

© 2019 г. В.Н. КРУТИКОВ, д-р техн. наук (krutikovvn@rambler.ru),
Н.С. САМОЙЛЕНКО (nostienataly@mail.ru),
В.В. МЕШЕЧКИН, канд. физ.-мат. наук (vvm@kemsu.ru)
(Кемеровский государственный университет)

О СВОЙСТВАХ МЕТОДА МИНИМИЗАЦИИ ВЫПУКЛЫХ ФУНКЦИЙ, РЕЛАКСАЦИОННОГО ПО РАССТОЯНИЮ ДО ЭКСТРЕМУМА

Представлен субградиентный метод минимизации — аналог метода минимальных итераций решения систем уравнений, наследующий от последнего свойства сходимости на квадратичных функциях. Предложенный алгоритм при определенном наборе параметров совпадает с известным ранее методом минимизации кусочно-линейных функций и является элементом разработанного Б.Т. Поляком семейства релаксационных по расстоянию до экстремума методов минимизации, длина шага которых вычисляется на основе заданного значения минимума функции. Параметры метода увязаны с ограничением на степень однородности функции, получены оценки его скорости сходимости на выпуклых функциях. Доказано, что на некоторых классах функций он сходится со скоростью геометрической прогрессии. Обсуждаются вычислительные возможности метода при решении задач высокой размерности.

Ключевые слова: субградиент, выпуклая функция, линейная алгебра, минимум функции, скорость сходимости.

DOI: 10.1134/S0005231019010094

1. Введение

Рассматривается задача минимизации выпуклой, но не обязательно дифференцируемой функции $f(x)$, $x \in R^n$, где R^n — конечномерное евклидово пространство. Основная цель работы заключается в том, чтобы представить релаксационный по расстоянию до экстремума субградиентный метод минимизации — аналог метода минимальных итераций, наследующий от последнего свойства сходимости на квадратичных функциях, сформулировать и обосновать ограничения на выбор параметров метода в зависимости от свойств функции, получить оценки его скорости сходимости на выпуклых функциях и привести некоторые обнадеживающие результаты вычислительных экспериментов относительно эффективности метода при решении задач высокой размерности.

Активные исследования в области субградиентных методов начались после появления работ [1, 2]. В области построения методов, релаксационных по функции, метод сопряженных градиентов для решения систем уравнений был распространен на решение задач гладкой (см., например, [3]) и негладкой оптимизации [4, 5]. Существенного повышения эффективности удалось достичь в результате создания методов негладкой оптимизации с изменением метрики пространства [6–9]. Совершенствование нересурсоемких релаксационных субградиентных методов отражено в [10–12].

Одно из активно развиваемых направлений работ по созданию методов негладкой оптимизации основано на построении и использовании приближений функции [13–16]. В [14–16] рассматривается подход создания гладких приближений для негладких функций. На этой основе удалось получить и теоретически обосновать широкий спектр методов [14–16], предназначенных для решения выпуклых задач оптимизации, задач композитной и стохастической композитной оптимизации. Методы этого класса применимы для широкого круга (по величине размерности и условиям гладкости) задач. Алгоритмы для решения сверхбольших задач оптимизации предложены в [17].

В работах Б.Т. Поляка [3, 18] предложено и исследовано семейство релаксационных по расстоянию до экстремума методов минимизации. В алгоритмах этого типа длина шага вычисляется (а не находится подбором) на основании известного минимального значения функции $f^* = \inf_{x \in R^n} f(x)$. Простейший из них следующий [3, 18]:

$$(1) \quad x_{k+1} = x_k - \frac{\gamma(f(x_k) - f^*)}{(g_k, g_k)} g_k, \quad g_k \in \partial f(x_k), \quad k = 0, 1, 2, \dots$$

Здесь и далее $g, g(x)$ — некоторый субградиент из субградиентного множества $\partial f(x)$ функции $f(x)$, $g_k = g(x_k)$, γ — шаговый множитель, определяемый свойствами минимизируемой функции $f(x)$, (x, y) — скалярное произведение векторов. Согласно [3, 18] метод (1) является релаксационным по расстоянию до экстремума. В [19] показано, что процесс (1) при $\gamma = 2$ является аналогом метода минимальных ошибок для решения систем линейных уравнений [20, 21], в котором при минимизации квадратичных функций выполняется операция минимизации расстояния до экстремума вдоль направления антиградиента. В [17] алгоритм (1) используется как один из представителей простейших субградиентных методов, встроенных в схему решения сверхбольших задач оптимизации [17].

В [22] для минимизации кусочно-линейных функций предложен метод — элемент отмеченного выше семейства [18], в котором направление движения формируется на основе текущего субградиента и предыдущего направления, а длина шага вычисляется на основе известного значения минимума функции, аналогично тому, как это делается в (1):

$$(2) \quad x_{k+1} = x_k - \gamma_k p_k, \quad \gamma_k = \gamma \frac{(f(x_k) - f^*)}{(p_k, p_k)}, \quad k = 0, 1, 2, \dots,$$

$$(3) \quad p_0 = g_0, \quad p_k = g_k + \beta_k p_{k-1},$$

$$(4) \quad \beta_k = \begin{cases} -\alpha_k \frac{(g_k, p_{k-1})}{(p_{k-1}, p_{k-1})} & \text{при } (g_k, p_{k-1}) < 0, \quad 0 \leq \alpha_k \leq 2; \\ 0 & \text{в противном случае,} \end{cases}$$

где наложено ограничение $0 < \gamma \leq 1$. Идея вычисления текущего направления движения на основе предыдущего множества субградиентов [18] здесь воплощена в виде использования аккумулированного вектора p_k . Ряд результатов о качестве направления движения в методе (2)–(4), полученных в [22] для кусочно-линейных функций, обоснован для выпуклых функций в [7].

В настоящей работе метод AA^* -минимальных итераций (см., например, [20, 21]) распространен на задачу минимизации. Полученный метод эквивалентен алгоритму (2)–(4) при $\alpha_k = 1$ и $\gamma = 2$ и наследует свойство конечной сходимости метода минимальных итераций на квадратичных функциях. На выпуклых функциях получена оценка скорости сходимости метода, где ограничения на выбор параметра γ в (2) увязаны с ограничением на степень однородности функции аналогично тому, как это сделано в [20] для процесса (1). Доказано, что на некоторых классах выпуклых функций метод имеет скорость сходимости геометрической прогрессии. Простота реализации, отсутствие необходимости выполнения вспомогательных операций минимизации и незначительные вычислительные затраты на итерации с учетом представленных результатов вычислительного эксперимента позволяют надеяться на эффективность метода при решении задач высокой размерности.

2. Аналог метода минимальных итераций в оптимизации

Покажем, что в случае специального вида квадратичной функции метод (2)–(4) при $\alpha_k = 1$ и $\gamma = 2$ является аналогом метода AA^* -минимальных итераций (см., например, [20, 21]). Для этого необходимо перейти от записи метода минимальных итераций в терминах линейной алгебры к записи в форме метода минимизации, использующего только характеристики функции (значения функции и градиента).

Используем способ вывода метода AA^* -минимальных итераций [20, с. 481; 21, раздел 8.36], описанный в [20, с. 481], для получения метода (2)–(4) при $\alpha_k = 1$ и $\gamma = 2$. Пусть требуется решить систему уравнений

$$(5) \quad Bx = b, \quad x, b \in R^n, \quad B \in R^{n \times n},$$

где B — невырожденная матрица. Образует вспомогательную систему уравнений посредством второй трансформации Гаусса, где после замены $BB^T y = x$ получим

$$(6) \quad BB^T y = b.$$

Обозначим:

$$A = BB^T, \quad r(x) = Bx - b, \quad r_k = Bx_k - b.$$

Для решения системы (6) используем метод сопряженных градиентов:

$$(7) \quad y_{k+1} = y_k - \gamma_k s_k, \quad \gamma_k = \frac{(r_k, r_k)}{(s_k, As_k)}, \quad k = 0, 1, 2, \dots,$$

$$(8) \quad s_k = r_k + b_k s_{k-1}, \quad b_k = -\frac{(r_k, As_{k-1})}{(p_k, p_k)} > 0.$$

Отметим, что в (7)–(8) использованы только те выражения для γ_k и b_k , которые приводят к схеме (2)–(4). Для того чтобы получить запись метода

минимальных итераций в форме (2)–(4), введем квадратичную функцию

$$(9) \quad f(x) = \frac{(Bx - b, Bx - b)}{2} + f^* = \frac{(r, r)}{2} + f^*,$$

где f^* — минимальное значение функции (9). Точка минимума x^* функции (9) совпадает с решением системы (5). Градиент функции задается выражением $g(x) = \nabla f(x) = B^T(Bx - b) = B^T r$.

Обозначим

$$B^T s_k = p_k.$$

Используя равенства $A = BB^T$, $(r_k, r_k) = 2(f(x_k) - f^*)$,

$$\begin{aligned} (As_k, s_k) &= (BB^T s_k, s_k) = (B^T s_k, B^T s_k) = (p_k, p_k), \\ (As_{k-1}, r_k) &= (BB^T s_{k-1}, r_k) = (B^T s_{k-1}, B^T r_k) = (p_{k-1}, g_k), \end{aligned}$$

умножая на B^T векторные равенства в (7), (8), с учетом $B^T y = x$ получим расчетные формулы, аналогичные (2)–(4) при $\alpha_k = 1$ и $\gamma = 2$:

$$(10) \quad x_{k+1} = x_k - \gamma_k p_k,$$

$$(11) \quad p_k = g_k + b_k p_{k-1},$$

$$(12) \quad \gamma_k = \frac{(r_k, r_k)}{(s_k, As_k)} = \frac{2(f(x_k) - f^*)}{(p_k, p_k)},$$

$$(13) \quad b_k = -\frac{(r_k, As_{k-1})}{(As_{k-1}, s_{k-1})} = -\frac{(g_k, p_{k-1})}{(p_{k-1}, p_{k-1})} > 0.$$

Поскольку для решения системы (6) используется метод сопряженных градиентов, минимум специальной функции (9) будет найден процессом (10)–(13) за конечное число шагов, не превосходящее размерности пространства n . Таким образом, для специального вида квадратичной функции удалось сформулировать метод минимизации, использующий в своей записи только выражения функции и градиента. Относительно сходимости метода на квадратичных функциях справедливо следующее.

Теорема 1. Минимум квадратичной функции со строго положительно определенной матрицей Гессе будет найден процессом (2)–(4) при $\alpha_k = 1$ и $\gamma = 2$ за конечное число шагов, не превосходящее размерности пространства n .

Доказательство. Квадратичная функция с невырожденной матрицей вторых производных A имеет вид

$$(14) \quad f(x) = (x, Ax)/2 - (b, x) + c.$$

Ее точка минимума равна $x^* = A^{-1}b$, а значение минимума — $f^* = c - (A^{-1}b, b)/2$. Обозначим: $d = B^{-1}b$, где B — матрица такая, что $B = B^T$,

$BB = A$. Тогда несложно установить, что функцию (14) можно представить в виде, аналогичном (9):

$$f(x) = (Bx - d, Bx - d)/2 + f^*.$$

Точка минимума последней $f(x)$ совпадает с решением системы $Bx = d$, которое согласно показанному ранее может быть найдено методом (2)–(4) при $\alpha_k = 1$ и $\gamma = 2$, использующим только характеристики функции. Теорема доказана.

3. Оценки скорости сходимости

Пусть $f(x)$ — выпуклая функция на R^n , а ее множество точек минимума X^* не пусто. В предыдущем разделе метод (2)–(4) при $\gamma = 2$ рассмотрен на квадратичной функции, степень однородности которой $\gamma_f = 2$. В [3] выбор γ метода (1) увязывается со степенью однородности функции. При этом показано, что для анализа сходимости метода (1) вместо условия однородности можно использовать неравенство [3]

$$(15) \quad \gamma_f(f(x) - f^*) \leq (g(x), x - x^*), \quad \forall x \in R^n, \quad \forall x^* \in X^*, \quad \gamma_f \geq 1.$$

Как увидим далее, метод (2)–(4) сравнительно с (1) более требователен к выбору шага γ . При условии (15) оценки скорости сходимости метода (2)–(4) удастся получить на интервале $0 < \gamma \leq \gamma_f$, что в 2 раза меньше интервала релаксационности по расстоянию для метода (1). В силу отмеченного ограничения на интервал для γ и необходимости анализа скорости сходимости метода, например на квадратичных функциях, на класс функций накладывалось условие (15).

В леммах 1 и 2 изложены предварительные результаты, где ограничение $0 < \gamma \leq \gamma_f$ существенно и позволяет получить их доказательство. В [7] подобные результаты получены для выпуклых функций, т.е. при $\gamma_f = 1$ в (15).

Лемма 1. Пусть $f(x)$ — выпуклая функция на R^n , ее множество точек минимума X^ непусто и выполнено (15). Тогда для последовательности, генерируемой процессом (2)–(4) при $0 < \gamma \leq \gamma_f$, справедлива оценка*

$$(16) \quad (x_k - x^*, p_k) \geq (x_k - x^*, g_k), \quad k = 0, 1, 2, \dots$$

Доказательство. Отметим, что согласно (4) $\beta_k \geq 0$. Докажем (16) методом индукции по k . Допустим, что оценка справедлива для $k = m$. Докажем ее для $k = m + 1$.

$$\begin{aligned} (x_{m+1} - x^*, p_{m+1}) &= (x_{m+1} - x^*, g_{m+1}) + \beta_{m+1}(x_m - \gamma_m p_m - x^*, p_m) = \\ &= (x_{m+1} - x^*, g_{m+1}) + \beta_{m+1}[(x_m - x^*, p_m) - \gamma_m(p_m, p_m)]. \end{aligned}$$

Преобразуем выражение в квадратных скобках:

$$\begin{aligned} (x_m - x^*, p_m) - \gamma_m(p_m, p_m) &= (x_m - x^*, p_m) - \gamma(f(x_m) - f^*) \geq \\ &\geq (x_m - x^*, g_m) - \gamma(f(x_m) - f^*) \geq (\gamma_f - \gamma)(f(x_m) - f^*) \geq 0. \end{aligned}$$

С учетом $\beta_k \geq 0$ и неравенства $0 < \gamma \leq \gamma_f$ преобразуем предыдущее равенство:

$$(x_{m+1} - x^*, p_{m+1}) \geq (x_{m+1} - x^*, g_{m+1}).$$

Последнее доказывает лемму.

Отметим, что неравенство (16) играет важную роль в обосновании сходимости метода. В силу ограничения леммы $0 < \gamma \leq \gamma_f$ при использовании $\gamma_f = 1$ не удалось бы получить доказательство сходимости метода на некоторых классах функций, где более эффективным является выбор параметра $\gamma > 1$.

Лемма 2. Пусть $f(x)$ — выпуклая функция на R^n , ее множество точек минимума X^* непусто и выполнено (15). Тогда для последовательности, генерируемой процессом (2)–(4) при $0 < \gamma \leq \gamma_f$, направление p_k сравнительно с g_k образует меньший угол с направлением $x_k - x^*$.

Доказательство. Для метода (2)–(4) справедлива оценка

$$(17) \quad \|p_k\| \leq \|g_k\|.$$

Действительно, при $\beta_k = 0$ имеем $(p_k, p_k) = (g_k, g_k)$. В противном случае с учетом ограничений на α_k в (4) получим:

$$\|p_k\|^2 = \left\| g_k - \alpha_k \frac{(g_k, p_{k-1})}{(p_{k-1}, p_{k-1})} p_{k-1} \right\|^2 = \|g_k\|^2 - (2 - \alpha_k) \frac{(g_k, p_{k-1})^2}{(p_{k-1}, p_{k-1})} \leq \|g_k\|^2.$$

Из (16) и (17) следует $\frac{(x_k - x^*, p_k)}{\|p_k\|} \geq \frac{(x_k - x^*, g_k)}{\|g_k\|}$. После деления последнего равенства на $\|x_k - x^*\|$ получим неравенство для косинусов соответствующих углов, которое доказывает утверждение леммы.

Обозначим достигнутое на множестве итераций процесса (2)–(4) «рекордное» значение функции $\varphi_k = \min_{i=0, \dots, k} f(x_i)$. В следующей теореме дана оценка скорости сходимости для последовательности «рекордных» значений φ_k .

Теорема 2. Пусть $f(x)$ — выпуклая функция на R^n , ее множество точек минимума X^* непусто и выполнено (15). Тогда для последовательности $x_k, k = 0, 1, 2, \dots$, генерируемой процессом (2)–(4) при $0 < \gamma \leq \gamma_f$, имеет место $x_k \rightarrow x^* \in X^*$, а для последовательности «рекордных» значений φ_k справедлива оценка

$$(18) \quad \varphi_k - f^* \leq \frac{c \|x_0 - x^*\|}{\sqrt{(k+1)\gamma(2\gamma_f - \gamma)}},$$

где c — константа, такая что $\|g(x)\| \leq c$ на множестве $\{x : \|x - x^*\| \leq \|x_0 - x^*\|\}$.

Доказательство. Пусть $y \in X^*$ — произвольная точка минимума функции $f(x)$. Тогда для процесса (2)–(4) получим:

$$\begin{aligned} \|x_{k+1} - y\|^2 &= \left\| x_k - y - \gamma \frac{(f(x_k) - f^*)}{(p_k, p_k)} (p_k) \right\|^2 = \\ &= \|x_k - y\|^2 - \gamma \frac{(f(x_k) - f^*)}{(p_k, p_k)} (x_k - y, p_k) + \gamma^2 \frac{(f(x_k) - f^*)^2}{(p_k, p_k)}. \end{aligned}$$

Отсюда с учетом (16) и (15) следует

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - \gamma(2\gamma_f - \gamma) \frac{(f(x_k) - f^*)^2}{(p_k, p_k)}.$$

Преобразуем последнее неравенство с учетом (17):

$$(19) \quad \|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - \gamma(2\gamma_f - \gamma) \frac{(f(x_k) - f^*)^2}{\|g(x_k)\|^2}.$$

Отсюда следует $(f(x_k) - f^*)/\|g(x_k)\| \rightarrow 0$. Так как последовательность x_k ограничена: $\|x_k - y\| \leq \|x_0 - y\|$, то $\|g(x_k)\| \leq c$ [3, лемма 8, § 1, гл. 5]. Поэтому $f(x_k) \rightarrow f^*$. Следовательно, имеется подпоследовательность $x_{k_i} \rightarrow x^*$, где x^* — некоторая точка минимума. Заменяя в (19) y на x^* , получим, что $\|x_k - x^*\|$ монотонно убывает, а $\|x_{k_i} - x^*\| \rightarrow 0$. Отсюда следует $x_k \rightarrow x^*$.

С учетом сказанного преобразуем (19):

$$\|x_{k+1} - x^*\|^2 \leq \|x_0 - x^*\|^2 - \frac{\gamma(2\gamma_f - \gamma)}{c^2} \sum_{i=0}^k (f(x_i) - f^*)^2.$$

Отсюда получим:

$$\frac{\gamma(2\gamma_f - \gamma)}{c^2} \sum_{i=0}^k (f(x_i) - f^*)^2 \leq \|x_0 - x^*\|^2.$$

Последнее неравенство приведем к следующему виду:

$$(20) \quad \frac{1}{k+1} \sum_{i=0}^k (f(x_i) - f^*)^2 \leq \frac{c^2 \|x_0 - x^*\|^2}{(k+1)\gamma(2\gamma_f - \gamma)}.$$

В левой части неравенства находится среднее арифметическое слагаемых. Следовательно, хотя бы одно из этих слагаемых с индексом i_k должно быть не больше среднего, что с учетом определения φ_k приводит к неравенству

$$(\varphi_k - f^*)^2 \leq (f(x_{i_k}) - f^*)^2 \leq \frac{c^2 \|x_0 - x^*\|^2}{(k+1)\gamma(2\gamma_f - \gamma)}.$$

Теорема доказана.

В случае острого минимума, определяемого соотношением

$$(21) \quad f(x) - f^* \geq \rho \|x - x^*\|, \quad \rho > 0,$$

справедлив результат, приводимый ниже.

Теорема 3. Пусть $f(x)$ — выпуклая на R^n функция, имеющая острый минимум. Тогда для последовательности точек x_k , $k = 0, 1, 2, \dots$, генерируемых процессом (2)–(4) при $0 < \gamma \leq 1$, имеет место оценка скорости сходимости геометрической прогрессии

$$\|x_{k+1} - x^*\|^2 \leq q \|x_k - x^*\|^2, \quad q = 1 - \gamma(2 - \gamma)\rho^2/c^2,$$

где c — константа, такая что $\|g(x)\| \leq c$ на множестве $\{x : \|x - x^*\| \leq \|x_0 - x^*\|\}$.

Доказательство. Преобразуем (19) с учетом (15) при $\gamma_f = 1$:

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 - \gamma(2 - \gamma)\rho^2 \|x_k - y\|^2/c^2 = q \|x_k - y\|^2,$$

что доказывает утверждение теоремы.

Теорема 4. Пусть $f(x)$ — сильно выпуклая на R^n с константой m , дифференцируемая функция, ее градиент удовлетворяет условию Липшица с константой M на множестве $S = \{x : \|x - x^*\| \leq \|x_0 - x^*\|\}$ и выполнено (15). Тогда для последовательности, генерируемой процессом (2)–(4) при $0 < \gamma \leq \gamma_f$, имеет место оценка скорости сходимости геометрической прогрессии

$$\|x_{k+1} - x^*\|^2 \leq q \|x_k - x^*\|^2, \quad q = 1 - \frac{m^2\gamma(2\gamma_f - \gamma)}{M^2}.$$

Доказательство. Выполнены все условия теоремы 2, поэтому справедлива оценка (19) при $y = x^*$. Из сильной выпуклости функции следует [3] $f(x_k) - f^* \geq m \|x_k - x^*\|^2$, а из условия Липшица имеем $\|g(x_k)\| \leq M \|x_k - x^*\|$. Объединяя эти оценки, получим

$$\|x_{k+1} - x^*\|^2 \leq \|x_k - x^*\|^2 - \frac{m^2\gamma(2\gamma_f - \gamma)}{M^2} \|x_k - x^*\|^2,$$

что доказывает утверждение теоремы.

4. Результаты численного исследования

Целью является изучение возможностей метода (2)–(4) (АММИ) на основе анализа и сравнения его результатов численного исследования с результатами для известных методов оптимизации. В методе АММИ задавались α_k в (4), $\gamma = \mu\gamma_f$ в (2) (параметр γ_f соответствует минимальной предполагаемой степени однородности функции) и предельное число итераций KO , протекающих при $\beta_k \neq 0$ в (3), после которого производилось обновление $\beta_k = 0$, т.е. набор $(\alpha_k, \mu\gamma_f, KO)$, например $(\alpha_k = 1,02, \gamma = 1,01\gamma_f, KO = \infty)$. Во всех методах функция и субградиент в точке вычислялись одновременно. Будем

обозначать через it число итераций метода, а через nfg количество вычислений функции и градиента, требуемые для достижения заданной точности

$$(22) \quad f_k - f^* \leq \varepsilon.$$

В табл. 1–3 приведены результаты для методов на функциях:

$$f_1(x) = \sum_{i=1}^n x_i^2 i^4, \quad x_{0,i} = 10, \quad x_i^* = 0, \quad i = 0, 1, 2, \dots, n,$$

$$f_2(x) = \sum_{i=1}^n |x_i| i, \quad x_{0,i} = 1, \quad x_i^* = 0, \quad i = 0, 1, 2, \dots, n,$$

$$f_3(x) = \sum_{i=1}^n x_i^2 (1 + (i-1)(100-1)/(n-1))^2,$$

$$x_{0,i} = 1, \quad x_i^* = 0, \quad i = 0, 1, 2, \dots, n,$$

$$f_4(x) = \sum_{i=1}^n |x_i| (1 + (i-1)(100-1)/(n-1)),$$

$$x_{0,i} = 1, \quad x_i^* = 0, \quad i = 0, 1, 2, \dots, n.$$

В качестве методов для сравнения использовались равноценные по затратам памяти и вычислений на итерации метод сопряженных градиентов (МСГ) и релаксационный субградиентный метод МССГ [12].

В табл. 1 приведено число итераций для методов АММИ, МСГ и МССГ на первых двух функциях, необходимое для выполнения неравенства (22). Функции характеризуются увеличением степени вытянутости линий уровня с ростом размерности. Коэффициенты квадратичной функции выбраны так, чтобы исчез эффект конечного окончания процесса для методов АММИ, МСГ. Для методов МСГ и МССГ затраты nfg примерно в 2 раза больше числа итераций. Для АММИ на $f_1(x)$ заданы параметры ($\alpha_k = 1,0$, $\gamma = 1,0\gamma_f$, $KO = \infty$), а на $f_2(x)$ — ($\alpha_k = 1,02$, $\gamma = 1,01\gamma_f$, $KO = 10000$).

На квадратичных функциях менее сложных, чем $f_1(x)$, методы АММИ и МСГ достигают необходимой точности за конечное число шагов, что свидетельствует о наличии у метода АММИ конечной сходимости. Результаты на $f_1(x)$ устанавливают, что для метода АММИ эффект потери скорости сходимости из-за ошибок округления при увеличении степени вырожденности функции проявляется в меньшей степени по сравнению с МСГ.

Результаты для $f_2(x)$ позволяют судить о возможностях метода на негладких функциях при различных масштабах переменных по осям координат и дают некоторое представление о границах возможностей метода. В данном случае масштабы переменных функции $f_2(x)$ имеют разброс в n раз меньший по сравнению с масштабами квадратичной функции $f_1(x)$. Для сравнения, методу на основе (1) при $n = 10$ и $n = 100$ требуется соответственно 2318 и 2046203 итераций для достижения заданной точности, что свидетельствует о наличии ускоряющих свойств у метода АММИ.

Таблица 1. Количество итераций

n		10	50	100	300	500	1000
$f_1(x)$ ($\varepsilon = 10^{-10}$)	МСГ	12	178	777	8496	25613	129356
$f_1(x)$ ($\varepsilon = 10^{-10}$)	АММИ	12	165	617	5682	17713	76204
$f_2(x)$ ($\varepsilon = 10^{-5}$)	АММИ	50	507	1948	6726	23970	23823
$f_2(x)$ ($\varepsilon = 10^{-5}$)	МССГ	1084	2529	4535	15333	32245	42977

Таблица 2. Количество вычислений функции и градиента при $\varepsilon = 10^{-8}$

n			5000	10000	25000	50000	100000	500000	1000000
$f_3(x)$	АММИ	$it = nfg$	642	658	679	696	713	753	771
$f_3(x)$	МССГ	it	656	672	695	711	728	767	783
		nfg	1036	1064	1109	1134	1196	1328	1388
$f_3(x)$	МСГ	it	641	657	679	696	713	753	770
		nfg	1287	1321	1368	1404	1441	1533	1574

Таблица 3. Количество вычислений градиента и функции при $\varepsilon = 10^{-4}$

n			5000	10000	25000	50000	100000	500000	1000000
$f_4(x)$	АММИ ₁	$it = nfg$	9166	15885	15033	14739	24563	41528	43054
$f_4(x)$	АММИ ₂	$it = nfg$	11830	10290	13349	19104	15202	26614	28834
$f_4(x)$	МССГ	it	52200	53329	54978	56051	57178	59981	61154
		nfg	104418	106682	109988	112145	114414	120062	122416

Число обусловленности функции $f_3(x)$ равно 10^{-4} и не зависит от размерности. Функция $f_4(x)$ имеет равный характер вытянутости линий уровня по осям координат сравнительно с $f_3(x)$. Тесты на этих функциях проведены с целью выявления характера скорости сходимости в зависимости от размерности при не зависящем от размерности характере вытянутости линий уровня по осям координат. Для АММИ на $f_3(x)$ заданы параметры ($\alpha_k = 1,02, \gamma = 1,0\gamma_f, KO = 1000$), см. табл. 2.

Число итераций метода МССГ несколько выше, чем у метода сопряженных градиентов, но общие затраты nfg у него в силу грубого одномерного поиска ниже. Таким образом, на квадратичных функциях при невысокой степени обусловленности метод АММИ равноценен по числу итераций методам МСГ и МССГ, но имеет более низкие затраты по числу обращений к функции и градиенту.

Результаты для функции $f_4(x)$ приведены в табл. 3. Точность минимизации функции выбиралась из расчета исключения возможности окончания вычислений вдалеке от точки минимума за счет получения нужного значения в подпространстве малой размерности. Методом АММИ сделано 2 варианта счета с параметрами ($\alpha_k = 1,0, \gamma = 1,0\gamma_f, KO = 500$) для АММИ₁ и ($\alpha_k = 1,02, \gamma = 1,01\gamma_f, KO = 1000$) для АММИ₂. Функция $f_4(x)$ имеет острый минимум. Число итераций, затраченных методами на данной функции, существенно ниже оценок, которые можно получить на основании теоремы 3. Согласно результатам табл. 3 число итераций методов АММИ и МССГ толь-

ко при росте размерности становится соизмеримым, но общие затраты обращений к алгоритму вычислений функции и градиента разнятся значительно.

Таким образом, на большеразмерных квадратичных тестах выявлено, что эффект потери скорости сходимости из-за ошибок округления при увеличении степени вырожденности функции у представленного алгоритма проявляется в меньшей степени по сравнению с методом сопряженных градиентов. Результаты решения негладких задач высокой размерности позволяют сделать вывод о стабильной работе метода и малой зависимости свойств его скорости сходимости от размерности при условии равного характера вытянутости линий уровня функций.

5. Заключение

Предложенный в статье субградиентный метод минимизации, полученный распространением метода минимальных итераций на решение задач минимизации, пополняет семейство релаксационных по расстоянию до экстремума методов [18], длина шага которых вычисляется на основе известного значения минимума функции, и наследует свойство конечной сходимости последнего на квадратичных функциях. Учет ограничения на степень однородности функции [3] позволил сформулировать границы его параметров. Предложенный алгоритм при определенном наборе параметров совпадает с известным ранее методом минимизации кусочно-линейных функций [22].

В статье на классе выпуклых функций получена оценка скорости сходимости метода, совпадающая с точностью до мультипликативной константы с соответствующей нижней оценкой для этого класса задач [6]. Показано, что на сильно выпуклых функциях с липшицевым градиентом [3] и функциях с острым минимумом [3] метод сходится линейно. Проведенный вычислительный эксперимент свидетельствует, что при умеренной степени вырожденности метод в меньшей мере подвержен проявлению неустойчивости из-за ошибок округления, нежели метод сопряженных градиентов, а на негладких задачах высокой размерности показывает устойчивые результаты. В силу отсутствия вспомогательных операций минимизации и незначительных вычислительных затрат на итерации метод может оказаться предпочтительным при решении определенных прикладных задач высокой размерности.

СПИСОК ЛИТЕРАТУРЫ

1. *Шор Н.З.* Применение метода градиентного спуска для решения сетевой транспортной задачи // Матер. научн. семинара по теорет. и прикл. вопр. кибернетики и исследования операций. Киев: Науч. совет по кибернетике АН УССР. 1962. Вып. 1. С. 9–17.
2. *Поляк Б.Т.* Один общий метод решения экстремальных задач // Докл. АН СССР. 1967. Т. 174. Вып. 1. С. 33–36.
3. *Поляк Б.Т.* Введение в оптимизацию. М.: Наука, 1983.
4. *Wolfe P.* Note on a method of conjugate subgradients for minimizing nondifferentiable functions // Math. Program. 1974. V. 7. No. 1. P. 380–383.
5. *Lemarechal C.* An extension of Davidon methods to non-differentiable problems // Math. Program. Study. 1975. V. 3. P. 95–109.

6. *Немировский А.С., Юдин Д.Б.* Сложность задач и эффективность методов оптимизации. М.: Наука, 1979.
7. *Шор Н.З.* Методы минимизации недифференцируемых функций и их приложения. Киев: Наук. думка, 1979.
8. *Крутиков В.Н., Петрова Т.В.* Релаксационный метод минимизации с растяжением пространства в направлении субградиента // Экономика и мат. методы. 2003. Т. 39. Вып. 1. С. 106–119.
9. *Крутиков В.Н., Горская Т.А.* Семейство релаксационных субградиентных методов с двухранговой коррекцией матриц метрики // Экономика и мат. методы. 2009. Т. 45. Вып. 4. С. 37–80.
10. *Нурминский Е.А., Тьен Д.* Метод сопряженных субградиентов с ограниченной памятью // АиТ. 2014. № 4. С. 67–80.
Nurminskii E.A., Tien D. Method of conjugate subgradients with constrained memory // Autom. Remote Control. 2014. V. 75. No. 4. P. 646–656.
11. *Крутиков В.Н., Вершинин Я.Н.* Многошаговый субградиентный метод для решения негладких задач минимизации высокой размерности // Вестн. Том. гос. ун-та. Математика и механика. 2014. № 3. С. 5–19.
12. *Крутиков В.Н., Вершинин Я.Н.* Субградиентный метод минимизации с коррекцией векторов спуска на основе пар обучающих соотношений // Вестн. КемГУ. 2014. № 1–1 (57). С. 46–54.
13. *Гольштейн Е.Г., Немировский А.С., Нестеров Ю.Е.* Метод уровней, его обобщения и приложения // Экономика и мат. методы. 1995. Т. 31. Вып. 3. С. 164–180.
14. *Nesterov Yu.E.* Smooth minimization of non-smooth functions // Math. Program. 2005. V. 103. No. 1. P. 127–152.
15. *Nesterov Yu.* Universal gradient methods for convex optimization problems // Math. Program. Ser. A. 2015. V. 152. P. 381–404.
16. *Гасников А.В., Нестеров Ю.Е.* Универсальный метод для задач стохастической композитной оптимизации // e-print. 2016.
<https://arxiv.org/ftp/arxiv/papers/1604/1604.05275.pdf>
17. *Nesterov Yu.* Subgradient methods for huge-scale optimization problems // Math. Program. Ser. A. 2013. V. 146. No. 1–2. P. 275–297.
18. *Поляк Б.Т.* Минимизация негладких функционалов // Журн. вычисл. мат. и матем. физики. 1969. Т. 9. № 3. С. 507–521.
19. *Самойленко Н.С., Крутиков В.Н., Мешечкин В.В.* Исследование одного варианта субградиентного метода // Вестн. КемГУ. 2015. № 5 (2). С. 55–58.
20. *Фадеев Д.К., Фадеева Д.К.* Вычислительные методы линейной алгебры. М.: Физматгиз, 1963.
21. *Воеводин В.В., Кузнецов Ю.А.* Матрицы и вычисления. М.: Наука, 1984.
22. *Camerini P., Fratta L., Maffioli F.* On improving relaxation methods by modified gradient techniques // Math. Program. 1975. No. 3. P. 26–34.

Статья представлена к публикации членом редколлегии Б.Т. Поляком.

Поступила в редакцию 15.02.2017

После доработки 15.03.2018

Принята к публикации 08.11.2018