

СЛАБАЯ СХОДИМОСТЬ ГЕОМЕТРИЧЕСКОГО ВАРИАНТА МЕТОДА СОПРЯЖЕННЫХ ГРАДИЕНТОВ*

А. В. Плоткин
avplotkin@gmail.com

28 января 2019 г.

1°. Пусть $f \in C^1(\mathbb{R}^n)$. Рассмотрим экстремальную задачу

$$f(x) \rightarrow \min_{x \in \mathbb{R}^n}.$$

Для нахождения стационарной точки функции $f(x)$ применяется метод сопряженных градиентов [1–3]. В данной работе доказывается слабая сходимость геометрического варианта метода сопряженных градиентов.

Обозначим $g(x) = f'(x)$ и перейдем к описанию вычислительной схемы.

Начальный шаг. Возьмем произвольное начальное приближение $x_0 \in \mathbb{R}^n$ и вычислим в нем значение градиента $g_0 = g(x_0)$. Если $g_0 = \mathbb{0}$, то x_0 — стационарная точка. Вычисления заканчиваются. В противном случае укажем направление спуска $d_0 = -g_0$.

Общий шаг. Пусть имеются x_k, d_k . Введем функцию

$$\varphi_k(\alpha) = f(x_k + \alpha d_k), \quad \alpha > 0.$$

Следующее приближение x_{k+1} вычислим по формуле

$$x_{k+1} = x_k + \alpha_k d_k,$$

где α_k является первой стационарной точкой функции $\varphi_k(\alpha)$:

$$\alpha_k = \min\{\alpha > 0 \mid \varphi_k'(\alpha) = 0\}.$$

В полученной точке найдем значение градиента $g_{k+1} = g(x_{k+1})$. Если $g_{k+1} = \mathbb{0}$, то x_{k+1} — стационарная точка. Вычисления заканчиваются. В противном случае укажем следующее направление спуска:

$$d_{k+1} = \frac{\|g_{k+1}\|^2 d_k + \|d_k\|^2 (-g_{k+1})}{\|g_{k+1}\|^2 + \|d_k\|^2}.$$

*Семинар по конструктивному негладкому анализу и недифференцируемой оптимизации «CNSA & NDO»: <http://www.apmath.spbu.ru/cnsa/>

Для доказательства слабой сходимости представленного метода потребуем выполнения двух условий.

УСЛОВИЕ 1. Множество $M = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ ограничено.

УСЛОВИЕ 2. В некоторой окрестности U множества M градиент $g(x)$ удовлетворяет условию Липшица, то есть существует константа $L > 0$, такая что

$$\|g(x) - g(y)\| \leq L \|x - y\| \quad \forall x, y \in U.$$

Сформулируем теорему о слабой сходимости геометрического варианта метода сопряженных градиентов.

ТЕОРЕМА (о слабой сходимости). Пусть выполнены условия 1 и 2, тогда

$$\lim_{k \rightarrow \infty} \|g_k\| = 0.$$

2°. Предварительно отметим некоторые свойства описанного метода.

СВОЙСТВО 1. Для всех $k \geq 1$ выполняется равенство $\langle -g_k, d_{k-1} \rangle = 0$.

Доказательство. Следует из определения α_k . □

СВОЙСТВО 2. Для всех $k \geq 0$ выполняется равенство $\langle -g_k, d_k \rangle = \|d_k\|^2$.

Доказательство. При $k = 0$ равенство очевидно. Рассмотрим случай $k \geq 1$. Воспользовавшись свойством 1, распишем обе части доказываемого равенства:

$$\begin{aligned} \langle -g_k, d_k \rangle &= \frac{\|g_k\|^2 \|d_{k-1}\|^2}{\|g_k\|^2 + \|d_{k-1}\|^2} \quad \forall k \geq 1, \\ \|d_k\|^2 &= \frac{\|g_k\|^4 \|d_{k-1}\|^2 + \|d_{k-1}\|^4 \|g_k\|^2}{(\|g_k\|^2 + \|d_{k-1}\|^2)^2} = \frac{\|g_k\|^2 \|d_{k-1}\|^2}{\|g_k\|^2 + \|d_{k-1}\|^2} \quad \forall k \geq 1. \end{aligned}$$

Отсюда следует требуемое равенство для всех $k \geq 0$. □

На рис. 1 приведено взаимное положение векторов d_{k-1} , d_k и $-g_k$.

При выполнении условия 1 метод является реализуемым. Действительно, на каждом шаге метода направление d_k сохраняет свойство убывания, так как

$$\varphi'_k(0) = \langle g_k, d_k \rangle = -\|d_k\|^2 < 0,$$

что совместно с условием 1 гарантирует существование стационарной точки функции $\varphi_k(\alpha)$ и положительность α_k .

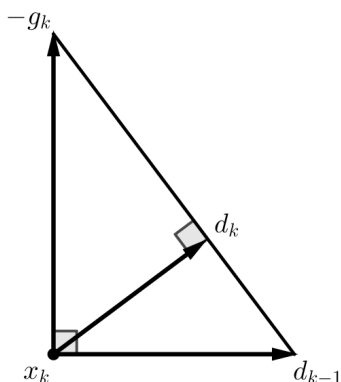


Рис. 1. Геометрия метода

ЛЕММА. Пусть выполнены условия 1 и 2, тогда

$$\sum_{k \geq 0} \|d_k\|^2 < +\infty.$$

Доказательство. В силу условия 2 и свойства 2 имеем

$$\begin{aligned} \varphi'_k(\alpha) &= \varphi'_k(0) + (\varphi'_k(\alpha) - \varphi'_k(0)) = \\ &= \varphi'_k(0) + (\langle g(x_k + \alpha d_k), d_k \rangle - \langle g(x_k), d_k \rangle) = \\ &= \varphi'_k(0) + \langle g(x_k + \alpha d_k) - g(x_k), d_k \rangle \leq \varphi'_k(0) + L\alpha \|d_k\|^2 = \\ &= (L\alpha - 1) \|d_k\|^2 \quad \forall k \geq 0. \end{aligned}$$

Подставив в полученное неравенство α_k , получим

$$\alpha_k \geq \frac{1}{L} \quad \forall k \geq 0.$$

Следовательно,

$$\begin{aligned} \varphi_k(0) - \varphi_k(\alpha_k) &= - \int_0^{\alpha_k} \varphi'_k(\alpha) d\alpha = - \int_0^{\frac{1}{L}} \varphi'_k(\alpha) d\alpha - \int_{\frac{1}{L}}^{\alpha_k} \varphi'_k(\alpha) d\alpha \geq \\ &\geq - \int_0^{\frac{1}{L}} \varphi'_k(\alpha) d\alpha \geq \int_0^{\frac{1}{L}} (1 - L\alpha) \|d_k\|^2 d\alpha = \frac{1}{2L} \|d_k\|^2 \quad \forall k \geq 0. \end{aligned}$$

Таким образом,

$$f(x_i) = f(x_0) - \sum_{k=0}^{i-1} (f(x_k) - f(x_{k+1})) \leq f(x_0) - \frac{1}{2L} \sum_{k=0}^{i-1} \|d_k\|^2 \quad \forall i \geq 0.$$

В силу условия 1 функция $f(x)$ ограничена снизу. Значит,

$$\sum_{k \geq 0} \|d_k\|^2 < +\infty.$$

□

Доказательство теоремы. Пусть вывод теоремы неверен. Тогда найдется константа $\varepsilon > 0$, такая что

$$\|g_k\|^2 \geq \varepsilon \quad \forall k \geq 0. \quad (1)$$

Докажем по индукции, что в этом случае

$$\|d_k\|^2 \geq \frac{\varepsilon}{k+1} \quad \forall k \geq 0. \quad (2)$$

При $k = 0$ неравенство верно. Сделаем индукционный переход от k к $k + 1$:

$$\|d_{k+1}\|^2 = \frac{\|g_{k+1}\|^2 \|d_k\|^2}{\|g_{k+1}\|^2 + \|d_k\|^2} = \frac{1}{\frac{1}{\|d_k\|^2} + \frac{1}{\|g_{k+1}\|^2}}.$$

Воспользовавшись индукционным предположением и неравенством (1), получим

$$\|d_{k+1}\|^2 \geq \frac{1}{\frac{k+1}{\varepsilon} + \frac{1}{\varepsilon}} = \frac{\varepsilon}{k+2}.$$

Таким образом, неравенство (2) установлено. Следовательно,

$$\sum_{k \geq 0} \|d_k\|^2 \geq \varepsilon \sum_{k \geq 0} \frac{1}{k+1} = +\infty.$$

Получили противоречие с доказанной леммой. Сходимость установлена. \square

3°. Рассмотрим вычислительные аспекты геометрического варианта метода сопряженных градиентов.

Нахождение первой стационарной точки функции $\varphi_k(\alpha) = f(x_k + \alpha d_k)$ является трудоемкой процедурой. Поэтому на практике используются более простые варианты линейного поиска. Одним из таких вариантов является линейный поиск с условием Армихо. Параметрами поиска выступают числа $0 < \varepsilon < 1$ и $\hat{\alpha} > 0$. Результатом является значение $\alpha_k \in (0, \hat{\alpha}]$, удовлетворяющее условию

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \varepsilon \alpha_k \langle g_k, d_k \rangle.$$

Для корректной работы линейного поиска необходимо убывание целевой функции в направлении поиска d_k . Но в общем случае это условие может нарушаться. Тогда метод сопряженных градиентов перезапускается с текущим приближением в качестве начального. Дополнительно перезапуск производится через каждые n итераций. Это объясняется тем фактом, что для квадратичной целевой функции метод сходится не более чем за n шагов. Приведем пример, демонстрирующий необходимость периодического обновления.

Рассмотрим задачу минимизации функции Химмельблау

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2.$$

В качестве начального приближения возьмем точку $(-2, 3.5)$. Параметрами линейного поиска выберем числа $\varepsilon = \frac{1}{4}$ и $\hat{\alpha} = 10$. Вычисления будем заканчивать, когда количество итераций достигнет 1000 или норма градиента уменьшится до 10^{-9} .

На рис. 2 и 3 приведены результаты вычислений в случае отсутствия обновления. После 1000 итераций норма градиента не опустилась ниже 0.1.

На рис. 4 и 5 приведены результаты вычислений в случае обновления. За 21 итерацию было получено решение с требуемой точностью.

4°. Перейдем к задаче минимизации многомерной функции BDEXP [4]

$$f(x) = \sum_{i=1}^{n-2} (x_i + x_{i+1}) \exp(-x_{i+2}(x_i + x_{i+1}))$$

при $n = 100$.

В качестве начального приближения возьмем точку $(1, \dots, 1)$. Параметр ε линейного поиска, как и ранее, выберем равным $\frac{1}{4}$. Второй параметр $\hat{\alpha}$ будем варьировать в зависимости от итерации. На k -й итерации будем вычислять значение $\hat{\alpha} = \hat{\alpha}_k$ следующим образом. Сначала положим $\hat{\alpha}_k := \hat{\alpha}_{k-1}$ в случае $k \geq 1$ и $\hat{\alpha}_0 := 1$ в случае $k = 0$. Далее, если неравенство Армихо выполняется при текущем значении $\hat{\alpha}_k$, то будем удваивать его до тех пор, пока оно не нарушится.

На рис. 6 приведены результаты вычислений. За 9 итераций было получено решение с требуемой нормой градиента.

ЛИТЕРАТУРА

1. W. W. Hager and H. Zhang. *A Survey of Nonlinear Conjugate Gradient Methods* // Pacific Journal of Optimization, Vol. 2, 2006, pp. 35–58.
2. В. Н. Малоземов. *О методе сопряженных градиентов* // Семинар «DHA & CAGD». Избранные доклады. 8 апреля 2012 г. (<http://dha.spb.ru/reps12.shtml#0428>)
3. В. Н. Малоземов. *Варианты метода сопряженных градиентов* // Семинар «CNSA & NDO». Избранные доклады. 29 октября 2015 г. (<http://www.apmath.spbu.ru/cnsa/reps15.shtml#1029>)
4. Andrei N. *An unconstrained optimization test functions collection*. // Advanced Modeling and Optimization. 2008. Vol. 10, No. 1, pp. 147–161.

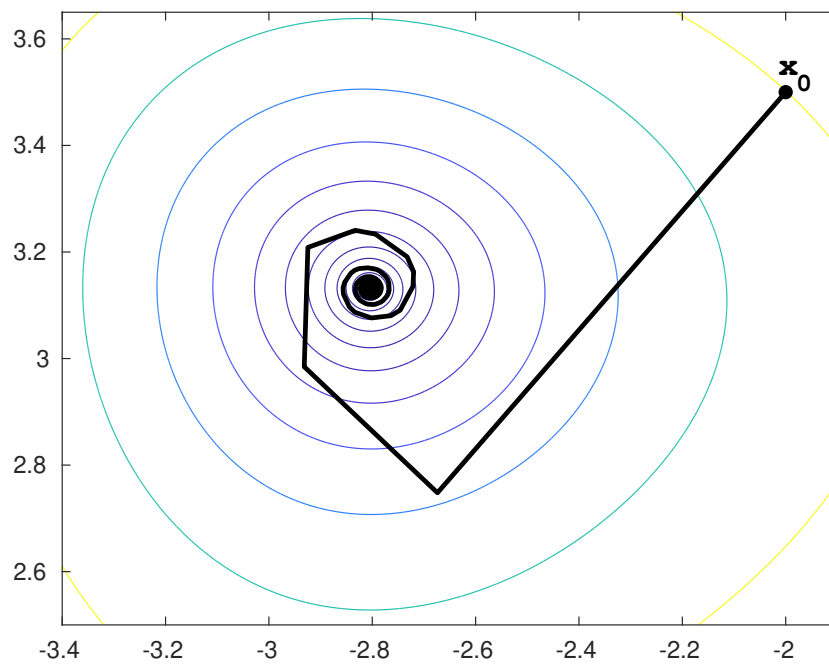


Рис. 2. Траектория спуска при отсутствии обновления

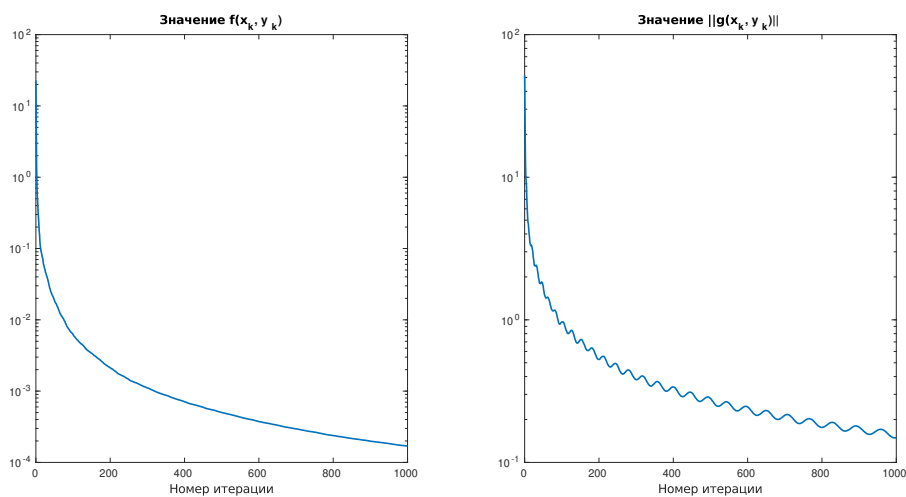


Рис. 3. Зависимость функции Химмельблау и нормы ее градиента от номера итерации

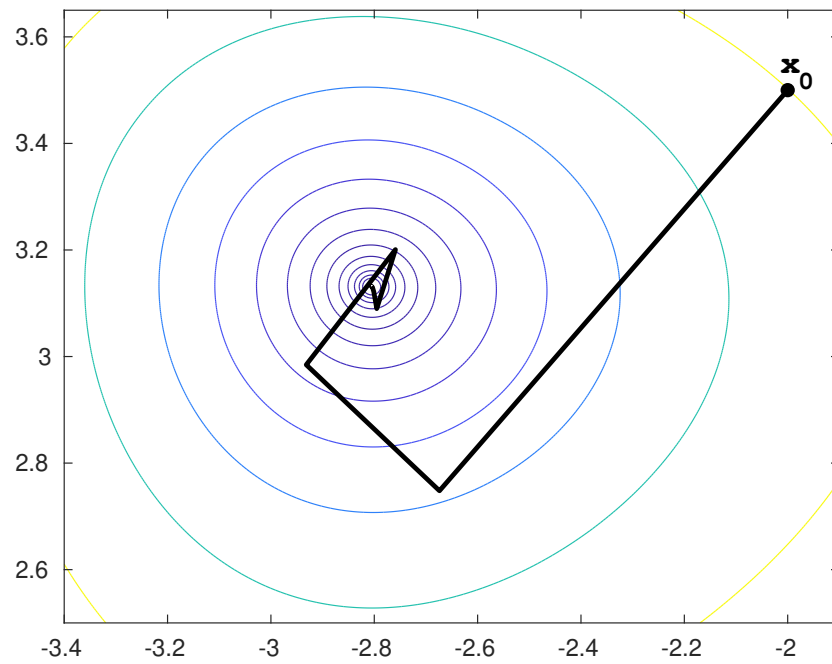


Рис. 4. Траектория спуска при использовании обновления

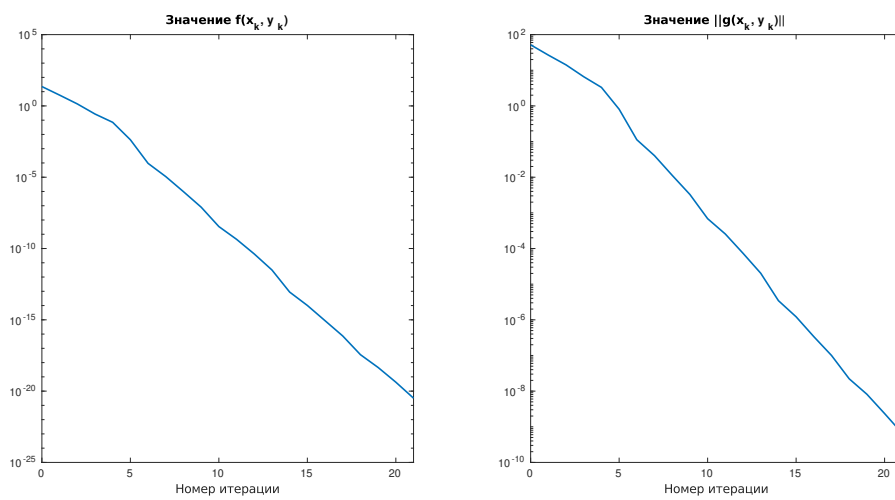


Рис. 5. Зависимость функции Химмельблау и нормы ее градиента от номера итерации

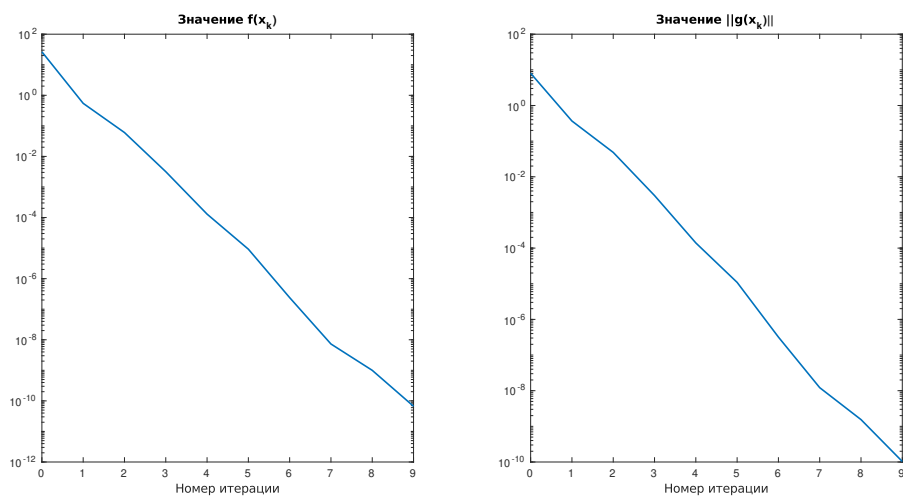


Рис. 6. Зависимость функции BDEXP и нормы ее градиента от номера итерации