

МАТЕМАТИКА

УДК 519.6

В.Н. Крутиков, Я.Н. Вершинин

МНОГОШАГОВЫЙ СУБГРАДИЕНТНЫЙ МЕТОД ДЛЯ РЕШЕНИЯ НЕГЛАДКИХ ЗАДАЧ МИНИМИЗАЦИИ ВЫСОКОЙ РАЗМЕРНОСТИ

Предложен многошаговый субградиентный метод для решения негладких задач минимизации высокой размерности и доказана его сходимость. По затратам памяти на хранение информации алгоритм сходен с методами сопряженных градиентов. В алгоритме используется новый метод решения неравенств, основанный на последовательной ортогонализации векторов обучения. Результаты численного исследования свидетельствуют о высокой скорости сходимости разработанного метода минимизации на негладких задачах высокой размерности.

Ключевые слова: алгоритм Качмажа, многошаговый алгоритм, метод минимизации, скорость сходимости.

1. Введение

Излагаемый в работе многошаговый релаксационный субградиентный метод минимизации (PCM), основанный на принципах организации методов «сопряженных субградиентов» [1, 2], принадлежит классу релаксационных методов ε -субградиентного типа (PCM) [1, 2] и предназначен для решения задач высокой размерности. Имеющиеся на настоящий момент PCM с растяжением пространства [4–8] соизмеримы по скорости сходимости на гладких функциях с квазиньютоновскими методами [6, 8] и эффективны при решении негладких задач овражного типа [6, 8]. В силу необходимости хранения и преобразования матрицы их эффективность по затратам времени резко снижается на задачах высокой размерности. Существующие многошаговые PCM [1, 3] существенно уступают в скорости сходимости субградиентным методам с растяжением пространства, подвержены циклированию на овражных задачах негладкой оптимизации, что определяет актуальность их совершенствования.

Пусть решается задача минимизации выпуклой на $\mathbf{R}^n$ функции $f(x)$. В PCM последовательные приближения строятся по формулам

$$x_{k+1} = x_k - \gamma_k s_k, \quad \gamma_k = \arg \min_{\gamma \in \mathbf{R}} f(x_k - \gamma s_k),$$

где направление спуска s_k выбирается как решение неравенств [2]:

$$(s, g) > 0, \quad \forall g \in G. \quad (1)$$

Здесь множество $G = \partial_\varepsilon f(x_k)$ – ε -субградиентное множество в точке x_k . Обозначим $S(G)$ – множество решений (1), $\partial f(x) \equiv \partial f_0(x)$ – субградиентное множе-

ство в точке x . В РСМ для решения систем неравенств (1) применяют итерационные методы (алгоритмы обучения), где в качестве элементов ε -субградиентных множеств, поскольку их явное задание отсутствует, используют субградиенты, вычисляемые на траектории спуска алгоритма минимизации.

В работах [3, 5–8] предложен и используется следующий подход сведения системы (1) к системе равенств. Пусть $G \subset \mathbf{R}^n$ принадлежит некоторой гиперплоскости, а его ближайший к началу координат вектор $\eta(G)$ является также и ближайшим к началу координат вектором гиперплоскости. Тогда решение системы $(s, g) = 1, \forall g \in G$, является также решением и для (1). Его можно найти как решение системы [5–8]

$$(s, g_i) = y_i, \quad i = 0, 1, \dots, k, \quad y_i \equiv 1. \quad (2)$$

В [3] (см. также [7, 8]) предложен метод минимизации, в котором для решения системы (2) используется алгоритм Качмажа [9] (см. также [10])

$$s_{k+1} = s_k + \frac{1 - (s_k, g_k)}{(g_k, g_k)} g_k. \quad (3)$$

Такой алгоритм минимизации при точном одномерном спуске на дифференцируемых функциях обладает свойствами метода сопряженных градиентов и эффективен при решении задач негладкой оптимизации [3, 7, 8].

В случае ортогональности векторов g_k метод (3) конечен [7, 8]. Алгоритм решения системы равенств с последовательной ортогонализацией векторов g_k предложен в [11]. В настоящей работе этот алгоритм распространен на решение неравенств и используется для поиска направления спуска в методе минимизации.

Основной целью построения направления s_k в субградиентных методах является поиск такого направления, которое обеспечивало бы возможность уменьшения функции из любой точки некоторой окрестности текущего приближения, т.е. решение системы неравенств (1), где множество G составлено из субградиентов окрестности текущего приближения x_k . Это означает возможность выхода из этой окрестности посредством минимизации функции вдоль этого направления. Чем шире окрестность, тем выше устойчивость метода к ошибкам округления, помехам, наличию малых локальных экстремумов и большее продвижение в направлении к экстремуму. В этой связи особую важность приобретают изучаемые в работе методы минимизации, в которых, в отличие от метода из [1] и его модификации из [2], встроенные алгоритмы решения систем неравенств используют субградиенты достаточно широкой окрестности текущего приближения минимума и не требуют точного одномерного спуска.

2. Многошаговый метод решения неравенств

В предлагаемом алгоритме строятся последовательные приближения решения системы (1).

Алгоритм А1.

1. Положить $k = 0$, $p_{k-1} = 0$ ($p_{k-1} \in \mathbf{R}^n$). Задать начальное приближение $s_0 \in \mathbf{R}^n$.

2. Выбрать произвольно $g_k \in G$, удовлетворяющий условию $(s_k, g_k) \leq 0$, если такого вектора не существует, то $s_k \in S(G)$, закончить работу алгоритма.

3. Получить новое приближение s_{k+1} :

$$s_{k+1} = s_k + \frac{1 - (s_k, g_k)}{(p_k, g_k)} p_k, \quad (4)$$

где
$$p_k = \begin{cases} g_k, & \text{если } (g_k, p_{k-1}) \geq 0, \\ g_k - \frac{(g_k, p_{k-1})}{\|p_{k-1}\|^2} p_{k-1}, & \text{если } (g_k, p_{k-1}) < 0. \end{cases} \quad (5)$$

4. Положить $k = k+1$. Перейти на пункт 2.

Алгоритм А1 отличается от алгоритма решения неравенств на основе алгоритма Качмажа [3, 7–8] (обозначим его А0) реализацией пункта 3. В А0 вместо (4), (5) используется формула (3).

На итерациях алгоритма А1 выполняются следующие соотношения:

$$\begin{aligned} & \text{а) } (p_k, p_{k-1}) = 0, \text{ если } (g_k, p_{k-1}) < 0; \\ & \text{б) } (p_k, g_k) \leq (g_k, g_k); \text{ в) } (p_k, p_k) = (p_k, g_k). \end{aligned} \quad (6)$$

При $k = 0$, поскольку в пункте 1 полагается $p_{k-1} = 0$, проводится преобразование (5а). В результате получим $p_0 = g_0$. В (5б) производится ортогонализация векторов p_k, p_{k-1} , что отражено в равенстве (6а). Непосредственно из (5) следует (6б). Из (5), с учетом (6а), получим (6с).

Обозначим $\eta_G \equiv \eta(G)$ – ближайший к началу координат вектор множества G , $\rho_G \equiv \rho(G) = \|\eta(G)\|$, $\mu_G = \eta(G) / \|\eta(G)\|$, $s^* = \mu_G / \rho_G$, $R_G \equiv R(G) = \max_{g \in G} \|g\|$, $v(G) = \rho_G / R_G$. Сделаем предположение относительно множества G .

Предположение 1. Множество G не пустое, выпуклое, замкнутое, ограниченное, $R_G < \infty$, и удовлетворяет условию делимости, то есть $\rho_G > 0$.

При этих условиях векторы μ_G и s^* являются решениями (1), а для векторов $g \in G$ выполняются ограничения [5–8]

$$1 \leq (s^*, g) \leq R_G / \rho_G, \quad \forall g \in G. \quad (7)$$

Мы изучим сходимость алгоритма А1 к решению s^* . Обозначим $\Delta_k = s_k - s^*$ – вектор невязки. Следующие результаты относительно алгоритма А1 получены в условиях справедливости предположения 1.

Лемма 1. Пусть последовательность $\{s_k\}$ получена в результате работы алгоритмом А1. Тогда для $k = 0, 1, 2, \dots$ имеют место оценки:

$$(\Delta_{k+1}, p_k) \leq 0; \quad (8)$$

$$(\Delta_k, p_k) \leq (\Delta_k, g_k). \quad (9)$$

Доказательство проведем по индукции. В силу равенства $p_k = g_k$ при $k = 0$ выполнено (9). Учитывая левое из неравенств (7) и условие пункта 2, $(s_k, g_k) \leq 0$, получим

$$-(\Delta_k, g_k) = (s^*, g_k) - (s_k, g_k) \geq 1 - (s_k, g_k) \geq 1. \quad (10)$$

Вычтем из обеих частей (4) s^* , умножим обе части равенства скалярно на p_k и

преобразуем правую часть с учетом (6с), (9) и (10):

$$(\Delta_{k+1}, p_k) = (\Delta_k, p_k) + \frac{1 - (s_k, g_k)}{(p_k, g_k)} (p_k, p_k) \leq (\Delta_k, p_k) - \frac{(\Delta_k, g_k)}{(p_k, g_k)} (p_k, g_k) \leq 0. \quad (11)$$

Здесь мы предположили, что последний переход в цепочке неравенств произведен при условии (9) для текущего k . Поскольку неравенство (9) выполняется при $k = 0$, то справедливо (11), откуда следует (8) при $k = 0$.

Предположим, что неравенства (8), (9) выполнены при $k = 0, 1, \dots, l-1$, где $l \geq 1$. Покажем, что они выполняются при $k = l$. В случае (5а) выполнено (9). Для доказательства (9) в случае (5б) умножим скалярно на Δ_k обе части равенства (5б). Отсюда, в силу справедливости (8) при $k = l-1$ и условия $(g_k, p_{k-1}) < 0$ из (5б), получим обоснование (9) при $k = l$

$$(\Delta_k, p_k) = (\Delta_k, g_k) - \frac{(g_k, p_{k-1})}{(p_{k-1}, p_{k-1})} (\Delta_k, p_{k-1}) \leq (\Delta_k, g_k).$$

Из (9) и (11) следует (8) при $k = l$. Лемма доказана.

В следующей теореме утверждается, что преобразование (5б) дает направление p_k на точку решения s^* с более острым углом по сравнению с g_k .

Теорема 1. Пусть последовательность $\{s_k\}$ получена в результате работы алгоритмом А1. Тогда для $k = 0, 1, 2, \dots$ имеет место оценка

$$\frac{(-\Delta_k, p_k)}{(p_k, p_k)^{0.5}} \geq \frac{(-\Delta_k, g_k)}{(g_k, g_k)^{0.5}} \geq \frac{1}{R_G}. \quad (12)$$

Доказательство. Из (9) и (10) следует

$$(-\Delta_k, p_k) \geq (-\Delta_k, g_k) \geq 1. \quad (13)$$

Отсюда, с учетом (6б), (6с), (9), (10) и определения величины R_G , имеем

$$\frac{(-\Delta_k, p_k)}{(p_k, p_k)^{0.5}} \geq \frac{(-\Delta_k, p_k)}{(g_k, g_k)^{0.5}} \geq \frac{(-\Delta_k, g_k)}{(g_k, g_k)^{0.5}} \geq \frac{1}{R_G}.$$

Теорема доказана.

Для обоснования сходимости алгоритма А1 нам потребуется следующий результат.

Лемма 2 [5, 7, 8]. Пусть множество G удовлетворяет предположению 1. Тогда $s_k \in S(G)$, если

$$\|\Delta_k\| < 1/R_G. \quad (14)$$

В следующей теореме обосновывается конечная сходимость алгоритма А1. Отметим, что полученные оценки полностью эквивалентны оценкам для алгоритма А0.

Теорема 2. Пусть множество G удовлетворяет предположению 1. Тогда для оценки скорости сходимости последовательности $\{s_k\}$ к точке s^* , генерируемой алгоритмом А1 до момента останова, справедливо соотношение

$$\|s_k - s^*\|^2 \leq (\|s_0\| + \rho_G^{-1})^2 - k/R_G^2, \quad (15)$$

для величины ρ_G^{-1} имеет место оценка

$$\rho_G^{-1} \geq \left(\sum_{j=0}^k (g_j, g_j)^{-1} \right)^{0,5} - \|s_0\| \geq \frac{k^{0,5}}{R_G} - \|s_0\|, \quad (16)$$

а при некотором значении k , удовлетворяющем неравенству

$$k \leq k^* \equiv R_G^2 (\|s_0\| + \rho_G^{-1})^2 + 1, \quad (17)$$

будет получен вектор $s_k \in S(G)$.

Доказательство. Найдем невязку Δ_{k+1} вычитанием s^* из обеих частей (4) и получим выражение квадрата ее нормы. Правую часть полученного выражения преобразуем с учетом неравенств (6b), (6c), (9), (10) и определения величины R_G :

$$\begin{aligned} (\Delta_{k+1}, \Delta_{k+1}) &= (\Delta_k, \Delta_k) + 2(\Delta_k, p_k) \frac{1 - (s_k, g_k)}{(p_k, g_k)} + (p_k, p_k) \frac{(1 - (s_k, g_k))^2}{(p_k, g_k)^2} \leq \\ &\leq (\Delta_k, \Delta_k) - 2 \frac{(1 - (s_k, g_k))^2}{(p_k, g_k)} + \frac{(1 - (s_k, g_k))^2}{(p_k, g_k)} \leq (\Delta_k, \Delta_k) - \frac{1}{(g_k, g_k)} \leq (\Delta_k, \Delta_k) - \frac{1}{R_G^2}. \end{aligned}$$

Отсюда, используя неравенство $\|s_0 - s^*\|^2 \leq (\|s_0\| + \|s^*\|)^2 = (\|s_0\| + \rho_G^{-1})^2$, которое следует из свойств нормы, получим оценки (15) и (16).

Согласно оценке (15), величина $\|\Delta_k\| \rightarrow 0$. Поэтому на некотором шаге k для вектора s_k будет выполнено неравенство (14), т.е. будет получен вектор $s_k \in S(G)$, являющийся решением системы (1). В качестве верхней оценки необходимого числа шагов можно взять k^* , равное значению k , при котором правая часть (15) обращается в нуль, увеличенному на 1. Это дает оценку (17). Теорема доказана.

3. Многошаговый субградиентный метод

Техника обоснования алгоритма соответствует [3, 5–8]. Пусть функция $f(x)$, $x \in \mathbf{R}^n$, выпукла. Обозначим $d(x) = \rho(\partial f(x))$, $D(z) = \{x \in \mathbf{R}^n \mid f(x) \leq f(z)\}$.

Примечание 1. Для выпуклой на $\mathbf{R}^n$ функции, при ограниченности множества $D(x_0)$ для точек $x^* \in D(x_0)$, удовлетворяющих условию $d(x^*) < d_0$, справедлива оценка [2, с. 291]

$$f(x^*) - f^* \leq D d_0, \quad (18)$$

где D – диаметр множества $D(x_0)$, $f^* = \inf_{x \in \mathbf{R}^n} f(x)$.

Дадим описание метода минимизации на основе алгоритма А1 для нахождения точек $x^* \in \mathbf{R}^n$, таких, что $d(x^*) \leq E_0$, где $E_0 > 0$.

Алгоритм М1.

1. Задать начальное приближение $x_0 \in \mathbf{R}^n$, целые $k = i = 0$.
2. Положить $k = k+1$, $q_k = i$, $s_i = 0$, $p_{i-1} = 0$, $\Sigma_i = 0$.
3. Задать ε_k, t_k .
4. Вычислить субградиент $g_i \in \partial f(x_i)$, удовлетворяющий условию $(s_i, g_i) \leq 0$, и если $g_i = 0$, то решение найдено, закончить работу алгоритма.

5. Получить новое приближение $s_{i+1} = s_i + \frac{1 - (s_i, g_i)}{(p_i, g_i)} p_i$, где

$$p_i = \begin{cases} g_i, & \text{если } (g_i, p_{i-1}) \geq 0, \\ g_i - \frac{(g_i, p_{i-1})}{\|p_{i-1}\|^2} p_{i-1}, & \text{если } (g_i, p_{i-1}) < 0. \end{cases}$$

6. Вычислить новое приближение критерия $\Sigma_{i+1} = \Sigma_i + (g_i, g_i)^{-1}$.

7. Вычислить новое приближение точки минимума:

$$x_{i+1} = x_i - \gamma_i s_{i+1}, \quad \gamma_i = \arg \min_{\gamma \in \mathbf{R}} f(x_i - \gamma_i s_{i+1}).$$

8. Положить $i = i + 1$.

9. Если $1/\sqrt{\Sigma_i} < \varepsilon_k$, то перейти на пункт 2.

10. Если $i - q_k > m_k$, то перейти на пункт 2.

11. Перейти на пункт 4.

Алгоритм М1 отличается от известного метода минимизации [3, 7, 8], основанного на алгоритме решения неравенств А0 с формулой Качмажа (3) (назовем его М0), реализацией пункта 5, где вместо формул (4), (5) используется преобразование (3). В алгоритм М1 в пунктах 2, 4, 5 встроены алгоритмы решения неравенств. Индекс q_k , $k = 0, 1, 2, \dots$, введен с целью обозначения номеров итераций i , при которых в пункте 2 при выполнении критериев пунктов 9, 10 происходит обновление для алгоритма решения неравенств ($p_{i-1} = 0$). Согласно (15), (17), алгоритм решения неравенств при $s_0 = 0$ имеет наилучшие оценки скорости сходимости. Поэтому при обновлении в пункте 2 алгоритма М1 задаем $s_i = 0$. Потребность в обновлении возникает вследствие того, что в результате смещения в пункте 7 происходит смена субградиентных множеств окрестности текущей точки достигнутого минимума, что приводит к необходимости решения системы неравенств на основе новой информации.

Отметим, что в силу условия точного одномерного спуска вдоль направления $(-s_{i+1})$ в пункте 7, в новой точке x_{i+1} вектор $g_{i+1} \in \partial f(x_{i+1})$, такой, что $(g_{i+1}, s_{i+1}) \leq 0$, всегда существует согласно необходимому условию минимума одномерной функции (см., например, [2, с. 287]). Следовательно, с учетом роста индекса i в пункте 8, условие $(g_i, s_i) \leq 0$ пункта 4 всегда удовлетворяется.

Доказательство сходимости метода М1 опирается на следующую лемму.

Лемма 3 [2]. Пусть функция $f(x)$ строго выпукла на $\mathbf{R}^n$, множество $D(x_0)$ ограничено, а последовательность $\{x_k\}_{k=0}^\infty$ такова, что

$$f(x_{k+1}) = \min_{\alpha \in [0,1]} f(x_k + \alpha(x_{k+1} - x_k)).$$

Тогда $\lim_{k \rightarrow \infty} \|x_{k+1} - x_k\| = 0$.

Обозначим $S_\varepsilon(G) = \{z \in \mathbf{R}^n \mid \|z - x\| \leq \varepsilon, \forall x \in G\}$ – ε -окрестность множества G , $U_\delta(x) = \{z \in \mathbf{R}^n \mid \|z - x\| \leq \delta\}$ – δ -окрестность точки x . В пункте 2 метода М1, после выполнения хотя бы одного из критериев пунктов 8 или 9 при некотором i , происходит обновление характеристик встроеного алгоритма решения неравенств. Подпоследовательности, выделяемые из последовательностей x_i , Σ_i ,

$i=1,2,\dots$, в пункте 2 алгоритма М1, обозначим соответственно $z_k = x_{q_k}$, $Q_k = \Sigma_{q_k}$, $k=1,2,\dots$, где значения $q_k = i$ задаются в моменты выполнения пункта 2.

Теорема 3. Пусть функция $f(x)$ строго выпукла на $\mathbf{R}^n$, множество $D(x_0)$ ограничено и параметры ε_k, m_k , задаваемые в пункте 2 алгоритма М1, фиксированы:

$$\varepsilon_k = E_0 > 0, \quad m_k = M_0, \quad (19)$$

Тогда, если x^* – предельная точка последовательности $\{x_{q_k}\}_{k=1}^{\infty}$, генерируемой алгоритмом М1, то

$$d(x^*) \leq \max \{E_0, R(x_0) / \sqrt{M_0}\} \equiv d_0, \quad (20)$$

где $R(x_0) = \max_{x \in D(x_0)} \max_{v \in \partial f(x)} \|v\|$. В частности, если $M_0 \geq R^2(x_0)E_0^{-2}$, то $d(x^*) \leq E_0$.

Доказательство. Существование предельных точек последовательности $\{z_k\}$ следует из ограниченности множества $D(x_0)$ и $z_k \in D(x_0)$. Допустим, что утверждение теоремы неверно: предположим, что подпоследовательность $z_{k_s} \rightarrow x^*$, но

$$d(x^*) = d^* > d_0 > 0. \quad (21)$$

Положим
$$\varepsilon = (d^* - d_0) / 2. \quad (22)$$

Обозначим $S_\varepsilon^* = S_\varepsilon(\partial f(x^*))$. Выберем $\delta > 0$, такое, что

$$\partial f(x) \subset S_\varepsilon^* \quad \forall x \in S_\delta(x^*). \quad (23)$$

Такой выбор возможен в силу полунепрерывности сверху точечно-множественного отображения $\partial f(x)$ (см. [2], с. 289).

Выберем номер K , такой, что при $k_s > K$ будет справедливо

$$z_{k_s} \in S_{\delta/2}(x^*), \quad x_i \in S_\delta(x^*), \quad q_{k_s} \leq i \leq q_{k_s} + M_0, \quad (24)$$

т.е. такой номер K , что точки x_i остаются в окрестности $S_\delta(x^*)$ в течение, по крайней мере, M_0 шагов алгоритма. Такой выбор возможен в силу сходимости $z_{k_s} \rightarrow x^*$ и результата леммы 3, условия которой выполняются при условиях теоремы 3 и наличии точного одномерного спуска в пункте 7 алгоритма М1.

Согласно предположению (21), условиям выбора ε (22), δ (23) и k (24) при $k_s > K$ будет выполняться неравенство

$$\rho(S_\varepsilon^*) \geq \rho(\partial f(x^*)) - \varepsilon = d^* - (d^* - d_0) / 2 > d_0. \quad (25)$$

При $k_s > K$, в силу справедливости соотношений (24), из (23) следует $g_i \in S_\varepsilon^*$, $q_{k_s} \leq i \leq q_{k_s} + M_0$. Алгоритм М1 содержит в своем составе алгоритм А1. Поэтому, с учетом оценок из (16), в зависимости от того, в каком из пунктов алгоритма М1 (9 или 10) произойдет обновление при некотором $i = j$, будет выполнено одно из неравенств:

$$\rho(S_\varepsilon^*) \leq \sum_j^{-0.5} \leq \varepsilon_k \leq E_0 \leq d_0; \quad (26)$$

$$\rho(S_\varepsilon^*) \leq R(x_0) / \sqrt{m_k} \leq R(x_0) / \sqrt{M_0} \leq d_0, \quad (27)$$

где последний переход в неравенствах вытекает из определения величины d_0 в (20). Но (25) противоречит как (26), так и (27). Полученное противоречие доказывает теорему.

Согласно оценке (20), для любой предельной точки последовательности $\{z_k\}$, генерируемой алгоритмом М1, будет выполнено $d(x^*) < d_0$, а следовательно, будет справедлива оценка (18).

В следующей теореме определяются условия, при которых алгоритм М1 генерирует последовательность $\{x_i\}$, сходящуюся к точке минимума.

Теорема 4. Пусть функция $f(x)$ строго выпукла, множество $D(x_0)$ ограничено и

$$\varepsilon_k \rightarrow 0, \quad m_k \rightarrow \infty. \quad (28)$$

Тогда любая предельная точка последовательности $\{x_{q_k}\}$, генерируемая алгоритмом М1, является точкой минимума функции $f(x)$ на $\mathbf{R}^n$.

Доказательство. Допустим, что утверждение теоремы неверно: предположим, что подпоследовательность $z_{k_s} \rightarrow x^*$, но при этом найдется такое $d_0 > 0$, что будет выполняться неравенство (21). Как и ранее зададим ε согласно (22). Выберем $\delta > 0$, такое, что будет выполнено (23).

В силу условий (28) найдется такое K_0 , что при $k > K_0$ будет выполняться соотношение

$$\max\{\varepsilon_k, R(x_0)/\sqrt{m_k}\} \leq d_0. \quad (29)$$

Обозначим $E_0 = d_0$ и M_0 – наименьшее значение m_k при $k > K_0$. Дальнейшие рассуждения аналогичны доказательствам теоремы 3.

Выберем номер $K > K_0$, такой, что при $k_s > K$ будет справедливо (24), т.е. такой номер K , что точки x_i остаются в окрестности $S_\delta(x^*)$ в течение, по крайней мере, M_0 шагов алгоритма. Согласно предположению (21), условиям выбора ε (22), δ (23) и k (24) при $k_s > K$ будет выполняться неравенство (25). При $k_s > K$, в силу справедливости соотношений (24), из (23) следует $g_i \in S_\varepsilon^*$, $q_{k_s} \leq i \leq q_{k_s} + M_0$. Алгоритм М1 содержит в своем составе алгоритм А1. Поэтому с учетом оценок из (16), в зависимости от того, в каком из пунктов алгоритма М1 (9 или 10) произойдет обновление при некотором $i = j$, будет выполнено одно из неравенств (26), (27), где последний переход в неравенствах следует из определения величин E_0 и M_0 . Но (25) противоречит как (26), так и (27). Полученное противоречие доказывает теорему.

Обозначим $\nabla f(x)$ – градиент функции, который в случае дифференцируемой выпуклой функции совпадает с субградиентом и является единственным элементом субградиентного множества [2]. Установим связь алгоритма М1 с методом сопряженных градиентов (МСГ) [13]:

$$\begin{aligned} \bar{x}_{i+1} &= \bar{x}_i - \bar{\gamma}_i \bar{s}_{i+1}, \quad \bar{\gamma}_i = \arg \min_{\bar{\gamma}} f(\bar{x}_i - \bar{\gamma} \bar{s}_{i+1}), \quad i = 0, 1, \dots, n-1, \\ \bar{s}_1 &= g_0, \quad \bar{s}_{i+1} = \bar{g}_i + \frac{(\bar{g}_i, \bar{g}_i)}{(\bar{g}_{i-1}, \bar{g}_{i-1})} \bar{s}_i, \quad i = 1, 2, \dots, n-1, \quad \bar{g}_i = \nabla f(\bar{x}_i). \end{aligned} \quad (30)$$

Теорема 5. Пусть функция $f(x)$, $x \in \mathbf{R}^n$, квадратичная, ее матрица вторых производных строго положительно определена, тогда алгоритм М1 при $\varepsilon_k = 0$, $m_k = n$ находит минимум за конечное число итераций, не превосходящее n , при этом для последовательностей, генерируемых алгоритмами МСГ и М1 при условии равенства начальных точек $x_0 = \bar{x}_0$ справедливы соотношения

$$а) p_i = g_i, \quad б) s_{i+1} = \bar{s}_{i+1} / (g_i, g_i), \quad в) x_{i+1} = \bar{x}_{i+1}, \quad i = 0, 1, \dots, n-1. \quad (31)$$

Доказательство проведем по индукции. В силу условий теоремы в пункте 4 алгоритма М1 $g_i = \nabla f(x_i)$. В результате итерации алгоритма М1 при $i = 0$ после выполнения действий пунктов 1–5 имеем: $p_{-1} = 0$, $s_0 = 0$, $p_0 = g_0$, $s_1 = g_0 / (g_0, g_0)$. Отсюда следует равенство (31а) при $i = 0$. Поскольку для МСГ на итерации при $i = 0$ $\bar{s}_1 = g_0$, то направления спуска в обоих алгоритмах удовлетворяют равенству (31б) при $i = 0$. В силу точного одномерного спуска и коллинеарности направлений спуска будет выполнено равенство (31в) при $i = 0$.

Ограничения в М1 $\varepsilon_k = 0$, $m_k = n$ необходимы для исключения преждевременных обновлений.

Предположим, что равенства (31) выполнены при $i = 0, 1, \dots, l$, где $l > 0$. Покажем, что они выполняются при $i = l+1$. В пункте 5 алгоритма М1 при $i = l+1$ в результате ортогонализации векторов g_{l+1} , p_l будет получен $p_{l+1} = g_{l+1}$, поскольку, согласно (31а), $p_l = g_l$, градиенты алгоритмов МСГ и М1 совпадают в силу идентичности точек (31в), в которых они вычисляются, а градиенты, используемые в МСГ, а следовательно и в М1, взаимно ортогональны [7]. Это доказывает (31а) при $i = l+1$.

В силу условия точного одномерного спуска выполняется равенство $(s_{l+1}, g_{l+1}) = 0$. Поэтому преобразование пункта 5 алгоритма М1 для вектора s с учетом (31а)) при $i = l+1$, (31б) при $i = l$ и (30) примет вид

$$s_{l+2} = s_{l+1} + \frac{g_{l+1}}{(g_{l+1}, g_{l+1})} = \frac{\bar{s}_{l+1}}{(g_l, g_l)} + \frac{g_{l+1}}{(g_{l+1}, g_{l+1})} = \frac{\bar{s}_{l+2}}{(g_{l+1}, g_{l+1})}.$$

Отсюда следует (31б). В силу точного одномерного спуска и коллинеарности направлений спуска будет выполнено равенство (31в) при $i = l+1$.

Из приведенного доказательства эквивалентности последовательностей, генерируемых алгоритмами МСГ и М1 и свойства окончания процесса минимизации методом МСГ не более чем через n шагов [13] следует доказательство теоремы.

4. Реализация алгоритма минимизации

Алгоритм М1 реализован согласно технике реализации РСМ [3, 5–8]. Рассмотрим версию алгоритма М1, включающую в себя процедуру одномерной минимизации вдоль направления s , функции которой заключаются в построении: а) текущего приближения минимума x_m ; б) точки y из окрестности x_m , такой, что для $g_1 \in \partial f(y)$ выполняется неравенство $(s, g_1) \leq 0$. Субградиент g_1 используется для решения системы неравенств.

Обращение к процедуре обозначим:

$$OM(\{x, s, g_x, f_x, h_0\}; \{\gamma_m, f_m, g_m, \gamma_1, g_1, h_1\}).$$

Блок входных параметров состоит из точки текущего приближения минимума x ,

направления спуска s , $g_x \in \partial f(x)$, $f_x = f(x)$, начального шага h_0 . Предполагается, что выполняется необходимое условие возможности спуска $(g_x, s) > 0$ в направлении s . Блок выходных параметров включает в себя γ_m – шаг в точку полученного приближения минимума $x^+ = x - \gamma_m s$, $f_m = f(x^+)$, $g_m \in \partial f(x^+)$, γ_1 – шаг вдоль s , такой, что в точке $y^+ = x - \gamma_1 s$ для $g_1 \in \partial f(y^+)$ выполняется неравенство $(g_1, s) \leq 0$, и h_1 – начальный шаг спуска для следующей итерации. В излагаемом ниже алгоритме векторы $g_1 \in \partial f(y^+)$ используются для решения множества неравенств, а точки $x^+ = x - \gamma_m s$ как точки приближений минимума.

Алгоритм одномерного спуска (ОМ). Пусть требуется найти приближение минимума одномерной функции $\phi(\beta) = f(x - \beta s)$, где x – некоторая точка, s – направление спуска. Возьмем возрастающую последовательность $\beta_0 = 0$ и $\beta_i = h_0 q_M^{i-1}$ при $i \geq 1$. Обозначим $z_i = x - \beta_i s$, $r_i \in \partial f(z_i)$, $i = 0, 1, 2, \dots$, l – номер i , при котором впервые выполнится соотношение $(r_i, s) \leq 0$. Зададим параметры отрезка локализации $[\gamma_0, \gamma_1]$ одномерного минимума: $\gamma_0 = \beta_{l-1}$, $f_0 = f(z_{l-1})$, $g_0 = r_{l-1}$, $\gamma_1 = \beta_l$, $f_l = f(z_l)$, $g_l = r_l$ и найдем точку минимума γ^* одномерной кубической аппроксимации функции на отрезке локализации. Вычислим

$$\gamma_m = \begin{cases} q_\gamma \gamma_1, & \text{если } l = 1 \text{ и } \gamma^* \leq q_\gamma \gamma_1, \\ \gamma_1, & \text{если } \gamma_1 - \gamma^* \leq q_\gamma (\gamma_1 - \gamma_0), \\ \gamma_0, & \text{если } l > 1 \text{ и } \gamma^* - \gamma_0 \leq q_\gamma (\gamma_1 - \gamma_0), \\ \gamma^*, & \text{в остальных случаях.} \end{cases} \quad (32)$$

Вычислим начальный шаг спуска для следующей итерации:

$$h_1 = q_m (h_0 \gamma_m)^{1/2}. \quad (33)$$

Алгоритм минимизации. В предлагаемом ниже варианте реализации алгоритма М1 обновление для метода решения неравенств не производится, а точный одномерный спуск заменен на приближенный.

Алгоритм.

1. Задать начальное приближение $x_0 \in \mathbf{R}^n$, начальный шаг одномерного спуска h_0 . Положить: $i = 0$, $g_0 = \tilde{g}_0 \in \partial f(x_0)$, $g_{i-1} = 0$, $p_{i-1} = 0$, $f_0 = f(x_0)$, $s_0 = \tilde{s}_0 = 0$. Задать параметры останова: N – максимально допустимое число итераций, ε_x – точность минимизации по аргументу, ε_g – точность минимизации по градиенту.

2. Получить приближение $\tilde{s}_{i+1} = s_i + \frac{1 - (s_i, \tilde{g}_i)}{(p_i, \tilde{g}_i)} p_i$,

$$\text{где } p_i = \begin{cases} \tilde{g}_i, & \text{если } (\tilde{g}_i, p_{i-1}) \geq 0, \\ \tilde{g}_i - \frac{(\tilde{g}_i, p_{i-1})}{\|p_{i-1}\|^2} p_{i-1}, & \text{если } (\tilde{g}_i, p_{i-1}) < 0. \end{cases}$$

Здесь осуществляется шаг метода решения неравенств.

3. Получить направление спуска

$$s_{i+1} = \begin{cases} \tilde{s}_{i+1}, & \text{если } (\tilde{s}_{i+1}, g_i) \geq 1, \\ \tilde{s}_{i+1} + g_i(1 - (\tilde{s}_{i+1}, g_i)) / (g_i, g_i), & \text{если } (\tilde{s}_{i+1}, g_i) < 1. \end{cases} \quad (34)$$

4. Произвести одномерный спуск вдоль $w_{i+1} = s_{i+1}(s_{i+1}, s_{i+1})^{-1/2}$:

$$OM(\{x_i, w_{i+1}, g_i, f_i, h_i\}; \{\gamma_{i+1}, f_{i+1}, g_{i+1}, \tilde{\gamma}_{i+1}, \tilde{g}_{i+1}, h_{i+1}\}).$$

Вычислить приближение точки минимума $x_{i+1} = x_i - \gamma_{i+1}w_{i+1}$.

5. Если $i > N$ или $\|x_{i+1} - x_i\| \leq \varepsilon_x$, или $\|g_{i+1}\| \leq \varepsilon_g$, то закончить вычисления, иначе положить $i = i + 1$ и перейти на пункт 2.

Поясним действия алгоритма. Из OM поступают два субградиента $\tilde{g}_{i+1}$ и g_{i+1} . Первый из них используется для решения неравенств в пункте 2, а второй – в пункте 3, для коррекции направления спуска с помощью формулы (3) с целью обеспечения необходимого условия $(s_{i+1}, g_i) > 0$ возможности спуска в направлении $(-s_{i+1})$. Как показано в [5, 7, 8] итерация (3) в (34) при $(\tilde{s}_{i+1}, g_i) < 1$ не ухудшает текущее приближения $\tilde{s}_i$ решения системы неравенств, поэтому это преобразование проводится. В пункте 3, когда $(\tilde{s}_{i+1}, g_i) > 1$, условие $(\tilde{s}_{i+1}, g_i) > 0$ выполнено и, согласно результатам работ [5, 7, 8], нет теоретических рекомендаций по улучшению решения системы неравенств. Поэтому преобразование коррекции не производится.

Хотя обоснование сходимости идеализированных версий РСМ [2–8] производится при условии точного одномерного спуска, реализации этих алгоритмов осуществляется с процедурами одномерной минимизации, в которых начальный шаг, в зависимости от прогресса, может увеличиваться или уменьшаться, что определяется заданными коэффициентами $q_M > 1$ и $q_m < 1$. При этом минимальный шаг на итерации не может быть меньше некоторой доли начального шага, величина которой задана в (32) параметрами $q_{\gamma 1} = 0,1$ и $q_{\gamma} = 0,2$, приведенные значения которых использовались нами при расчетах.

5. Численный эксперимент

Алгоритм М1 реализован согласно технике реализации алгоритма М0 [5–8], в которой ключевое значение играют коэффициенты уменьшения $q_m < 1$ и увеличения $q_M > 1$ начального шага одномерного спуска на итерации. Значения q_m близкие к 1 обеспечивают малую скорость убывания шага и соответственно малую скорость сходимости метода. При этом малая скорость убывания шага устраняет заикливание метода в силу того, что субградиенты функции, участвующие в решении неравенств берутся из более широкой окрестности. Выбор параметра q_m должен соизмеряться с возможной скоростью сходимости метода минимизации. Чем выше скоростные возможности алгоритма, тем меньшим может быть выбран этот параметр. Например, в РСМ с растяжением пространства [4–8] выбирается $q_m = 0,8$. Для гладких функций выбор этого параметра не критичен и его можно брать из интервала $[0,8–0,98]$. От параметра возрастания шага скорость сходимости практически не зависит, поэтому его можно взять постоянным $q_M = 1,5$.

Исследование проводилось на следующих функциях.

$$1. f_1(x) = \sum_{k=1}^n |x_k| \cdot k, x^* = (0, 0, \dots, 0), x_0 = (10, \frac{10}{2}, \frac{10}{3}, \dots, \frac{10}{n});$$

$$2. f_2(x) = \sum_{k=1}^n x_k^2 \cdot k^2, x^* = (0, 0, \dots, 0), x_0 = (10, \frac{10}{2}, \frac{10}{3}, \dots, \frac{10}{n});$$

$$3. f_3(x) = \sum_{k=1}^{n-1} [1000(x_k - x_{k+1})^2 + (1 - x_{k+1})^2], x^* = (1, \dots, 1), x_0 = (0, \dots, 0).$$

Функция $f_3(x)$ взята из [12]. В таблице приведено количество затраченных методом вычислений функции и субградиента, которое соответствует моменту выполнения условия $f_k - f^* \leq \varepsilon$. Знаком N помечены задачи, которые не удалось решить за количество итераций, не превышающее заданное максимальное, т.е. найти точку приближения x_k , в которой бы выполнялось условие останова алгоритма по значению функции, т.е. $f_k - f^* \leq \varepsilon$. При тестировании другие критерии останова не использовались.

Результаты численного эксперимента

n	f_2	$f_1, \varepsilon = 10^{-5}$ $q_m = 0,999, q_M = 1,5$		$f_2, \varepsilon = 10^{-10}$ $q_m = 0,98, q_M = 1,5$		$f_3, \varepsilon = 10^{-10}$ $q_m = 0,85, q_M = 1,5$	
		M0	M1	M0	M1	M0	M1
100	938	27573	26646	2064	1649	760	604
200	2359	84837 (N)	51203	4008	3096	869	612
300	3891	84773 (N)	54203	5781	4364	903	627
400	5929	84705 (N)	54070	7804	5884	885	605
500	7632	84731 (N)	53654	10086	7245	947	665
600	9264	84767 (N)	54290	12457	8598	935	621
700	10914	84677 (N)	68003	14837	10564	975	631
800	12563	84957 (N)	51794	17345	11822	960	658
900	14272	84649 (N)	66241	19839	14073	948	653
1000	16008	87499 (N)	56017	22478	16042	967	703

Функция 2 квадратичная с отношением собственных значений $1/n^2$. Во втором столбце приведены результаты счета для метода сопряженных градиентов [13] с кубической интерполяцией при поиске одномерного минимума с заданной точностью, которая выбрана экспериментально из условия минимизации количества вычислений функции и градиента на решение комплекса задач. Отметим, что использование в МСГ грубого одномерного спуска из M1 приводит к многократному увеличению числа итераций. Здесь алгоритм M1 эффективнее метода M0 и сравним с МСГ при размерностях выше 400. Поэтому при решении гладких задач минимизации высокой размерности с высокой степенью вытянутости поверхностей уровня наряду с МСГ возможно применение многошаговых РСМ.

Кусочно-линейная функция 1 имеет одинаковую вытянутость линий уровня с функцией 2, но неизмеримо сложнее для метода минимизации. Здесь, в методе M0 происходит заикливание, а увеличение числа итераций не приводит к решению задачи. На этом примере заметно существенное повышение эффективности за счет ортогонализации векторов обучения в алгоритме M1.

На квадратичной функции 3 с небольшим разбросом собственных значений методы М0 и М1 практически эквивалентны.

Заключение

В работе на задачу решения множества неравенств распространен итерационный метод решения системы равенств [11]. Разработанный алгоритм обоснован теоретически. На его основе сформулирован и обоснован релаксационный субградиентный метод минимизации, который, в силу незначительных затрат памяти (пропорционально размерности задачи) и отсутствия матричных вычислений, пригоден для решения задач высокой размерности.

По свойствам сходимости на квадратичных функциях высокой размерности, при больших разбросах собственных значений, разработанный алгоритм превосходит имеющиеся многошаговые релаксационные субградиентные методы и соизмерим по эффективности с методом сопряженных градиентов.

Новый метод позволяет расширить круг решаемых негладких задач. Численные результаты свидетельствуют о повышении эффективности метода минимизации при введении ортогонализации векторов обучения в алгоритм решения неравенств, что особенно проявляется при решении негладких задач высокой размерности с высокой степенью вытянутости поверхностей уровня.

ЛИТЕРАТУРА

1. Wolfe P. Note on a method of conjugate subgradients for minimizing nondifferentiable functions // Math. Programming. 1974. V. 7. No. 3. P. 380–383.
2. Демьянов В.Ф., Васильев Л.В. Недифференцируемая оптимизация. М.: Наука, 1972. 368 с.
3. Крутиков В.Н., Петрова Т.В. Новый метод решения задач минимизации большой размерности // Вестник КемГУ. Кемерово, 2001. Вып. 4. С. 65–71.
4. Шор Н.З. Методы минимизации недифференцируемых функций и их приложения. Киев: Наукова думка, 1979. 199 с.
5. 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. 33–49.
6. Крутиков В.Н., Горская Т.А. Семейство релаксационных субградиентных методов с двухранговой коррекцией матриц метрики // Экономика и мат. методы. 2009. Т. 45. № 4. С. 37–80.
7. Крутиков В.Н. Релаксационные методы безусловной оптимизации, основанные на принципах обучения: учеб. пособие / ГОУ ВПО «Кемеровский государственный университет». Кемерово: Кузбассвуиздат, 2004. 171 с.
8. Крутиков В.Н. Обучающиеся методы безусловной оптимизации и их применение. Томск: Изд-во Том. гос. педагогического ун-та, 2008. 264 с.
9. Kaczmarz S. Approximate solution of systems of linear equations // Int. J. Control. 1993. V. 54. No. 3. P. 1239–1241.
10. Цыпкин Я.З. Основы теории обучающихся систем. М.: Наука, 1981. 251 с.
11. Крутиков В.Н., Вершинин Я.Н. Алгоритмы обучения на основе ортогонализации последовательных векторов // Вестник КемГУ. 2012. Вып. 2 (50). С. 37–42.
12. Скоков В.А. Варианты метода уровней для минимизации негладких выпуклых функций и их численное исследование // Экономика и математические методы. 1997. Т. 33. № 1.
13. Поляк Б.Т. Введение в оптимизацию. М.: Наука, 1983. 384 с.

Krutikov V.N., Vershinin Ya.N. THE SUBGRADIENT MULTISTEP MINIMIZATION METHOD FOR NONSMOOTH HIGH-DIMENSIONAL PROBLEMS

In this paper, a new multistep relaxation subgradient minimization method is proposed. It is based on principles of organization of "conjugate subgradients" methods. The presented method belongs to the class of relaxation methods of the ε -subgradient type (RSM) and is intended for solving nonsmooth high-dimensional problems.

The space tension RSMs available at present are comparable in the rate of convergence for smooth functions with quasi-Newton methods and are efficient in solving nonsmooth problems of the ravine type. At high dimensional problems, its effectiveness is reduced due to the necessity of storage and transformation of the metric matrix. In the smooth case, the conjugate gradient method substitutes quasi-Newton methods at high-dimensional problems. Existing multistep RSMs are significantly inferior to the subgradient space tension methods in the rate of convergence and loop at ravine type nonsmooth optimization problems. That is why they are practically not applied for even for small dimension problems. These circumstances determine the importance of establishing effective multistage RSMs. In the considered relaxation subgradient method, additional learning relations are used at iterations with the aim to improve the efficiency of the learning algorithm for a known method based on extending the Kaczmarz algorithm to inequality systems. This innovation expands the range of solved nonsmooth optimization problems and increases the rate of convergence in solving smooth and non-smooth minimization problems.

Numerical results indicate an increase in the minimization method efficiency due to orthogonalization of learning vectors in the algorithm that solves the inequalities, which is particularly evident when solving nonsmooth problems of high dimensionality with a high degree of elongation of the level surfaces. According to the convergence properties at high dimension quadratic functions, at a large scatter of eigenvalues, the developed algorithm is superior to existing multi-step relaxation subgradient methods and is comparable in the effectiveness to the conjugate gradients method.

Keywords: Kaczmarz algorithm, multistep algorithm, minimization method, convergence rate

Krutikov Vladimir Nikolaevich (Doctor of technical Sciences, Prof.,
Kemerovo State University, Kemerovo, Russian Federation)
E-mail: krutikovvn@gmail.com

Vershinin Yaroslav Nikilaevich (M. Sc, Kemerovo State University, Kemerovo, Russian Federation)
E-mail: Azimus88@gmail.com

REFERENCES

1. Wolfe P. Note on a method of conjugate subgradients for minimizing nondifferentiable functions (1974) *Math. Programming*. V. 7. No. 3, pp. 380–383.
2. Dem'yanov V.F., Vasil'ev L.V. *Nedifferentsiruemaya optimizatsiya*. Moscow, Nauka Publ., 1972. 368 p. (in Russian)
3. Krutikov V.N., Petrova V.V. Novyy metod resheniya zadach minimizatsii bol'shoy razmernosti (2001) *Vestnik KemGU*. No. 4, pp. 65–71. (in Russian)
4. Shor N.Z. *Metody minimizatsii nedifferentsiruemykh funktsiy i ikh prilozheniya*. Kiev: Naukova dumka Publ., 1979. 199 p. (in Russian)
5. Krutikov V.N., Petrova V.V. Relaksatsionnyy metod minimizatsii s rastyazheniem prostranstva v napravlenii subgradienta (2003) *Ekonomika i mat. metody*. V. 39. No. 1, pp. 33–49. (in Russian)
6. Krutikov V.N., Gorskaya V.A. Semeystvo relaksatsionnykh subgradientnykh metodov s dvukhrangovoy korrektsiey matritsy metriki (2009) *Ekonomika i mat. metody*. V. 45. No. 4, pp. 37–80. (in Russian)
7. Krutikov V.N. *Relaksatsionnye metody bezuslovnoy optimizatsii, osnovannye na printsipakh obucheniya*. Kemerovo, Kuzbassvuzizdat Publ., 2004. 171 p. (in Russian)

8. Krutikov V.N. Obuchayushchiesya metody bezuslovnoy optimizatsii i ikh primeneniye. Tomsk, Izd-vo Tom. gosudarstvennogo pedagogicheskogo un-ta, 2008. 264 p. (in Russian)
9. Kaczmarz S. Approximate solution of systems of linear equations (1993) Int. J. Control. V. 54. No. 3, pp. 1239–1241.
10. Tsypkin Ya.Z. Osnovy teorii obuchayushchikhsya sistem. Moscow, Nauka Publ., 1981. 251 p. (in Russian)
11. Krutikov V.N., Vershinin Ya.N. Algoritmy obucheniya na osnove ortogonalizatsii posledovatel'nykh vektorov (2012) Vestnik KemGU. No. 2 (50), pp. 37–42. (in Russian)
12. Skokov V.A. Varianty metoda urovney dlya minimizatsii negladkikh vypuklykh funktsiy i ikh chislennoe issledovanie (1997) Ekonomika i matematicheskie metody. V. 33. No 1. (in Russian)
13. Polyak B.V. Vvedenie v optimizatsiyu. Moscow, Nauka Publ., 1983. 384 p. (in Russian)