, 83, faizlievar1983@mail.ru

В статье представлены жадные алгоритмы, которые используют подход типа Франка – Вульфа для нахождения разреженной монотонной регрессии. Проблема нахождения монотонной регрессии возникает при сглаживании эмпирических данных, в задачах динамического программирования, математической статистике и во многих других прикладных задачах. Для решения данной задачи требуется найти неубывающую последовательность точек, имеющую наименьшую ошибку приближения к заданному множеству точек на плоскости. Задача построения монотонной регрессии может быть сформулирована в виде задачи выпуклого программирования с линейными ограничениями и имеет неполиномиальную сложность. В статье также находятся оценки скорости сходимости представленных жадных алгоритмов.

Ключевые слова: жадные алгоритмы, алгоритм Франка – Вульфа, монотонная регрессия.

DOI: 10.18500/1816-9791-2017-17-4-431-440



НАУЧНЫЙ ОТДЕЛ

ВВЕДЕНИЕ

Начиная с 1980-х гг. наблюдается повышенный интерес к оценкам с ограничениями в математической статистике [1]. Одной из таких задач является задача построения монотонной регрессии, которая состоит в том, чтобы найти неубывающую последовательность точек, имеющую наименьшую ошибку приближения к заданному множеству точек на плоскости. Обзор результатов по задаче построения монотонной регрессии можно найти в книге Т. Робертсона (Robertson), Ф. Райта (Wright), и Р. Декстры (Dykstra) [2]. В статьях [3–6] рассматривается проблема нахождения монотонной регрессии в рамках квадратичного и выпуклого программирования.

Используя подход на основе математического программирования, авторы статей [7–9] недавно предложили новые результаты по этой проблеме. В статьях [10, 11] рассматривалась задача такого типа для случая частичных порядков, определенных переменными множественной регрессии.

Монотонная регрессия используется в различных областях:

- в математической статистике, например, при непараметрическом восстановлении функций распределения или функций плотности [1, 12];
- при сглаживании эмпирических данных, например в статье [13], используется монотонная регрессия как метод анализа данных при подсчете оценки риска заболевания по отношению к точечному источнику загрязнения окружающей среды;
- в задачах динамического программирования; так, в статье [14] представлены алгоритмы решения задач динамического программирования на основе методов формосохраняющей аппроксимации и показана применимость алгоритмов, сохраняющих конус, для решения задач оптимального роста.

При построении монотонной регрессии мы предполагаем, что существует (неубывающая) зависимость между сигналом x и ответом $y(x)$.

Пусть $\mathbb{B}[0,1]$ есть пространство ограниченных функций с нормой $\|f\| = \max_{x\in[0,1]}|f(x)|$. Функция $z \in \mathbb{B}[0,1]$ называется монотонной, если для каждой пары точек $x_2 > x_1$ имеет место неравенство

$$z(x_2) - z(x_1) \geqslant 0.$$

Обозначим через Δ_1 множество всех монотонных функций из $\mathbb{B}[0,1]$.

Задача построения монотонной регрессии может быть сформулирована в виде задачи выпуклой оптимизации следующим образом: необходимо найти функцию $z \in \mathbb{B}[0,1]$ с наименьшей ошибкой приближения функции $y \in \mathbb{B}[0,1]$ в норме $L_q[0,1]$, $q \in (1,2]$, при условии монотонности z:

$$E(z) = \int_0^1 (z(x) - y(x))^q dx \to \min_{z\in\Delta_1}. \tag{1}$$

В этой статье рассматриваются жадные алгоритмы, использующие подход типа Франка – Вульфа для нахождения разреженной монотонной регрессии, и находятся оценки скорости сходимости этих алгоритмов.

1. СЛАБЫЙ ЖАДНЫЙ АЛГОРИТМ В $L_1[0,1]$

Множество элементов $\mathscr{D}$ пространства $\mathbb{B}[0,1]$ называется *словарем*, если каждый элемент $g \in \mathscr{D}$ ограничен по норме единицей, $\|g\| \leqslant 1$, и замыкание оболочки $\mathscr{D}$ есть $\mathbb{B}[0,1]$, т. е. $\overline{\operatorname{span}\mathscr{D}} = \mathbb{B}[0,1]$.

Обозначим:

$$\theta_a(x) = \begin{cases} 0, & x < a, \\ 1, & x \geqslant a. \end{cases}$$

Тогда множество $\{\theta_a :\ a \in (0,1)\}$ является словарем в $\mathbb{B}[0,1]$, будем в дальнейшем обозначать его $\mathscr{D}$.

Таким образом, решение задачи ищется в виде $z = \sum\limits_{i=1}^{n} b_i \theta_{a_i}$, где $a_1 < a_2 < \ldots < a_n$ и $b_i \geqslant 0$, $i = 1, \ldots, n$, $n \in \mathbb{N}$, и

$$z(1) - z(0) = \sum_{i=1}^{n} b_i = \sup_{x \in [0,1]} y(x) - \inf_{x \in [0,1]} y(x).$$

Обозначим через $S(\mathscr{D})$ замыкание в $\mathbb{B}[0,1]$ всех таких z.

Одним из вариантов конструктивных методов поиска наилучших m-членных приближений являются жадные алгоритмы. Жадные алгоритмы позволяют получить разреженные решения по отношению к словарю $\mathscr{D}$. Возможно, метод Франка – Вульфа [15], который также известен как метод условного градиента [16], является одним из наиболее известных алгоритмов для нахождения оптимальных решений задач условной выпуклой оптимизации. Важный вклад в разработку алгоритмов типа алгоритма Франка – Вульфа можно найти в [17–19]. В статье [19] приведены результаты двойственной сходимости для алгоритмов типа Франка – Вульфа на основе развития концепции двойственности, представленной в работе [17]. Недавние результаты по сходимости жадных алгоритмов можно найти в работах [20–29].

В настоящей работе рассматривается жадный алгоритм для поиска решений задачи выпуклой оптимизации (1), которые являются разреженными относительно некоторого словаря, в пространстве $\mathbb{B}[0,1]$. Полученные в этом параграфе результаты по оценке скорости сходимости основаны на использовании результатов статей [28, 30].

Заметим, что функция E, определенная в (1), является дифференцируемой по Фреше. Обозначим через $\langle E'(x), y\rangle$ значение функционала $E'(x)$ (производной по Фреше функции E в точке x) в точке y. Для решения задачи (1) мы будем использовать алгоритм 1. Алгоритм для каждого $m \geqslant 1$ находит следующий элемент G_m по индукции с использованием текущего элемента G_{m-1} и элемента словаря θ_{a_m}, полученного на шаге жадного спуска.

Алгоритм 1: Слабый жадный алгоритм

начало алгоритма
- $\cdot$ Пусть $G_0 = 0$;
- **цикл для каждого** $m = 1, 2, \ldots, M$
  - $\cdot$ (*Шаг жадного спуска*) Найти точку $a_m \in (0,1)$ (т. е. функцию $\theta_{a_m} \in \mathscr{D}$) такую, что $\langle -E'(G_{m-1}), \theta_{a_m}\rangle \geqslant t_m \sup\limits_{s \in \mathscr{D}} \langle -E'(G_{m-1}), s\rangle$;
  - $\cdot$ (*Линейный поиск*) Найти число $0 \leqslant \lambda_m \leqslant 1$, такое, что $E\left((1-\lambda_m)G_{m-1} + \lambda_m \theta_{a_m}\right) = \inf\limits_{0 \leqslant \lambda \leqslant 1} E\left((1-\lambda)G_{m-1} + \lambda \theta_{a_m}\right)$;
  - $\cdot$ (*Переход к следующей точке*) $G_m = (1-\lambda_m)G_{m-1} + \lambda_m \theta_{a_m}$;

конец алгоритма

На шаге жадного спуска мы максимизируем функционал, который использует градиентную информацию в точке G_{m-1}, полученной на предыдущей итерации алгоритма. Алгоритм 1 принадлежит классу методов типа метода Франка – Вульфа, так как в каждом текущем элементе G_{m-1} он движется в направлении минимума линеаризованной целевой функции E на допустимом множестве, которым в данном случае является словарь $\mathscr{D}$.

На шаге жадного спуска мы ищем супремум по словарю $\mathscr{D}$ (а не множеству $S(\mathscr{D})$), так как элементы $S(\mathscr{D})$ являются в основном линейными комбинациями бесконечного числа элементов словаря. Таким образом, полученное оптимальное решение не обязательно будет разреженным по отношению к словарю $\mathscr{D}$.

Найти точное решение подзадачи $\sup_{s\in\mathscr{D}}\langle -E'(G_{m-1}), s\rangle$ практически невозможно за конечное время для большинства функций y. Поэтому алгоритм 1 использует ослабляющую последовательность τ на шаге жадного спуска для нахождения приближенного решения, которое имеет качество приближения, по меньшей мере t_m, на шаге m. По этой причине алгоритм называется «слабым».

Шаг линейного поиска алгоритма 1 находит наилучший элемент на отрезке, соединяющем текущую точку G_{m-1} и элемент θ_{a_m}.

Положим $\Omega := \{f \in \mathbb{B}[0,1] :\ E(x) \leqslant E(0)\}$, и заметим, что Ω ограничено. Анализ сходимости жадных алгоритмов существенно использует меру «нелинейности» целевой функции E на Ω, которая может быть оценена с помощью модуля гладкости функции E.

Напомним, что модуль гладкости ρ функции E на ограниченном множестве Ω определяется следующим образом:

$$\rho(E,u) = \frac{1}{2}\sup_{f\in\Omega,\|g\|=1}|E(f+ug)+E(f-ug)-2E(f)|, \qquad u>0. \tag{2}$$

E называется равномерно гладкой на Ω, если $\lim_{u\to 0}\rho(E,u)/u = 0$.

Следующая лемма легко может быть установлена.

Лемма 1. *Функция E, определенная в* (1)*, является равномерно гладкой функцией с модулем гладкости, удовлетворяющем неравенству $\rho(E,u) \leqslant \gamma u^q$, $1 < q \leqslant 2$, $\gamma > 0$.*

Используя геометрические свойства функции E и результаты статьи [30], мы можем доказать следующую оценку скорости сходимости жадного алгоритма 1.

Теорема 1. *Пусть $\tau = \{t_k\}_{k=1}^{\infty}$, $0 < t_k \leqslant 1$, $k = 1,2,\ldots$, есть ослабляющая последовательность. Тогда*

$$E(G_m) - E(z^*) \leqslant \left(1 + C_1(q,\gamma)\sum_{k=1}^{m} t_k^p\right)^{1-q}, \qquad p := \frac{q}{q-1}, \qquad m \geqslant 2, \tag{3}$$

где $z^* \in S(\mathscr{D})$ есть оптимальное решение и $C_1(q,\gamma)$ есть положительное число, не зависящее от m.

2. ЖАДНЫЙ АЛГОРИТМ ДЛЯ РЕШЕНИЯ ЗАДАЧИ ПОСТРОЕНИЯ МОНОТОННОЙ РЕГРЕССИИ ДЛЯ ДИСКРЕТНОГО НАБОРА ТОЧЕК

Рассмотрим задачу построения монотонной регрессии для дискретного набора точек: сигналу $x = (x_1,\ldots,x_n)$ соответствуют значения $y = (y_1,\ldots,y_n)$. Полагаем, что $x_{i+1} - x_i \neq \mathrm{const}$, $x_i < x_{i+1}$, $i = 1,\ldots,n-1$.

Последовательность $z = (z_1, \dots, z_n) \in \mathbb{R}^n$ называется монотонной, если $z_i - z_{i-1} \geqslant 0$, $i = 2, \dots, n$.

Обозначим через Δ_1^n множество всех монотонных последовательностей в $\mathbb{R}^n$.

Тогда задача построения монотонной регрессии может быть записана как задача условной выпуклой оптимизации: необходимо найти вектор $z \in \mathbb{R}^n$ с наименьшим значением квадратичной ошибки приближения данного вектора $y \in \mathbb{R}^n$ при условии монотонности z:

$$f(z) = \frac{1}{n}\sum_{i=1}^{n}(z_i - y_i)^2 \to \min_{z \in \Delta_1^n}. \tag{4}$$

Простой итеративный алгоритм для решения задачи (4) предложен в [31, 32] и называется PAVA (Pool-Adjacent-Violators Algorithm). В работе [5] рассматривается обобщение этого алгоритма.

Для построения жадного алгоритма для решения задачи (4) мы перейдем от точек z_i к их скачкам $\zeta_i = z_{i+1} - z_i$, $i = 1, \dots, n-1$, $\zeta_0 = z_1$. Тогда задача (4) может быть переписана в следующем виде:

$$g(\zeta) := \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=0}^{i-1}\zeta_j - y_i\right)^2 \to \min_{\zeta \in S}, \tag{5}$$

где S есть множество всех $\zeta = (\zeta_0, \zeta_1, \dots, \zeta_{n-1}) \in \mathbb{R}^n$ таких, что $\min_i y_i \leqslant \zeta_0 \leqslant \max_i y_i$, $(\zeta_1, \dots, \zeta_{n-1}) \in \mathbb{R}_+^{n-1}$ и $\sum\limits_{j=0}^{n-1}\zeta_j \leqslant \max_i y_i$.

Пусть $\nabla g(\zeta)$ означает градиент g в точке ζ,

$$\nabla g(\zeta) = \left(\frac{\partial g}{\partial \zeta_0}, \frac{\partial g}{\partial \zeta_1}, \dots, \frac{\partial g}{\partial \zeta_{n-1}}\right),$$

где

$$\frac{\partial g}{\partial \zeta_k} = \frac{2}{n}\sum_{i=k}^{n}\Big(\sum_{j=0}^{i-1}\zeta_j - y_i\Big), \qquad k = 0, \dots, n-1.$$

Следует отметить, что для задач большой размерности задачи (5) классическими алгоритмами градиентного спуска решение может оказаться сложным с вычислительной точки зрения. В этой связи в настоящем исследовании предлагается использовать жадный алгоритм типа алгоритма Франка – Вульфа для решения задачи (5). Псевдокод жадного алгоритма приведен ниже (алгоритм 2).

Оценка скорости сходимости алгоритма 2 приведена в следующей теореме.

Теорема 2. *Пусть последовательность $\{\zeta^t\}$ получена в соответствии с алгоритмом 2. Тогда для любого $t \geqslant 2$*

$$g(\zeta^t) - g(\zeta^*) \leqslant 4\sqrt{\frac{n(n+1)(2n+1)}{6n^2}}\,\frac{(\max_i y_i - \min_i y_i)^2}{t+2}, \tag{6}$$

где ζ^* есть оптимальное решение задачи (5).

Доказательство. Известно [15], что для всех $t \geqslant 2$ имеет место неравенство

$$g(\zeta^t) - g(\zeta^*) \leqslant \frac{2L(Diam(S))^2}{t+2},$$

Алгоритм 2: Жадный алгоритм для решения задачи (5)

начало алгоритма

- Зададим число итераций N;
- Начальная точка $\zeta^0 = (\zeta_0^0, 0, \dots, 0)$, где $\zeta_0^0 = \frac{1}{n}\sum\limits_{i=1}^{n} y_i$;
- Счетчик числа итераций $t = 0$;
- **цикл пока** $t < N$ **выполнять**
  - Вычислить $\nabla g(\zeta^t)$, градиент функции g в точке ζ^t;
  - Пусть $\widetilde{\zeta}^t$ есть решение линейной оптимизационной задачи $\nabla g(\zeta^t)^T \cdot \zeta \to \min\limits_{\zeta \in S}$, где $\nabla g(\zeta^t)^T \cdot \zeta$ есть скалярное произведение векторов;
  - Положим $\zeta^{t+1} = \zeta^t + \alpha_t(\widetilde{\zeta}^t - \zeta^t)$, $\alpha_t = \frac{2}{t+2}$ и $t := t+1$;
- Восстановить $z = (z_1, \dots, z_n)$ на основе вектора скачков ζ^N;

конец алгоритма

где L есть константа Липшица и $Diam(S)$ есть диаметр множества S.

Пусть

$$\nabla^2 g(\zeta) := \left(\frac{\partial^2 g}{\partial \zeta_0^2}, \frac{\partial^2 g}{\partial \zeta_1^2}, \dots, \frac{\partial^2 g}{\partial \zeta_{n-1}^2}\right).$$

Известно, что если ∇g является дифференцируемой, то ее константа Липшица L удовлетворяет неравенству

$$L \leqslant \sup_{\zeta} \|\nabla^2 g(\zeta)\|_2.$$

Тогда

$$L \leqslant \sup_{\zeta} \sqrt{\sum_{k=0}^{n-1}\left(\frac{\partial^2 g}{\partial \zeta_k^2}\right)^2} = \frac{1}{n}\sqrt{\sum_{k=1}^{n}(2(n-k+1))^2} = \frac{2}{n}\sqrt{\sum_{k=1}^{n} k^2} = 2\sqrt{\frac{n(n+1)(2n+1)}{6n^2}}.$$

Легко видеть, что $Diam(S) := \sqrt{2}(\max_i y_i - \min_i y_i)$. □

Недостатком этого метода является зависимость теоретической степени сходимости от размерности задачи.

Благодарности. Работа выполнена при финансовой поддержке РФФИ (проекты № 16-01-00507, 18-01-00408).

Библиографический список

1. *Chen Y.* Aspects of Shape-constrained Estimation in Statistics : PhD Thesis. University of Cambridge, 2013. 143 p.
2. *Robertson T., Wright F., Dykstra R.* Order Restricted Statistical Inference. N. Y. : John Wiley & Sons, 1988. 544 p.
3. *Barlow R., Brunk H.* The Isotonic Regression Problem and Its Dual // Jurnal of the American Statistical Association. 1972. Vol. 67, № 1. P. 140–147. DOI: 10.2307/2284712.
4. *Dykstra R.* An Isotonic Regression Algorithm // Journal of Statistical Planning and Inference. 1981. Vol. 5, № 1. P. 355–363. DOI: 10.1214/aos/1176345866.

5. *Best M. J., Chakravarti N.* Active set algorithms for isotonic regression: a unifying framework // Math. Program. 1990. Vol. 47, № 3. P. 425–439. DOI: 10.1007/BF01580873.
6. *Best M., Chakravarti N., Ubhaya V.* Minimizing Separable Convex Functions Subject to Simple Chain Constraints // SIAM Journal on Optimization. 2000. Vol. 10, № 3. P. 658–672. DOI: 10.1137/S1052623497314970.
7. *Stromberg U.* An Algorithm for Isotonic Regression with Arbitrary Convex Distance Function // Computational Statistics & Data Analysis. 1991. Vol. 11, № 1. P. 205–219.
8. *Ahuja R., Orlin J.* A Fast Scaling Algorithm for Minimizing Separable Convex Functions Subject to Chain Constraints // Operations Research. 2001. Vol. 49, № 1. P. 784–789. DOI: 10.1287/opre.49.5.784.10601
9. *Hansohm J.* Algorithms and Error Estimations for Monotone Regression on Partially Preordered Sets // Journal of Multivariate Analysis. 2007. Vol. 98, № 1. P. 1043–1050. DOI: 10.1016/j.jmva.2006.11.001.
10. *Dykstra R., Robertson T.* An Algorithm for Isotonic Regression for Two or More Independent Variables // Ann. Statist. 1982. Vol. 10, № 1. P. 708–719. DOI: 10.1214/aos/1176345866.
11. *Burdakov O., Grimvall A., Hussian M.* A Generalised PAV Algorithm for Monotonic Regression in Several Variables // COMPSTAT: Proc. 16th Symposium in Computational Statistics. Prague, Czech Republic, 2004. P. 761–767.
12. *Bach F.* Efficient Algorithms for Non-convex Isotonic Regression through Submodular Optimization // Online Journal CoRR. 2017. Vol. abs/1707.09157. URL: http://arxiv.org/abs/1707.09157 (accessed 10, September, 2017).
13. *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, № 13. P. 1605–1613. DOI: 10.1002/(SICI)1097-0258(19990715)18:13<1605::AID-SIM146>3.0.CO;2-V.
14. *Cai Y., Judd K. L.* Shape-preserving dynamic programming // Math. Meth. Oper. Res. 2013. Vol. 77. P. 407–421. DOI: 10.1007/s00186-012-0406-5.
15. *Frank M., Wolfe Ph.* An algorithm for quadratic programming // Naval Research Logistics Quarterly. 1956. Vol. 3, № 1–2. P. 95–110. DOI: 10.1002/nav.3800030109.
16. *Levitin E. S., Polyak B. T.* Constrained minimization methods // USSR Comp. Math. & M. Phys. 1966. Vol. 6, № 5. P. 1–50.
17. *Clarkson K. L.* Coresets, sparse greedy approximation, and the Frank–Wolfe algorithm // ACM Transactions on Algorithms. 2010. Vol. 6, № 4. P. 1–30. DOI: 10.1145/1824777.1824783.
18. *Freund R. M., Grigas P.* New analysis and results for the Frank-Wolfe method // Math. Program. 2016. Vol. 155, № 1–2. P. 199–230. DOI: 10.1007/s10107-014-0841-6.
19. *Jaggi M.* Revisiting Frank – Wolfe: Projection-free sparse convex optimization // ICML'13 : Proc. 30th International Conference on Machine Learning. Atlanta, GA, USA, 2013. P. 427–435.
20. *Friedman J.* Greedy Function Approximation: A Gradient Boosting Machine // Ann. Statist. 2001. Vol. 29, № 5. P. 1189–1232. DOI: 10.1214/aos/1013203451.
21. *Davis G., Mallat S., Avellaneda M.* Adaptive greedy approximation // Constr. Approx. 1997. Vol. 13. P. 57–98. DOI: 10.1007/BF02678430.
22. *Jones L.* On a conjecture of Huber concerning the convergence of projection pursuit regression // Ann. Statist. 1987. Vol. 15, № 2. P. 880–882. DOI: 10.1214/aos/1176350382.
23. *Barron A. R., Cohen A., Dahmen W., DeVore R. A.* Approximation and Learning by Greedy Algorithms // Ann. Statist. 2008. Vol. 36, № 1. P. 64–94. DOI: 10.1214/009053607000000631.

24. *DeVore R. A., Temlyakov V. N.* Some remarks on greedy algorithms // Advances in Computational Mathematics. 1996. Vol. 5. P. 173–187. DOI: 10.1007/BF02124742.
25. *Konyagin S. V., Temlyakov V. N.* A remark on greedy approximation in Banach spaces // East J. Approx. 1999. Vol. 5, № 3. P. 365–379.
26. *Nguyen H., Petrova G.* Greedy Strategies for Convex Optimization // Calcolo. 2017. Vol. 54, № 1. P. 207–224. DOI: 10.1007/s10092-016-0183-2.
27. *Temlyakov V. N.* Dictionary descent in optimization // Analysis Mathematica. 2016. Vol. 42, № 1. P. 69–89. DOI: 10.1007/s10476-016-0106-0.
28. *DeVore R. A., Temlyakov V. N.* Convex optimization on Banach spaces // Found. Comput. Math. 2016. Vol. 16, № 2. P. 369–394. DOI: 10.1007/s10208-015-9248-x.
29. *Sidorov S., Mironov S., Pleshakov M.* Dual Greedy Algorithm for Conic Optimization Problem // CEUR-WS. 2016. Vol. 1623. P. 276–283.
30. *Temlyakov V. N.* Greedy approximation in convex optimization // Constr. Approx. 2015. Vol. 41, № 2. P. 269–296. DOI: 10.1007/s00365-014-9272-0.
31. *De Leeuw J., Hornik K., Mair P.* Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods // Journal of Statistical Software. 2009. Vol. 32, № 5. P. 1–24. DOI: 10.18637/jss.v032.i05.
32. *Chepoi V., Cogneau D., Fichet B.* Polynomial algorithms for isotonic regression // Lecture Notes–Monograph Series. 1997. Vol. 31. P. 147–160. DOI: 10.1214/lnms/1215454134

Образец для цитирования:
Гудков А. А., Миронов С. В., Файзлиев А. Р. О сходимости жадного алгоритма для решения задачи построения монотонной регрессии // Изв. Сарат. ун-та. Нов. сер. Сер. Математика. Механика. Информатика. 2017. Т. 17, вып. 4. С. 431–440. DOI: 10.18500/1816-9791-2017-17-4-431-440.

On the Convergence of a Greedy Algorithm for the Solution of the Problem for the Construction of Monotone Regression

A. A. Gudkov, S. V. Mironov, A. R. Faizliev

Alexander A. Gudkov, orcid.org/0000-0002-6472-4855, Saratov State University, 83, Astrakhanskaya Str., Saratov, Russia, 410012, alex-good96@mail.ru
Sergei V. Mironov, orcid.org/0000-0003-3699-5006, Saratov State University, 83, Astrakhanskaya Str., Saratov, Russia, 410012, MironovSV@info.sgu.ru
Alexey R. Faizliev, orcid.org/0000-0001-6442-4361, Saratov State University, 83, Astrakhanskaya Str., Saratov, Russia, 410012, faizlievar1983@mail.ru

The paper presents greedy algorithms that use the Frank-Woolf-type approach for finding a sparse monotonic regression. The problem of finding monotonic regression arises in smoothing an empirical data, in problems of dynamic programming, mathematical statistics and in many other applied problems. The problem is to find a non-decreasing sequence of points with the lowest error of approximation to the given set of points on the plane. The problem of constructing monotonic regression can be formulated as a convex programming problem with linear constraints and is NP-hard problem. The paper also contains estimates of the rate of convergence for the presented greedy algorithms.

Key words: greedy algorithms, Frank – Wolfe algorithm, monotone regression.

Acknowledgements: This work was supported by the Russian Foundation for Basic Research (projects nos. 16-01-00507, 18-01-00408).

References

1. Chen Y. *Aspects of Shape-constrained Estimation in Statistics*: PhD Thesis. University of Cambridge, 2013. 143 p.
2. Robertson T., Wright F., Dykstra R. *Order Restricted Statistical Inference*. New York, John Wiley & Sons, 1988. 544 p.
3. Barlow R., Brunk H. The Isotonic Regression Problem and Its Dual. *Journal of the American Statistical Association*, 1972, vol. 67, no. 1, pp. 140–147. DOI: 10.2307/2284712.
4. Dykstra R. An Isotonic Regression Algorithm. *Journal of Statistical Planning and Inference*, 1981, vol. 5, no. 1, pp. 355–363. DOI: 10.1214/aos/1176345866.
5. Best M. J., Chakravarti N. Active set algorithms for isotonic regression: a unifying framework. *Math. Program.*, 1990, vol. 47, no. 3, pp. 425–439. DOI: 10.1007/BF01580873.
6. Best M., Chakravarti N., Ubhaya V. Minimizing Separable Convex Functions Subject to Simple Chain Constraints. *SIAM Journal on Optimization*, 2000, vol. 10, no. 3, pp. 658–672. DOI: 10.1137/S1052623497314970.
7. Stromberg U. An Algorithm for Isotonic Regression with Arbitrary Convex Distance Function. *Computational Statistics & Data Analysis*, 1991, vol. 11, no. 1, pp. 205–219.
8. Ahuja R., Orlin J. 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
9. Hansohm J. Algorithms and Error Estimations for Monotone Regression on Partially Preordered Sets. *Journal of Multivariate Analysis*, 2007, vol. 98, no. 1, pp. 1043–1050. DOI: 10.1016/j.jmva.2006.11.001.
10. Dykstra R., Robertson T. An Algorithm for Isotonic Regression for Two or More Independent Variables. *Ann. Statist.*, 1982, vol. 10, no. 1, pp. 708–719. DOI: 10.1214/aos/1176345866.
11. Burdakov O., Grimvall A., Hussian M. A Generalised PAV Algorithm for Monotonic Regression in Several Variables. *COMPSTAT: Proc. 16th Symposium in Computational Statistics*. Prague, Czech Republic, 2004, pp. 761–767.
12. Bach F. Efficient Algorithms for Non-convex Isotonic Regression through Submodular Optimization. *Online journal CoRR*, 2017, vol. abs/1707.09157. Available at: http://arxiv.org/abs/1707.09157 (accessed 10 September, 2017).
13. 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, no. 13, pp. 1605–1613. DOI: 10.1002/(SICI)1097-0258(19990715)18:13<1605::AID-SIM146>3.0.CO;2-V.
14. Cai Y., Judd K. L. Shape-preserving dynamic programming. *Math. Meth. Oper. Res.*, 2013, vol. 77, pp. 407–421. DOI: 10.1007/s00186-012-0406-5.
15. Frank M., Wolfe Ph. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 1956, vol. 3, no. 1–2, pp. 95–110. DOI: 10.1002/nav.3800030109.
16. Levitin E. S., Polyak B. T. Constrained minimization methods. *USSR Comp. Math. & M. Phys.*, 1966, vol. 6, no. 5, pp. 1–50.
17. Clarkson K. L. Coresets, sparse greedy approximation, and the Frank – Wolfe algorithm. *ACM Transactions on Algorithms*, 2010, vol. 6, no. 4, pp. 1–30. DOI: 10.1145/1824777.1824783.
18. Freund R. M., Grigas P. New analysis and results for the Frank-Wolfe method. *Math. Program.*, 2016, vol. 155, no. 1–2, pp. 199–230. DOI: 10.1007/s10107-014-0841-6.
19. Jaggi M. Revisiting Frank–Wolfe: Projection-free sparse convex optimization. *ICML'13: Proc. 30th International Conference on Machine Learning*. Atlanta, GA, USA, 2013, pp. 427–435.

20. Friedman J. Greedy Function Approximation: A Gradient Boosting Machine. *Ann. Statist.*, 2001, vol. 29, no. 5, pp. 1189–1232. DOI: 10.1214/aos/1013203451.
21. Davis G., Mallat S., Avellaneda M. Adaptive greedy approximation. *Constr. Approx.*, 1997, vol. 13, pp. 57–98. DOI: 10.1007/BF02678430.
22. Jones L. On a conjecture of Huber concerning the convergence of projection pursuit regression. *Ann. Statist.*, 1987, vol. 15, no. 2, pp. 880–882. DOI: 10.1214/aos/1176350382.
23. Barron A. R., Cohen A., Dahmen W., DeVore R. A. Approximation and Learning by Greedy Algorithms. *Ann. Statist.*, 2008, vol. 36, no. 1, pp. 64–94. DOI: 10.1214/009053607 000000631.
24. DeVore R. A., Temlyakov V. N. Some remarks on greedy algorithms. *Advances in Computational Mathematics*, 1996, vol. 5, pp. 173–187. DOI: 10.1007/BF02124742.
25. Konyagin S. V., Temlyakov V. N. A remark on greedy approximation in Banach spaces. *East J. Approx.*, 1999, vol. 5, no. 3, pp. 365–379.
26. Nguyen H., Petrova G. Greedy Strategies for Convex Optimization. *Calcolo*, 2017, vol. 54, no. 1, pp. 207–224. DOI: 10.1007/s10092-016-0183-2.
27. Temlyakov V. N. Dictionary descent in optimization. *Analysis Mathematica*, 2016, vol. 42, no. 1, pp. 69–89. DOI: 10.1007/s10476-016-0106-0.
28. DeVore R. A., Temlyakov V. N. Convex optimization on Banach spaces. *Found. Comput. Math.*, 2016, vol. 16, no. 2, pp. 369–394. DOI: 10.1007/s10208-015-9248-x.
29. Sidorov S., Mironov S., Pleshakov M. Dual Greedy Algorithm for Conic Optimization Problem. *CEUR-WS*, 2016, vol. 1623, pp. 276–283.
30. Temlyakov V. N. Greedy approximation in convex optimization. *Constr. Approx.*, 2015, vol. 41, no. 2, pp. 269–296. DOI: 10.1007/s00365-014-9272-0.
31. De Leeuw J., Hornik K., Mair P. Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods. *Journal of Statistical Software*, 2009, vol. 32, no. 5, pp. 1–24. DOI: 10.18637/jss.v032.i05.
32. Chepoi V., Cogneau D., Fichet B. Polynomial algorithms for isotonic regression. *Lecture Notes–Monograph Series*, 1997, vol. 31, pp. 147–160. DOI: 10.1214/lnms/1215454134.

Cite this article as:
Gudkov A. A., Mironov S. V., Faizliev A. R. On the Convergence of a Greedy Algorithm for the Solution of the Problem for the Construction of Monotone Regression. *Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform.*, 2017, vol. 17, iss. 4, pp. 431–440 (in Russian). DOI: 10.18500/1816-9791-2017-17-4-431-440.