Math-Net.Ru

Общероссийский математический портал

А. В. Деревенцов, Сравнение скорости сходимости чисто жадного и ортогонального жадного алгоритмов, *Матем. заметки*, 2012, том 92, выпуск 4, 528–532

DOI: <https://doi.org/10.4213/mzm9091>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением <http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.162.27.143

26 апреля 2023 г., 09:49:14

Сравнение скорости сходимости чисто жадного и ортогонального жадного алгоритмов

А. В. Деревенцов

В работе рассматриваются два вида жадных алгоритмов: чисто жадный алгоритм (PGA) и ортогональный жадный алгоритм (OGA). С точки зрения оценок скорости сходимости на всем классе $\mathcal{A}_1(\mathcal{D})$ ортогональный жадный алгоритм оптимален и существенно превосходит чисто жадный алгоритм. Основным результатом настоящей работы является утверждение, показывающее, что для отдельных элементов класса $\mathcal{A}_1(\mathcal{D})$ (и даже $\mathcal{A}_0(\mathcal{D})$) ситуация может быть и противоположной: скорость сходимости ортогонального жадного алгоритма может быть существенно ниже скорости сходимости чисто жадного алгоритма.

Библиография: 16 названий.

Жадные алгоритмы изучались в статистике (см., например, [1], [2]) и в теории обработки сигналов (см., например, [3]) с 1980-х гг. Активное изучение общей теории жадных алгоритмов было начато в 1990-е гг. (см., например, [4]–[7]).

В данной работе рассматриваются два вида жадных алгоритмов: чисто жадный алгоритм (PGA) и ортогональный жадный алгоритм (OGA) (см., например, [4]).

Напомним основные определения и некоторые результаты, связанные с этими видами жадных алгоритмов.

Рассмотрим действительное гильбертово пространство H со скалярным произведением $\langle \cdot, \cdot \rangle$. Множество $\mathcal{D} \subset H$ будем называть *словарем*, если норма всех элементов $\mathcal{D}$ равна 1 и $\text{span } \mathcal{D} = H$.

Для заданного словаря $\mathcal{D}$ и функции $f \in H$ чисто жадный и ортогональный жадный алгоритмы определяются следующим образом.

Чисто жадный алгоритм (PGA). Положим

$$f_0^{\text{PGA}, \mathcal{D}} = f, \quad G_0^{\text{PGA}}(f, \mathcal{D}) = 0.$$

Если уже определен $f_m^{\text{PGA}, \mathcal{D}}$, найдем элемент $g_{m+1} \in \mathcal{D}$, удовлетворяющий условию

$$|\langle f_m^{\text{PGA}, \mathcal{D}}, g_{m+1} \rangle| = \sup_{g \in \mathcal{D}} |\langle f_m^{\text{PGA}, \mathcal{D}}, g \rangle|,$$

и положим

$$\begin{aligned} G_{m+1}^{\text{PGA}}(f, \mathcal{D}) &= G_m^{\text{PGA}}(f, \mathcal{D}) + \langle f_m^{\text{PGA}, \mathcal{D}}, g_{m+1} \rangle g_{m+1}, \\ f_{m+1}^{\text{PGA}, \mathcal{D}} &= f_0^{\text{PGA}, \mathcal{D}} - G_{m+1}^{\text{PGA}}(f, \mathcal{D}) = f_m^{\text{PGA}, \mathcal{D}} - \langle f_m^{\text{PGA}, \mathcal{D}}, g_{m+1} \rangle g_{m+1}. \end{aligned}$$

Работа выполнена при поддержке Министерства образования и науки РФ (государственный контракт № 07.514.12.4030).

ОРТОГОНАЛЬНЫЙ ЖАДНЫЙ АЛГОРИТМ (OGA). Положим

$$f_0^{\text{OGA}, \mathcal{D}} = f, \quad G_0^{\text{OGA}}(f, \mathcal{D}) = 0.$$

Если уже определен $f_m^{\text{OGA}, \mathcal{D}}$, найдем элемент $g_{m+1} \in \mathcal{D}$, удовлетворяющий условию

$$|\langle f_m^{\text{OGA}, \mathcal{D}}, g_{m+1} \rangle| = \sup_{g \in \mathcal{D}} |\langle f_m^{\text{OGA}, \mathcal{D}}, g \rangle|,$$

и положим

$$G_{m+1}^{\text{OGA}}(f, \mathcal{D}) = \text{Proj}_{g_1, \dots, g_{m+1}} f_0^{\text{OGA}, \mathcal{D}}, \quad f_{m+1}^{\text{OGA}, \mathcal{D}} = f_0^{\text{OGA}, \mathcal{D}} - G_{m+1}^{\text{OGA}}(f, \mathcal{D}).$$

Сходимость чисто жадного алгоритма для всех словарей $\mathcal{D}$ и функций $f \in H$ была доказана Рейтё и Вальтером [8] и, независимо, Джонсом [9]. Сходимость ортогонального жадного алгоритма для всех словарей $\mathcal{D}$ и функций $f \in H$ была установлена Дубининым [10].

Напомним также определение класса $\mathcal{A}_1(\mathcal{D})$. Для заданного словаря $\mathcal{D}$

$$\mathcal{A}_1(\mathcal{D}) = \bigcup_{M \geq 0} \overline{\mathcal{A}_1^0(\mathcal{D}, M)},$$

где

$$\mathcal{A}_1^0(\mathcal{D}, M) = \left\{ f \in H : f = \sum_{\lambda \in \Lambda} c_\lambda g_\lambda, g_\lambda \in \mathcal{D}, |\Lambda| < \infty, \sum_{\lambda \in \Lambda} |c_\lambda| \leq M \right\}.$$

Для $f \in \mathcal{A}_1(\mathcal{D})$ норма $|f|_{\mathcal{A}_1(\mathcal{D})}$ полагается равной $\inf \{M \geq 0 \mid f \in \mathcal{A}_1(\mathcal{D}, M)\}$. Напомним также, что класс $\mathcal{A}_0(\mathcal{D})$ – это класс, состоящий из всех конечных линейных комбинаций элементов словаря $\mathcal{D}$.

Оценка сверху на скорость сходимости ортогонального жадного алгоритма была получена ДеВором и Темляковым [4]. А именно, установлено, что для произвольного словаря $\mathcal{D}$ и элемента $f \in \mathcal{A}_1(\mathcal{D})$ и $m \geq 1$ выполняется оценка

$$\|f - G_m^{\text{OGA}}(f, \mathcal{D})\| \leq |f|_{\mathcal{A}_1(\mathcal{D})} m^{-1/2}.$$

Из приведенной оценки, а также результатов Стечкина [11] и Бэррона [12] следует оптимальность ортогонального жадного алгоритма на классе $\mathcal{A}_1(\mathcal{D})$.

Первая оценка сверху на скорость сходимости чисто жадного алгоритма была получена ДеВором и Темляковым [4], затем улучшена Конягиным и Темляковым [13], затем улучшена Сильниченко [14]. Последний результат может быть сформулирован следующим образом: существует такая константа $C > 0$, что для произвольного словаря $\mathcal{D}$, элемента $f \in \mathcal{A}_1(\mathcal{D})$ и $m \geq 1$ выполняется неравенство

$$\|f - G_m^{\text{PGA}}(f, \mathcal{D})\| \leq C |f|_{\mathcal{A}_1(\mathcal{D})} m^{-0.182}.$$

Первая оценка снизу на скорость сходимости чисто жадного алгоритма была получена ДеВором и Темляковым [15], затем улучшена Лившицем и Темляковым [15], затем улучшена Лившицем [16]. Последний результат заключается в следующем: существуют такой словарь $\mathcal{D}$, элемент $f \in \mathcal{A}_0(\mathcal{D})$ и константа $C > 0$, что для любого $m \geq 1$ выполняется неравенство

$$\|f - G_m^{\text{PGA}}(f, \mathcal{D})\| > C |f|_{\mathcal{A}_1(\mathcal{D})} m^{-0.1898}.$$

Таким образом, с точки зрения оценок скорости сходимости на всем классе $\mathcal{A}_1(\mathcal{D})$ ортогональный жадный алгоритм оптимален и существенно превосходит чисто жадный алгоритм.

Основным результатом настоящей работы является утверждение, показывающее, что для отдельных элементов класса $\mathcal{A}_1(\mathcal{D})$ (и даже $\mathcal{A}_0(\mathcal{D})$) ситуация может быть и противоположной: скорость сходимости ортогонального жадного алгоритма может быть существенно ниже скорости сходимости чисто жадного алгоритма. Это утверждение может быть формализовано следующим образом.

ТЕОРЕМА 1. *Существуют такие словарь $\mathcal{D}$ и элемент $f \in \mathcal{A}_0(\mathcal{D})$, что для любого $m \geq 3$ выполняется*

$$\|f_m^{\text{PGA}}\| = 0, \quad \text{но} \quad \|f_m^{\text{OGA}}\| \sim m^{-1/2}.$$

ДОКАЗАТЕЛЬСТВО. Рассмотрим пространство l^2 . Пусть $\{e_k\}_{k=1}^\infty$ — стандартный ортонормированный базис в этом пространстве. Положим $\mathcal{D} = \bigcup_{k=0}^\infty \{g_k\}$, где

$$g_0 = \alpha \left(-\frac{3\sqrt{3}}{4}, \frac{3}{4}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots \right), \quad g_1 = e_1, \quad g_2 = \left(\frac{1}{2}, \frac{\sqrt{3}}{2}, 0, 0, 0, \dots \right)$$

и $g_k = e_k$ для любого $k \geq 3$. Значение α найдем из условия $\|g_0\| = 1$:

$$\alpha = \left(\frac{9}{4} + \sum_{n=4}^\infty \frac{1}{n^2} \right)^{-1/2} < \frac{2}{3}.$$

Положим

$$f = \left(6, 3, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots \right) = \alpha^{-1} g_0 + 6g_1 + \frac{3\sqrt{3}}{2} g_2.$$

Так как f является трехчленной комбинацией элементов словаря $\mathcal{D}$, $f \in \mathcal{A}_0(\mathcal{D})$.

Осуществим чисто жадное разложение элемента f по словарю $\mathcal{D}$. Отметим, что в чисто жадном разложении скалярное произведение текущего остатка с элементом, выбранным на предыдущем шаге, равно нулю, поэтому нет нужды его считать. Имеем

$$\begin{aligned} f_0^{\text{PGA}} &= f = \left(6, 3, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots \right), \\ |\langle f_0^{\text{PGA}}, g_0 \rangle| &= \alpha \left| \left(-\frac{9\sqrt{3}}{2} + \frac{9}{4} + \sum_{n=4}^\infty \frac{1}{n^2} \right) \right| < 6, \quad |\langle f_0^{\text{PGA}}, g_1 \rangle| = 6, \\ |\langle f_0^{\text{PGA}}, g_2 \rangle| &= 3 + \frac{3\sqrt{3}}{2} < 6, \quad |\langle f_0^{\text{PGA}}, g_k \rangle| = \frac{1}{k+1} < 6 \quad \text{для любого } k \geq 3. \end{aligned}$$

Таким образом,

$$g_1^{\text{PGA}} = g_1 \quad \text{и} \quad f_1^{\text{PGA}} = \left(0, 3, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots \right).$$

Далее,

$$\begin{aligned} |\langle f_1^{\text{PGA}}, g_0 \rangle| &= \alpha \left(\frac{9}{4} + \sum_{n=4}^\infty \frac{1}{n^2} \right) < \frac{3\sqrt{3}}{2}, \quad |\langle f_1^{\text{PGA}}, g_2 \rangle| = \frac{3\sqrt{3}}{2}, \\ |\langle f_1^{\text{PGA}}, g_k \rangle| &= \frac{1}{k+1} < \frac{3\sqrt{3}}{2} \quad \text{для любого } k \geq 3. \end{aligned}$$

Следовательно, $g_2^{\text{PGA}} = g_2$ и второй остаток

$$f_2^{\text{PGA}} = \left(-\frac{3\sqrt{3}}{4}, \frac{3}{4}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots \right),$$

т.е. совпадает с $\alpha^{-1}g_0$.

Очевидно, что $g_3^{\text{PGA}} = g_0$ и $f_3^{\text{PGA}} = (0, 0, 0, 0, 0, \dots)$, т.е. на третьем шаге чисто жадное разложение завершено и $\|f_m^{\text{PGA}}\| = 0$ для любого $m \geq 3$.

Осуществим ортогональное жадное разложение элемента f по словарю $\mathcal{D}$. Отметим, что в ортогональном жадном разложении скалярное произведение текущего остатка с элементами, уже участвовавшими в разложении, равно нулю, поэтому нет нужды его считать. Также отметим, что из определений ортогонального и чисто жадного алгоритмов следует, что $g_1^{\text{OGA}} = g_1^{\text{PGA}}$ и $g_2^{\text{OGA}} = g_2^{\text{PGA}}$, поэтому

$$g_1^{\text{OGA}} = g_1, \quad g_2^{\text{OGA}} = g_2, \quad f_2^{\text{OGA}} = \left(0, 0, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots \right).$$

Далее, имеем

$$|\langle f_2^{\text{OGA}}, g_0 \rangle| = \alpha \sum_{n=4}^{\infty} \frac{1}{n^2} < \frac{2}{3} \int_3^{\infty} \frac{1}{x^2} dx = \frac{2}{9} < \frac{1}{4},$$

$$|\langle f_2^{\text{OGA}}, g_k \rangle| = \frac{1}{k+1} \quad \text{для любого } k \geq 3,$$

поэтому

$$g_3^{\text{OGA}} = g_3 \quad \text{и} \quad f_3^{\text{OGA}} = \left(0, 0, 0, \frac{1}{5}, \frac{1}{6}, \frac{1}{7}, \dots \right).$$

Далее, если

$$f_{m-1}^{\text{OGA}} = \left(\underbrace{0, 0, \dots, 0}_{m-1}, \frac{1}{m+1}, \frac{1}{m+2}, \dots \right), \quad m \geq 3,$$

то

$$|\langle f_{m-1}^{\text{OGA}}, g_0 \rangle| = \alpha \sum_{n=m+1}^{\infty} \frac{1}{n^2} < \frac{2}{3} \int_m^{\infty} \frac{1}{x^2} dx = \frac{2}{3m} < \frac{1}{m+1},$$

$$|\langle f_{m-1}^{\text{OGA}}, g_k \rangle| = \frac{1}{k+1} \quad \text{для любого } k \geq m;$$

следовательно,

$$g_m^{\text{OGA}} = g_m \quad \text{и} \quad f_m^{\text{OGA}} = \left(\underbrace{0, 0, \dots, 0}_m, \frac{1}{m+2}, \frac{1}{m+3}, \dots \right).$$

Таким образом, для любого $m \geq 3$ имеет место равенство

$$\|f_m^{\text{OGA}}\|^2 = \sum_{n=m+2}^{\infty} \frac{1}{n^2} \sim m^{-1},$$

из которого сразу следует соотношение

$$\|f_m^{\text{OGA}}\| \sim m^{-1/2}.$$

Теорема 1 доказана.

Утверждение следующей теоремы дополняет утверждение теоремы 1.

ТЕОРЕМА 2. Для любых $\alpha \geq \beta \geq 1/2$ существуют словарь $\mathcal{D}$, элемент $f \in \mathcal{A}_0(\mathcal{D})$ и константы C_1, C_2 такие, что

$$\|f_m^{\text{PGA}}\| \sim C_1 m^{-\alpha}, \quad \text{но} \quad \|f_m^{\text{OGA}}\| \sim C_2 m^{-\beta}.$$

СПИСОК ЦИТИРОВАННОЙ ЛИТЕРАТУРЫ

- [1] J. H. Friedman, W. Stuetzle, “Projection pursuit regression”, *J. Amer. Statist. Assoc.*, **76**:376 (1981), 817–823.
- [2] L. K. Jones, “On a conjecture of Huber concerning the convergence of projection pursuit regression”, *Ann. Statist.*, **15**:2 (1987), 880–882.
- [3] S. G. Mallat, Z. Zhang, “Matching pursuit with time-frequency dictionaries”, *IEEE Trans. Signal Process.*, **41**:12 (1993), 3397–3415.
- [4] R. A. DeVore, V. N. Temlyakov, “Some remarks on greedy algorithms”, *Adv. Comput. Math.*, **5**:2-3 (1996), 173–187.
- [5] V. N. Temlyakov, “Weak greedy algorithms”, *Adv. Comput. Math.*, **12**:2-3 (2000), 213–227.
- [6] V. N. Temlyakov, “Greedy algorithms in Banach spaces”, *Adv. Comput. Math.*, **14**:3 (2001), 277–292.
- [7] В. В. Галатенко, Е. Д. Лившиц, “Обобщенные приближенные слабые жадные алгоритмы”, *Матем. заметки*, **78**:2 (2005), 186–201.
- [8] L. Rejtő, G. G. Walter, “Remarks on projection pursuit regression and density estimation”, *Stochastic Anal. Appl.*, **10**:2 (1992), 213–222.
- [9] L. K. Jones, “A simple lemma on greedy approximation in Hilbert space and convergence rates for projection pursuit regression and network training”, *Ann. Statist.*, **20**:1 (1992), 608–613.
- [10] V. V. Dubinin, *Greedy Algorithms and Applications*, Ph.D. Thesis, Univ. South Carolina, 1997.
- [11] С. Б. Стечкин, “Об абсолютной сходимости ортогональных рядов”, *ДАН СССР*, **102** (1955), 37–40.
- [12] A. R. Barron, “Universal approximation bounds for superpositions of a sigmoidal function”, *IEEE Trans. Inform. Theory*, **39**:3 (1993), 930–945.
- [13] S. V. Konyagin, V. N. Temlyakov, “Rate of convergence of Pure Greedy Algorithm”, *East J. Approx.*, **5**:4 (1999), 493–499.
- [14] А. В. Сильниченко, “О скорости сходимости жадных алгоритмов”, *Матем. заметки*, **76**:4 (2004), 628–632.
- [15] E. D. Livshitz, V. N. Temlyakov, “Two lower estimates in greedy approximation”, *Constr. Approx.*, **19**:4 (2003), 509–523.
- [16] Е. Д. Лившиц, “О нижних оценках скорости сходимости жадных алгоритмов”, *Изв. РАН. Сер. матем.*, **73**:6 (2009), 125–144.

А. В. Деревенцов

Московский государственный университет

им. М. В. Ломоносова

E-mail: dereventsov@gmail.com

Поступило

01.12.2010

Исправленный вариант

22.03.2011