ГРАДИЕНТНЫЙ МЕТОД С РЕГУЛИРОВКОЙ ШАГА ДЛЯ БЕЗУСЛОВНОЙ МИНИМИЗАЦИИ*

В. Н. Малозёмов — v.malozemov@spbu.ru

Г. Ш. Тамасян — g.tamasyan@spbu.ru

21 марта 2019 г.

Аннотация. С современных позиций анализируется градиентный метод с регулировкой шага для безусловной минимизации гладких функций, предложенный Ю. М. Данилиным в 1975 г. (см. [1, с. 52–58 и комментарий на с. 171]).

1°. Опишем градиентный метод с регулировкой шага для безусловной минимизации функции $f(x)$ из класса $C^1(\mathbb{R}^n)$.

Параметрами метода являются числа $\varepsilon \in (0,1)$ и $\alpha > 0$.

В качестве начального приближения берём произвольную точку $x_0 \in \mathbb{R}^n$.

Пусть имеется k-е приближение x_k. Если градиент $f'(x_k)$ равен нулю, то вычисления заканчиваются. В противном случае переходим к очередному приближению

$$x_{k+1} = x_k - \alpha_k f'_k, \tag{1}$$

где $f'_k = f'(x_k)$ и α_k — первое число в последовательности $\alpha, \frac{1}{2}\alpha, \frac{1}{4}\alpha, \ldots,$ при котором выполняется неравенство

$$f(x_k - \alpha_k f'_k) \leqslant f(x_k) - \varepsilon\alpha_k \|f'(x_k)\|^2. \tag{2}$$

После этого вычисления повторяются.

2°. По описанию видно, что метод ориентирован на получение стационарной точки функции $f(x)$ — точки, в которой градиент равен нулю. Это может случится на конечной итерации, если $f'(x_k) = \mathbf{0}$. Нас интересует основной случай, когда последовательность $\{x_k\}$ бесконечная, то есть когда $f'(x_k) \neq \mathbf{0}$ при всех $k = 0, 1, 2, \ldots$

ТЕОРЕМА 1 (о слабой сходимости градиентного метода с регулировкой шага). *Пусть функция $f(x)$ ограничена снизу на $\mathbb{R}^n$ и её градиент удовлетворяет условию Липшица*

$$\|f'(x) - f'(y)\| \leqslant L\|x - y\| \quad \textit{при всех } x, y \textit{ из } \mathbb{R}^n. \tag{3}$$

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

Тогда справедливо предельное соотношение

$$\lim_{k\to\infty} f'(x_k) = \mathbf{0}. \tag{4}$$

Доказательство. Введём обозначение (u, v) для открытого интервала, соединяющего точки u и v из $\mathbb{R}^n$:

$$(u, v) = \{x = u + t(v - u) \mid t \in (0, 1)\}.$$

По теореме о среднем имеем

$$f(x_k - tf'_k) = f(x_k) - t\|f'_k\|^2 - t\langle f'_{kc} - f'_k, f'_k\rangle,$$

где $f'_{kc} = f'(x_{kc})$ и $x_{kc} \in (x_k, x_k - tf'_k)$. Согласно (3)

$$f(x_k - tf'_k) \leqslant f(x_k) - t\|f'_k\|^2 + Lt^2\|f'_k\|^2 = \\ = f(x_k) - t\|f'_k\|^2(1 - Lt).$$

Отсюда следует, что при $t > 0$ и $1 - Lt \geqslant \varepsilon$, то есть при

$$0 < t \leqslant \frac{1-\varepsilon}{L},$$

выполняется неравенство

$$f(x_k - tf'_k) \leqslant f(x_k) - \varepsilon t\|f'_k\|^2. \tag{5}$$

Напомним, что шаг спуска α_k выбирается из условия (2). Если $\alpha \leqslant \dfrac{1-\varepsilon}{L}$, то в силу (5) будет $\alpha_k = \alpha$ при всех $k = 0, 1, 2, \ldots$, т. е. градиентный метод с регулировкой шага переходит в градиентный метод с постоянным шагом.

Пусть $\alpha > \dfrac{1-\varepsilon}{L}$. Тогда найдётся натуральное число s, такое, что

$$\alpha\frac{1}{2^s} \leqslant \frac{1-\varepsilon}{L} < \alpha\frac{1}{2^{s-1}}.$$

По правилу выбора α_k и (5) имеем

$$\alpha_k \geqslant \alpha\frac{1}{2^s} = \alpha\frac{1}{2^{s-1}2} > \frac{1-\varepsilon}{2L}.$$

Во всех случаях

$$\alpha_k \geqslant \min\{\alpha, \tfrac{1-\varepsilon}{2L}\}. \tag{6}$$

На основании (1), (2) и (6) получаем

$$f(x_{k+1}) \leqslant f(x_k) - \varepsilon\min\{\alpha, \tfrac{1-\varepsilon}{2L}\}\|f'_k\|^2.$$

Перепишем последнее неравенство в виде

$$f(x_k) - f(x_{k+1}) \geqslant \varepsilon\min\{\alpha, \tfrac{1-\varepsilon}{2L}\}\|f'_k\|^2, \quad k = 0, 1, 2, \ldots$$

Отсюда в силу ограниченности снизу монотонно убывающей последовательности $\{f(x_k)\}$ следует, что $\|f'_k\|^2 \longrightarrow 0$ при $k \longrightarrow \infty$. Это равносильно (4). Теорема доказана. □

3°. Исследование сильной сходимости и скорости сходимости градиентного метода с регулировкой шага будем проводить для сильно выпуклых функций. По определению функция $f \in C^2(\mathbb{R}^n)$ называется *сильной выпуклой*, если при всех x и h из $\mathbb{R}^n$ выполняется условие

$$\langle f''(x)h, h\rangle \geqslant m\|h\|^2, \tag{7}$$

где $m > 0$. Поясним смысл введённого термина

ЛЕММА 1. *Если $f(x)$ — сильно выпуклая функция, то*

$$f(tx_1 + (1-t)x_0) \leqslant tf(x_1) + (1-t)f(x_0) - \tfrac{1}{2}mt(1-t)\|x_1 - x_0\|^2 \tag{8}$$

при всех x_0, x_1 из $\mathbb{R}^n$ и $t \in [0, 1]$.

Доказательство.[1] По теореме о среднем имеем

$$f(v) - f(u) = \langle f'(u), v - u\rangle + \tfrac{1}{2}\langle f''(\xi)(v-u), v-u\rangle,$$

где $\xi \in (u, v)$. В силу (7)

$$f(v) - f(u) \geqslant \langle f'(u), v - u\rangle + \tfrac{1}{2}m\|v - u\|^2. \tag{9}$$

Зафиксируем x_0, x_1 из $\mathbb{R}^n$ и $t \in [0, 1]$. Обозначим $u = tx_1 + (1-t)x_0$. Дважды воспользуемся неравенством (9). Первый раз при $v = x_1$ и второй раз при $v = x_0$. Запишем

$$\begin{gathered} f(x_1) - f(u) \geqslant \langle f'(u), x_1 - u\rangle + \tfrac{1}{2}m(1-t)^2\|x_1 - x_0\|^2, \\ f(x_0) - f(u) \geqslant \langle f'(u), x_0 - u\rangle + \tfrac{1}{2}mt^2\|x_1 - x_0\|^2. \end{gathered}$$

Теперь первое неравенство умножим на t, второе — на $1 - t$, после чего полученные неравенства сложим «в столбик». Придём к неравенству

$$\begin{gathered} tf(x_1) + (1-t)f(x_0) - f(u) \geqslant \tfrac{1}{2}m\|x_1 - x_0\|^2\left(t(1-t)^2 + (1-t)t^2\right) = \\ = \tfrac{1}{2}mt(1-t)\|x_1 - x_0\|^2, \end{gathered}$$

что равносильно (8). Лемма доказана. □

Нам потребуется следующее свойство сильно выпуклых функций.

ЛЕММА 2. *У сильно выпуклой функции $f(x)$ существует и единственна точка глобального минимума x_*. В ней и только в ней градиент функции $f(x)$ обращена в ноль.*

Доказательство имеется, например, в [2, лемма 2].

[1] Это доказательство предложила студентка мат-мех факультета Настя Кукульская.

4°. Обратимся к анализу сильной сходимости градиентного метода с регулировкой шага для безусловной минимизации

ТЕОРЕМА 2. *Пусть функция $f(x)$ принадлежит классу $C^2(\mathbb{R}^n)$ и при всех x и h из $\mathbb{R}^n$ выполняются неравенства*

$$m\|h\|^2 \leqslant \langle f''(x)h, h\rangle \leqslant M\|h\|^2, \tag{10}$$

где m и M — положительные константы. Пусть также параметры ε и α связаны соотношением

$$\alpha > \frac{2(1-\varepsilon)}{M}.$$

Тогда последовательность $\{x_k\}$, построенная градиентным методом с регулировкой шага, сходится к точке глобального минимума x_* функции $f(x)$. При этом справедлива следующая оценка скорости сходимости

$$\|x_k - x_*\|^2 \leqslant Cq^k, \quad k = 0, 1, \dots,$$

где

$$q = 1 - \frac{\varepsilon(1-\varepsilon)m}{M}\left(1 + \frac{m}{M}\right) \quad \text{и} \quad C = \frac{2}{m}\left(f(x_0) - f(x_*)\right).$$

5°. Предварительно докажем одно вспомогательное утверждение.

ЛЕММА 3. *В условиях теоремы 2 справедливо неравенство*

$$\|f'(x)\|^2 \geqslant m\left(1 + \frac{m}{M}\right)[f(x) - f(x_*)] \quad \forall\, x \in \mathbb{R}^n. \tag{11}$$

Доказательство. Так как $f'(x_*) = \mathbf{0}$, то по теореме о среднем имеем

$$f(x) - f(x_*) = \tfrac{1}{2}\langle f''(\xi)(x - x_*), x - x_*\rangle, \quad \xi \in (x, x_*).$$

В силу условия (10)

$$\tfrac{1}{2}m\|x - x_*\|^2 \leqslant f(x) - f(x_*) \leqslant \tfrac{1}{2}M\|x - x_*\|^2. \tag{12}$$

В частности,

$$\|x - x_*\|^2 \geqslant \frac{2}{M}[f(x) - f(x_*)]. \tag{13}$$

Снова воспользуемся теоремой о среднем:

$$f(x_*) - f(x) = \langle f'(x), x_* - x\rangle + \tfrac{1}{2}\langle f''(\eta)(x_* - x), x_* - x\rangle, \quad \eta \in (x_*, x).$$

В силу условия (10)

$$f(x) - f(x_*) \leqslant \|f'(x)\| \cdot \|x - x_*\| - \tfrac{1}{2}m\|x - x_*\|^2. \tag{14}$$

Из (12) и (14) следует, что

$$\tfrac{1}{2}m\|x-x_*\|^2 \leqslant f(x)-f(x_*) \leqslant \|f'(x)\|\cdot\|x-x_*\| - \tfrac{1}{2}m\|x-x_*\|^2. \tag{15}$$

После очевидных преобразований придём к неравенству

$$\|x-x_*\| \leqslant \frac{1}{m}\|f'(x)\|. \tag{16}$$

Вернёмся к формуле (14). С учётом (13) и (16) получаем

$$f(x)-f(x_*) \leqslant \frac{1}{m}\|f'(x)\|^2 - \frac{m}{M}\left[f(x)-f(x_*)\right],$$

что равносильно (11). Лемма доказана. □

6°. Переходим к доказательству теоремы 2. По теореме о среднем имеем

$$f(x_k - tf'_k) = f(x_k) - t\|f'_k\|^2 + \tfrac{1}{2}t^2\langle f''(x_{kc})f'_k, f'_k\rangle,$$

где $x_{kc} \in (x_k, x_k - tf'_k)$. В силу (10)

$$\begin{aligned} f(x_k - tf'_k) \leqslant f(x_k) - t\|f'_k\|^2 + \tfrac{1}{2}Mt^2\|f'_k\|^2 = \\ = f(x_k) - t\|f'_k\|^2(1-\tfrac{1}{2}Mt). \end{aligned}$$

Отсюда следует, что при $t>0$ и $1-\frac{1}{2}Mt \geqslant \varepsilon$, то есть при

$$0 < t \leqslant \frac{2(1-\varepsilon)}{M},$$

выполняется неравенство

$$f(x_k - tf'_k) \leqslant f(x_k) - \varepsilon t\|f'_k\|^2. \tag{17}$$

По условию теоремы $\alpha > \dfrac{2(1-\varepsilon)}{M}$. Выберем натуральное s так, чтобы

$$\alpha\frac{1}{2^s} \leqslant \frac{2(1-\varepsilon)}{M} < \alpha\frac{1}{2^{s-1}}.$$

Тогда по правилу выбора шага α_k и (17) получаем

$$\alpha_k \geqslant \alpha\frac{1}{2^s} = \alpha\frac{1}{2^{s-1}2} > \frac{1-\varepsilon}{M}. \tag{18}$$

На основании (1), (2) и (18) приходим к неравенству

$$f(x_{k+1}) \leqslant f(x_k) - \frac{\varepsilon(1-\varepsilon)}{M}\|f'_k\|^2.$$

Оценка (11) позволяет продолжить преобразования:

$$f(x_{k+1}) - f(x_*) \leqslant \left[1 - \frac{\varepsilon(1-\varepsilon)m}{M}\left(1 + \frac{m}{M}\right)\right]\left(f(x_k) - f(x_*)\right).$$

Обозначим

$$q = 1 - \frac{\varepsilon(1-\varepsilon)m}{M}\left(1 + \frac{m}{M}\right)$$

и перепишем последнее неравенство в виде

$$f(x_{k+1}) - f(x_*) \leqslant q\left(f(x_k) - f(x_*)\right).$$

Отсюда следует, что

$$f(x_k) - f(x_*) \leqslant q^k\left(f(x_0) - f(x_*)\right), \quad k = 0, 1, 2, \ldots$$

Теперь воспользуемся первым неравенством в (15). Получим

$$\|x_k - x_*\|^2 \leqslant \frac{2}{m}\left(f(x_k) - f(x_*)\right) \leqslant \frac{2}{m}q^k\left(f(x_0) - f(x_*)\right),$$

что и требовалось доказать. □

З а м е ч а н и е. Очевидно, что $q < 1$. По условию $\varepsilon \in (0, 1)$, так что

$$\frac{\varepsilon(1-\varepsilon)m}{M}\left(1 + \frac{m}{M}\right) \leqslant 2\varepsilon(1-\varepsilon) \leqslant \frac{1}{2} \quad \text{и} \quad q \geqslant \frac{1}{2}.$$

Значит, $q \in [\frac{1}{2}, 1)$.

Наименьшее значение q достигается при $\varepsilon = \frac{1}{2}$ и равно следующей величине

$$q_{\min} = 1 - \frac{1}{4}\frac{m}{M}\left(1 + \frac{m}{M}\right).$$

7°. В этом разделе на примерах демонстрируется работа градиентного метода с регулировкой шага. Обратимся к банку тестовых задач безусловной оптимизации [2].

ПРИМЕР 1. Рассмотрим задачу минимизации функции Химмельблау

$$f(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

(см. [2, с. 151]). Зафиксируем следующие значения параметров метода: $\varepsilon = \frac{1}{2}$ и $\alpha = 1$. В качестве начального приближения возьмем точку $x_0 = (-2, 3.5)$, предложенную в [3]. Вычисления будем заканчивать, когда количество итераций достигнет 1000 или норма градиента станет меньше 10^{-10}. На рис. 1 и 2 приведены результаты вычислений.

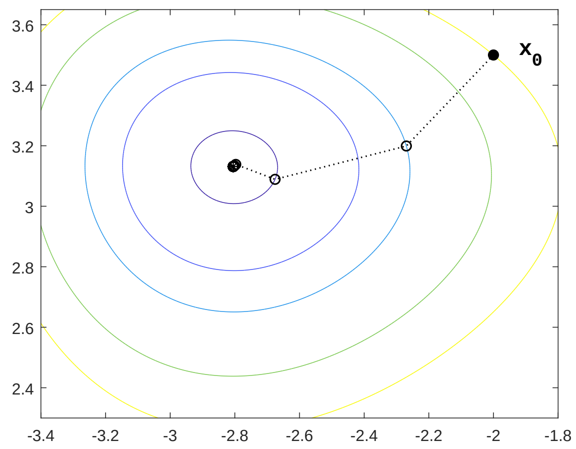

Рис. 1. Траектория спуска

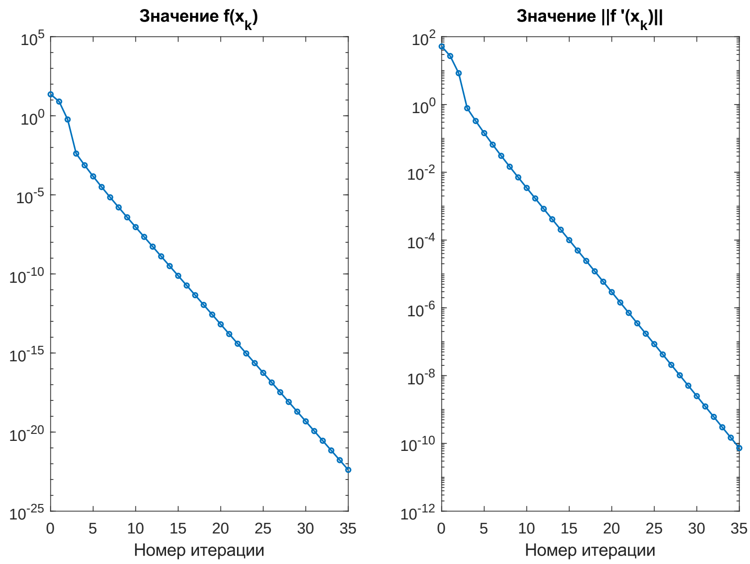

Рис. 2. Поведение функции Химмельблау по итерациям

За 35 итераций было получено решение с требуемой точностью. Обратим внимание на характер изменения величины шага α_k:

$$\alpha_0 = 2^{-7}, \quad \alpha_1 = \alpha_2 = 2^{-6}, \quad \alpha_j = 2^{-7} \text{ при } \; j = 3:34.$$

Общее число проверок выполнения неравенства (5) составляет 278.

ПРИМЕР 2. Применим рассматриваемый метод к многомерной функции BDEXP [2, с. 157]:

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

при $n = 100$.

В качестве начального приближения возьмем точку $(1, \ldots, 1)$. Значения параметров метода оставим прежними, т. е. $\varepsilon = \frac{1}{2}$ и $\alpha = 1$. Как видно из рис. 3 после 1000 итераций норма градиента не достигла 10^{-3}.

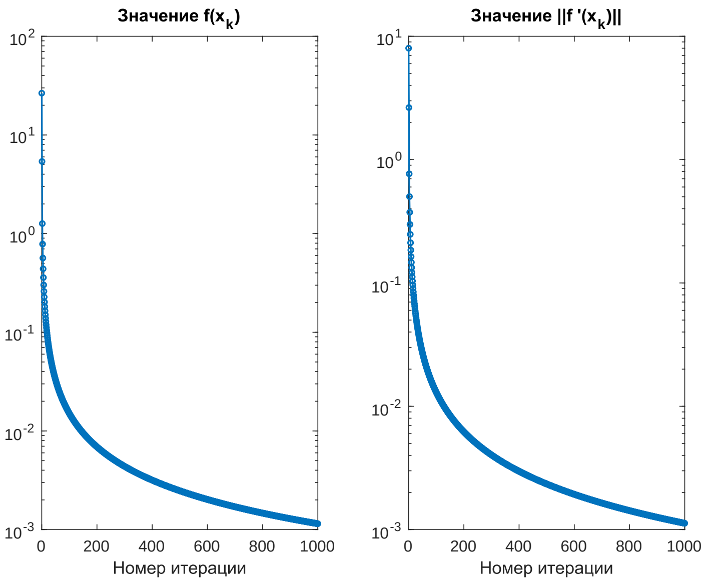

Рис. 3. Поведение функции BDEXP по итерациям

Приведём значения α_k:

$$\alpha_0 = \tfrac{1}{2}, \quad \alpha_j = 1 \text{ при } \ j \geqslant 1.$$

8°. Согласно описанию метода (см. п. 1) величина параметра α остается неизменной на протяжении всех итераций. Напомним, что при построении очередного приближения x_{k+1} приходится проверять выполнения неравенства (5) при различных значениях параметра t. Рассмотренные примеры позволяют сделать вывод о целесообразности варьирования параметра α от итерации к итерации. Действительно, соседние величины α_k и α_{k+1} «мало» отличаются друг от друга, поэтому на $(k+1)$-й итерации в качестве начального шага α можно взять α_k. А в случае выполнения неравенства (5) при $t = \alpha$, проверить его справедливость при бóльших значениях параметра α.

Приведенные ниже результаты вычислений подтверждают существенное увеличение эффективности градиентного метода с регулировкой шага, если величину α на $(k+1)$-й итерации выбирать согласно следующему правилу.

Пусть уже имеется α_k, $k \geqslant 0$. Переопределим α, положив $\alpha := \alpha_k$. Проверим выполнение неравенства (5) при $t = \alpha$.

- Если оно выполнено, то в качестве α_{k+1} выбирается наибольшее число в последовательности $\alpha,\ 2\alpha, 2^2\alpha, \ldots,$ при котором справедливо неравенство (5).

- В противном случае α_{k+1} равно первому числу в последовательности $\frac{1}{2}\alpha, \frac{1}{4}\alpha, \ldots,$ при котором выполнено неравенство (5).

9°. На тех же примерах рассмотрим работу градиентного метода с регулировкой шага по предложенному правилу.

ПРИМЕР 3. На рис. 4 приведены результаты вычислений в случае варьирования параметра α в задаче минимизации функции BDEXP. После 18 итераций получено решение с требуемой точностью. Приведём значения α_k:

$$\alpha_0 = \tfrac{1}{2}, \quad \alpha_1 = 1, \quad \alpha_2 = 2, \quad \alpha_j = 4^{j-2} \text{ при } \ j = 3 : 17.$$

Отметим, что неравенство (5) проверялось 67 раз.

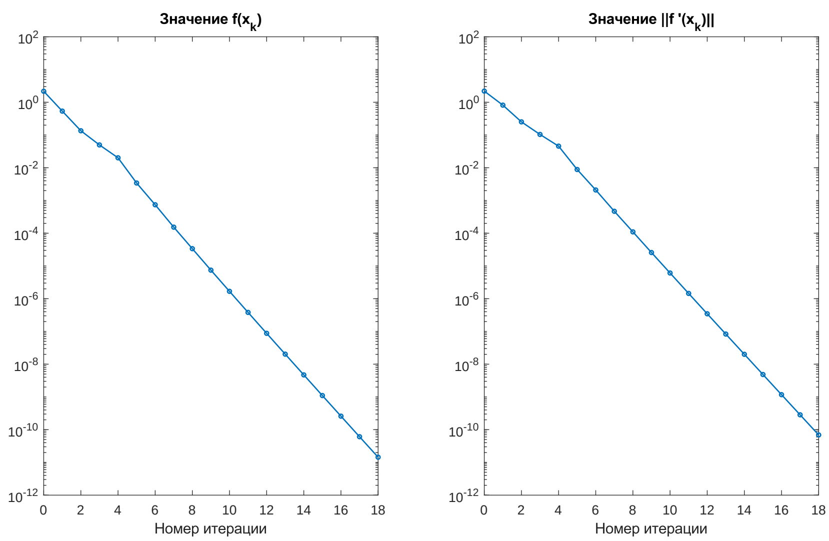

Рис. 4. Поведение функции BDEXP по итерациям

ПРИМЕР 4. Для функции Химмельблау (см. пример 1), в случае варьирования параметра α, результаты вычислений не изменились, однако общее число проверок неравенства (5) сократилось с 278 до 77.

ЛИТЕРАТУРА

1. Пшеничный Б. Н., Данилин Ю. М. Численные методы в экстремальных задачах. М.: Изд-во «Наука», 1975. 320 с.

2. Andrei N. *An unconstrained optimization test functions collection.* // Advanced Modeling and Optimization. 2008. Vol. 10, No. 1, pp. 147–161.

3. Плоткин А. В. *Слабая сходимость геометрического варианта метода сопряжённых градиентов* // Семинар «CNSA & NDO». Избранные доклады. 28 января 2019 г. (`http://apmath.spbu.ru/cnsa/reps19.shtml#0128`)