ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Июль–сентябрь 2022. Т. 29, № 3. С. 24–44

УДК 519.8 DOI 10.33048/daio.2022.29.739

ОПТИМИЗАЦИЯ ПАРАМЕТРОВ СУБГРАДИЕНТНОГО МЕТОДА НА ОСНОВЕ ДВУХРАНГОВОЙ КОРРЕКЦИИ МАТРИЦ МЕТРИКИ

В. Н. Крутиков[1, a], *П. С. Станимирович*[2, b], *О. Н. Инденко*[1, c], *Е. М. Товбис*[3, d], *Л. А. Казаковцев*[3, e]

[1] Кемеровский гос. университет,
ул. Красная, 6, 650043 Кемерово, Россия

[2] Факультет естественных наук и математики, Нишский университет,
ул. Вышеградска, 33, 18000 Ниш, Сербия

[3] Сибирский гос. университет науки и технологий им. акад. Решетнёва,
пр. Красноярский рабочий, 31, 660031 Красноярск, Россия

E-mail: [a] krutikovvn@rambler.ru, [b] pecko@pmf.ni.ac.rs,
[c] oksana230805@mail.ru, [d] sibstu2006@rambler.ru, [e] levk@bk.ru

Аннотация. Предлагается релаксационный субградиентный метод, включающий оптимизацию параметров с использованием коррекции матриц метрики второго ранга, со структурой, аналогичной квазиньютоновским методам. Преобразование матрицы метрики заключается в подавлении ортогональных и усилении коллинеарных компонентов вектора субградиента минимальной длины. Задача построения матрицы метрики формулируется как задача решения системы неравенств. Решение такой системы основано на новом алгоритме обучения. Получена оценка скорости его сходимости в зависимости от параметров множества субградиентов. На этой основе разработан и исследован новый релаксационный субградиентный метод. Вычислительные эксперименты над сложными функциями большой размерности подтверждают эффективность предложенного алгоритма. Табл. 4, библиогр. 32.

Ключевые слова: выпуклая оптимизация, негладкая оптимизация, релаксационный субградиентный метод.

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования России (гос. контракт № FEFE–2020–0013). Работа второго автора выполнена при поддержке Научного фонда Республики Сербия (грант № 7750185) и Министерства образования, науки и технологического развития Республики Сербия (контракт № 451–03–68/2020–14/200124).

Введение

Предметом нашего исследования является минимизация выпуклой, но не обязательно дифференцируемой функции $f(x)$, $x \in \mathbb{R}^n$, в конечномерном евклидовом пространстве $\mathbb{R}^n$. Начало исследований в области субградиентных методов положено в работах Н. З. Шора, Б. Т. Поляка и др. [1–4]. Существует ряд направлений построения методов негладкой оптимизации. Одно из них основано на построении и использовании приближений функции [5–12]. На базе этого удалось получить и теоретически обосновать широкий спектр методов [5–7], 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. Ряд эффективных подходов в области негладкой оптимизации возник в результате создания первых субградиентных методов с растяжением пространства [13–15], к числу которых относится и субградиентный метод с растяжением пространства в направлении субградиента [14] на классе релаксационных по расстоянию до экстремума методов минимизации [3, 16, 17].

Первые релаксационные субградиентные методы (РСМ) предложены в [4, 18, 19]. В [14] авторами разработан эффективный РСМ с растяжением пространства в направлении разности субградиентов (r-алгоритм). Последующие работы по разработке эффективного РСМ с растяжением пространства связаны с выявлением происхождения r-алгоритма и его теоретическим обоснованием [20–22]. Формализация модели субградиентных множеств и использование идей и алгоритмов машинного обучения [20] позволили выявить принципы РСМ с растяжением пространства [21, 23] и получить теоретическую основу для их развития. Оказалось, что задачу нахождения направления спуска в РСМ можно свести к задаче решения системы неравенств на субградиентных множествах и математически сформулировать как решение задачи минимизации некоторого функционала качества. В этом случае свойства алгоритма обучения определяют скорость сходимости метода минимизации. Учитывая высокую скорость сходимости РСМ с растяжением пространства и их возможности для решения негладких невыпуклых задач минимизации, представляется актуальным развитие теории и практики этих негладких методов оптимизации.

Авторами в [21] было предложено семейство методов решения неравенств с двухранговой коррекцией матриц метрики. Наша цель — выделить из этого семейства алгоритм обучения, который использует самую свежую информацию о субградиентах текущего приближения минимума и оптимизирует свои параметры для построения на этой основе эффективного РСМ, аналогичного квазиньютоновским методам [24, 25].

Обозначим множество допустимых направлений через $S(G) = \{s \in \mathbb{R}^n \mid \min_{g \in G}(s, g) > 0\}$, $G \subset \mathbb{R}^n$. Пусть $\partial_\varepsilon f(x_k)$ — это ε-субградиентное множество в точке x_k и $\partial f(x_k) \equiv \partial f_{\varepsilon=0}(x_k)$ — субградиентное множество (субдифференциал) в x_k.

Последовательные приближения в РСМ ε-субградиентного типа строятся с помощью итераций [4, 18–21, 26]

$$x_{k+1} = x_k - \gamma_k s_{k+1} \tag{1}$$

с начальной аппроксимации x_0. Параметр $\gamma_k > 0$ — это размер шага, определяемый одномерным линейным поиском из [20], в то время как направление спуска s_{k+1} выбирается из множества $S(\partial_\varepsilon f(x_k))$ [18]. Если $S(G) \neq \varnothing$, то любой вектор $s \in S(G)$ является решением системы неравенств

$$(s, g) > 0, \quad g \in G, \tag{2}$$

т. е. определяет нормаль плоскости, разделяющей начало координат и множество G. Одним из решений (2) является $\eta(G)$, обозначающий вектор минимальной длины из G. В методе ε-наискорейшего спуска $s_{k+1} = \eta(\partial_\varepsilon f(x_k))$ [18]. Из-за отсутствия явного описания множества ε-субградиентов в (1) вектор $\eta(G)$ оболочки субградиентов G, полученный на траектории спуска, используется как направление спуска [4, 18, 19]. Эффективные методы такого типа могут быть получены с использованием алгоритмов обучения [20, 27] для оценки параметров разделяющей плоскости.

В данной работе мы предлагаем релаксационный субградиентный метод, в котором решается система неравенств (2) на основе коррекции матриц метрики второго ранга, аналогичных используемым в квазиньютоновских методах оптимизации. Цель преобразования матриц метрики состоит в том, чтобы отобразить широкий пучок субградиентов в текущей окрестности минимума в узкий пучок направлений, образующих острые углы со всеми субградиентами вблизи текущего решения. Использование таких направлений позволяет выйти за пределы этой окрестности. Для решения неравенств (2) используем формализованную модель субградиентных множеств и задействуем концепции и алгоритмы машинного обучения, в частности, итерационный метод наименьших квадратов. Для демонстрации эффективности предложенных алгоритмов проводятся вычислительные эксперименты на сложных большеразмерных тестовых функциях.

В разд. 1 представлен обзор основных предварительных результатов. Разд. 2 содержит описание алгоритма для решения (2). В разд. 3 представлен предлагаемый субградиентный метод минимизации, а иллюстративные численные примеры включены в разд. 4. Некоторые заключительные замечания изложены в последнем разделе.

1. Обзор предварительных результатов

Пусть множество $G \subset \mathbb{R}^n$ принадлежит некоторой гиперплоскости и $\eta(G)$ — его минимальный по длине вектор. Существует решение системы уравнений $(s,g)=1$, $g\in G$, которое одновременно удовлетворяет (2). Следовательно, такое решение может быть использовано в качестве решения для (2). Одно из возможных решений этой системы имеет вид $s_{k+1} = \arg\min_s F_k(s)$, где

$$F_k(s) = \sum_{i=0}^{k} w_i Q_i(s) + \frac{1}{2}\sum_{i=1}^{n} s_i^2, \quad Q_i(s) = \frac{1}{2}(q_i - (s,g_i))^2.$$

С учётом регуляризующей составляющей $\frac{1}{2}\sum\limits_{i=1}^{n} s_i^2$ такое решение может быть получено модифицированным итерационным методом наименьших квадратов (ИМНК) [28] со специальным масштабированием данных [20]

$$s_{k+1} = s_k + \frac{H_k g_k (q_k - (s_k, g_k))}{(1+A)(g_k H_k g_k)}, \quad s_0 = 0, \quad A > 0, \tag{3}$$

$$H_{k+1} = H_k - \frac{H_k g_k g_k^T H_k^T}{(1+A)(g_k, H_k g_k)}, \quad H_0 = I. \tag{4}$$

В отличие от обычного ИМНК, модифицированный ИМНК (3), (4) можно применять, начиная с первой итерации. Применяя (3), (4), получим уравнения r-алгоритма из [1] в форме из [22]. Обозначим разности $y_k = g_k - g_{k+1}$. Преобразуем данные $(s, g_i) = q_i$, $i = 0,1,\ldots,k$, для $q_i = 1$ вычитанием соседних равенств. Получим $(s, y_i) = 0$, $i = 0,1,\ldots,k-1$, и $(s, g_k) = 1$. Для части данных в $i = 0,1,\ldots,k-1$ путём замены $1/(1+A) = \left(1 - 1/\alpha_k^2\right)$ выполним преобразования (3), (4). В результате (3) в силу $s_0 = 0$ и $q_i = 0$ получим $s_i = 0$, $i = 0,1,\ldots,k-1$. Последовательно выполним преобразование (4):

$$H_{i+1} = H_i - \left(1 - \frac{1}{\alpha^2}\right)\frac{H_i y_i y_i^T H_i^T}{(y_i, H_i y_i)}, \quad H_0 = I, \quad i = 0,1,\ldots,k-1. \tag{5}$$

Обновление (4) дополнительных данных $\{g_k, q_k = 1\}$ может быть опущено, поскольку в результате (3) для этих данных будет получен вектор, коллинеарный вектору $H_k g_k$. Таким образом, используя (3), (4) для объединённых данных, получили уравнения преобразования матриц метрики r-алгоритма из [1] в виде (5), предложенном в [22].

Алгоритмы решения системы неравенств с коррекцией матриц метрики ранга 2 обладают лучшими характеристиками скорости сходимости [21]. В данной работе рассматривается частный случай этого семейства алгоритмов, который, встраиваясь в релаксационный метод минимизации, позволяет получить метод, близкий по структуре к квазиньютоновским методам. В этом методе целевая функция в (1) убывает на k-й итерации вместе с $s_{k+1} = H_k g_k$, $g_k \in \partial f(x_k)$, после чего выполняется следующее преобразование двухранговой метрики:

$$H_{k+1} = H_k - \left(1 - \frac{1}{\alpha_k^2}\right)\frac{H_k y_k y_k^T H_k^T}{(y_k, H_k y_k)} - \left(1 - \frac{1}{\beta_k^2}\right)\frac{H_k p_k p_k^T H_k^T}{(p_k, H_k p_k)}, \tag{6}$$

$$y_k = g_k - g_{k+1}, \quad p_k = g_{k+1} + t_k y_k, \quad t_k = -\frac{(y_k, H_k g_{k+1})}{(y_k, H_k y_k)},$$

$$\alpha_k > 1, \quad 0 < \beta_k \leqslant 1, \quad \alpha_k \beta_k > 1. \tag{7}$$

Здесь обновление (6) — это встроенный метод решения системы неравенств (2). Решение системы (2) на k-й итерации — это вектор $H_{k+1} g_{k+1}$, который используется как направление спуска. Для фиксированных параметров $\alpha_k = \alpha_{\text{const}}$, $\beta_k = \beta_{\text{const}}$ оценки, обеспечивающие сходимость, были получены в [21]. Скорость сходимости метода решения (2) и основанного на нём РСМ можно значительно увеличить, если использовать параметры α_k, β_k, которые настраиваются в зависимости от текущей ситуации. В данной работе мы формулируем и теоретически обосновываем алгоритм с оптимальным выбором параметров α_k, β_k.

Доказана сходимость предложенного алгоритма обучения за ограниченное число итераций при решении задачи (2) на отделимых множествах. С помощью алгоритма обучения разработан метод минимизации негладких функций. Обоснована его сходимость на строго выпуклых функциях. Алгоритм реализован и исследован численно. Вычислительная стоимость алгоритма при расчёте значений функции и её градиента снижена более чем вдвое по сравнению со стоимостью r-алгоритма. Предлагаемый метод минимизации будет полезен при решении различных задач негладкой невыпуклой минимизации с высокой степенью вырождения, возникающих, например, в ситуациях оценивания параметров математических моделей в условиях негладкой регуляризации [29–32].

2. Алгоритм решения системы неравенств

Обозначим

$$\eta_G = \min_{g \in G} \|g\|, \quad \rho_G = \|\eta_G\|, \quad \mu_G = \frac{\eta_G}{\|\eta_G\|}, \quad s^* = \frac{\mu_G}{\rho_G},$$

$$R_G = \max_{g \in G} \|g\|, \quad R_S = \max_{g \in G} (\mu_G, g), \quad r_G = \frac{\rho_G}{R_S}, \quad V_G = \frac{\rho_G}{R_G}, \quad M_G = \frac{R_S}{\rho_G}.$$

Введём зависимость $\theta(M)=(M-1)^2/(M+1)^2$ и обратную для неё функцию $m(\theta)=(1+\theta^{1/2})/(1-\theta^{1/2})$, удовлетворяющую равенству $m(\theta(M))=M$. Для некоторого $\theta<1/2$ определим зависимости $a(\theta)=1/(2\theta)$, $b(\theta)=1/(2(1-\theta))$. Будем использовать

Предположение 1. *Множество G выпуклое, замкнутое, ограниченное ($R_G<\infty$) и удовлетворяет условию отделимости, т. е. $\rho_G>0$.*

Векторы μ_G и s^* являются решениями системы (2). Параметры ρ_G и R_S характеризуют толщину множества G в направлении μ_G, выражаемую в виде двустороннего неравенства

$$\rho_G\leqslant(\mu_G,g)\leqslant R_S,\quad g\in G. \tag{8}$$

Из (8) с учётом определения вектора s^* получим

$$1\leqslant(s^*,g)\leqslant\frac{R_S}{\rho_G}=M_G,\quad g\in G. \tag{9}$$

Граница M_G в (9) существенно влияет на скорость сходимости рассматриваемых в работе методов решения неравенств. Величина R_s согласно (8) удовлетворяет ограничениям

$$\rho_G\leqslant R_S\leqslant\|\mu_G\|\max_{g\in G}\|g\|\leqslant R_G.$$

Для произвольной симметричной строго положительно определённой матрицы H размера $n\times n$ будем использовать обозначение $H>0$. Вместо (6) для краткости будем обозначать $H_{k+1}=(H_k,\alpha_k,\beta_k,y_k,p_k)$.

В предлагаемом алгоритме строятся последовательные приближения решения системы (2) в виде $s_k=H_kg_k$, где g_k — произвольный вектор из G, а матрицы $H_k>0$, $k=0,1,\dots$, корректируются так, что при определённых ограничениях на характеристики множества G через конечное число итераций будет получено решение системы (2).

Вектор p_k в (12) найден из условия ортогональности векторов $v_k=H_kp_k$ и y_k, что подразумевает равенство

$$(y_k,H_kp_k)=(y_k,v_k)=0. \tag{14}$$

В силу выпуклости множества G справедливы свойства

$$p_k=g_k(1+t_k)-t_ku_k\in G,\quad 0>t_k>-1,\quad \|p_k\|\geqslant\rho_G. \tag{15}$$

Исследуем зависимости $a(\theta)$ и $b(\theta)$, которые используются в алгоритме для вычисления величин параметров α_k^2, β_k^2. Обозначим $Q(\theta,a,b)=a\theta+b(1-\theta)$. Зависимости $a(\theta)$ и $b(\theta)$ являются решением задачи

$$\max_{a,b}\{ab\mid Q(\theta,a,b)\leqslant 0,\ \theta<1/2,\ a>1,\ b\leqslant 1,\ ab>1\}, \tag{16}$$

причём решение достигается на границе $Q(\theta,a(\theta),b(\theta))=0$ [21]. Как будет показано далее, приемлемые характеристики скорости сходимости

Алгоритм 1. $\mathrm{A}(\alpha_k, \beta_k)$

Вход: $k = 0$, $g_0 \in G$, начальная матрица $H_0 = I$, $q \geqslant 1$.

1: Выбрать θ_A такое, что

$$\theta(M_G) \leqslant \theta_A < 1/2. \tag{10}$$

2: Найти вектор $u_k \in G$ такой, что

$$(H_k u_k, g_k) \leqslant 0. \tag{11}$$

3: Если $u_k \in G$ не существует, то $H_k g_k \in S(G)$ — решение. Закончить алгоритм.

4: Вычислить

$$y_k = g_k - u_k, \quad t_k = -\frac{(y_k, H_k g_k)}{(y_k, H_k y_k)}, \quad p_k = g_k + t_k y_k. \tag{12}$$

5: Вычислить $C_k = \min\{|(y_k, H_k g_k)|, |(y_k, H_k u_k)|\}$ и $\theta_{gk}(m(\theta_A))$, где

$$\theta_{gk}(M) = \left(1 + \frac{(y_k, H_k y_k)}{(M-1)^2 (p_k, H_k p_k)} \left(1 + \frac{C_r}{(y_k, H_k y_k)}(M-1)\right)^2\right)^{-1}, \tag{13}$$

и параметр $\theta_k = \max\{\theta_A/q^2, \min\{\theta_{gk}(m(\theta_A)), \theta_A\}\}$ как проекцию $\theta_{gk}(m(\theta_A))$ на отрезок $[\theta_A/q^2, \theta_A]$.

6: Найти параметры $\alpha_k^2 = a(\theta_k)$, $\beta_k^2 = b(\theta_k)$ и рассчитать $H_{k+1} = (H_k, \alpha_k, \beta_k, y_k, p_k)$ из (6).

7: Выбрать произвольный вектор $g_{k+1} \in G$.

8: Положить $k := k + 1$ и перейти на шаг 2.

алгоритма $\mathrm{A}(\alpha_k, \beta_k)$ обеспечиваются при $Q(\theta, a, b) \leqslant 0$ и максимально возможном произведении $ab > 1$.

В следующей лемме приведены соотношения для характеристик параметров α_k^2, β_k^2 при условии (10). Здесь и далее будем обозначать $\alpha^2 = a(\theta_A)$, $\beta^2 = b(\theta_A)$.

Лемма 1. *При условии* (10) *для характеристик алгоритма* $\mathrm{A}(\alpha_k, \beta_k)$ *справедливы соотношения*

$$q^2\alpha^2 \geqslant \alpha_k^2 \geqslant \alpha^2 > 1, \quad \beta_k^2 \leqslant \beta^2 < 1, \quad \alpha_k^2\beta_k^2 \geqslant \alpha^2\beta^2 > 1, \tag{17}$$

$$\begin{aligned} &Q\big(\min\{\theta(M_G), \theta_{gk}(M_G)\}, \alpha_k^2, \beta_k^2\big) \\ &\qquad \leqslant Q\big(\min\{\theta_A, \theta_{gk}(m(\theta_A))\}, \alpha_k^2, \beta_k^2\big) \leqslant Q\big(\theta_k, \alpha_k^2, \beta_k^2\big) = 0. \end{aligned} \tag{18}$$

Обозначим $A_k = H_k^{-1}$, $\operatorname{tr} A$ и $\det A$ — след и определитель матрицы A соответственно. Для произвольной матрицы $A > 0$ будем обозначать через $A^{1/2}$ матрицу, для которой $A^{1/2} > 0$ и $A^{1/2}A^{1/2} = A$. Для характеристик матриц H_k, A_k используем результат из [21], справедливый для произвольных параметров α_k, β_k, удовлетворяющих условию (7).

Лемма 2 [21]. *Пусть $H_k > 0$, матрица H_{k+1} получена в результате (6), где параметры α_k, β_k удовлетворяют условию (7), а для произвольных векторов $y_k \neq 0$ и $p_k \neq 0$ выполняется равенство (14). Тогда $H_{k+1} > 0$ и*

$$A_{k+1} = A_k + \left(\alpha_k^2 - 1\right)\frac{y_k y_k^T}{(y_k, H_k y_k)} + \left(\beta_k^2 - 1\right)\frac{p_k p_k^T}{(p_k, H_k p_k)}, \tag{19}$$

$$\operatorname{tr} A_{k+1} = \operatorname{tr} A_k + \left(\alpha_k^2 - 1\right)\frac{(y_k, y_k)}{(y_k, H_k y_k)} + \left(\beta_k^2 - 1\right)\frac{(p_k, p_k)}{(p_k, H_k p_k)}, \tag{20}$$

$$\det H_{k+1} = \frac{\det H_k}{\alpha_k^2 \beta_k^2}, \quad \det A_{k+1} = \alpha_k^2 \beta_k^2 \det A_k. \tag{21}$$

Изложим идею обоснования алгоритма $\mathrm{A}(\alpha_k, \beta_k)$. В силу (9) и неравенства Шварца для произвольных $g_k \in G$ и $H_k > 0$ получим

$$1 \leqslant (s^*, g_k^*) \leqslant \left(s^*, A_k^{1/2} H_k^{1/2} g_k\right)^2 \leqslant (s^*, A_k s^*)(g_k H_k g_k). \tag{22}$$

Предположим, что шаг 2 алгоритма 1 создаёт вектор $u_k \in G$, удовлетворяющий (11). Предположим, что в этом случае для последовательностей $\{g_k\}$ и $\{H_k\}$, генерируемых алгоритмом 1, существует область параметров α_k, β_k при которых $(s^*, A_k s^*)(g_k H_k g_k)$ убывает и через конечное число итераций станет меньше 1. Однако в силу (22) это невозможно. Следовательно, через конечное число итераций на шаге 2 невозможно будет найти вектор $u_k \in G$, удовлетворяющий (11), т. е. будет найдено решение $H_k g_k \in S(G)$ системы (2).

В лемме 3 получена оценка скорости убывания величин $(g_k, H_k g_k)$ в (22).

Лемма 3. *Пусть множество G удовлетворяет предположению 1, при этом $\theta(M_G) \leqslant \theta_A$, а последовательность $\{\pi_k = \min\limits_{0 \leqslant j \leqslant k-1} (g_j, H_j g_j)\}$ вычисляется на основе характеристик алгоритма $\mathrm{A}(\alpha_k, \beta_k)$. Тогда*

$$\pi_k \leqslant \frac{4kR_G^2(q^2\alpha^2 - 1)}{n[(\alpha^2\beta^2)^{k/n} - 1]}, \quad k \geqslant 1. \tag{23}$$

Исследуем поведение сомножителя $(s^*, A_k s^*)$ в (22). Из (19) получим

$$(s^*, A_{k+1} s^*) = (s^*, A_k s^*) + \left(\alpha_k^2 - 1\right)\frac{(s^*, y_k)^2}{(y_k, H_k y_k)} + \left(\beta_k^2 - 1\right)\frac{(s^*, p_k)^2}{(p_k, H_k p_k)}. \tag{24}$$

Для некоторого фиксированного k обозначим

$$\tilde{s}^* = A_k^{1/2} s^*, \quad \tilde{v} = A_k^{1/2} v_k = H_k^{1/2} p_k,$$
$$r_1 = H_k^{1/2} g_k, \quad r_2 = H_k^{1/2} u_k, \quad z = H_k^{1/2} y_k.$$

В дальнейшем P — плоскость, образованная векторами r_1 и r_2, $\tilde{s}$ — проекция вектора $\tilde{s}^*$ на плоскость P, $\widetilde{G} = P \cap \{H_k^{1/2} g \mid g \in G\}$, φ — угол, образованный векторами $\tilde{s}, \tilde{v}$. Из (24) получим

$$(s^*, A_{k+1} s^*) = (s^*, A_k s^*) \\ + \|\tilde{s}\|^2 \big[\big(\alpha_k^2 - 1\big) \sin^2 \varphi + \big(\beta_k^2 - 1\big)(1 - \sin^2 \varphi)\big]. \quad (25)$$

Основываясь на свойствах множества G, с использованием характеристик алгоритма $\mathrm{A}(\alpha_k, \beta_k)$ получим две оценки для $\sin^2 \varphi$.

Лемма 4. *Если множество G удовлетворяет предположению 1, то*

$$\sin^2 \varphi \leqslant \theta(M_G) = \frac{(M_G - 1)^2}{(M_G + 1)^2}. \quad (26)$$

Вторая оценка для $\sin^2 \varphi$ имеет вид $\sin^2 \varphi \leqslant \theta_{gk}(M_G)$, где зависимость $\theta_{gk}(M)$ определена в (13). На основании этих оценок получается

Лемма 5. *Если множество G удовлетворяет предположению 1, то*

$$\sin^2 \varphi \leqslant \min\{\theta(M_G), \theta_{gk}(M_G)\}, \quad (27)$$

где зависимость $\theta_{gk}(M)$ определена в (13).

Оценка (27) и способ задания параметров алгоритма $\mathrm{A}(\alpha_k, \beta_k)$ приводят к следующему результату.

Лемма 6. *Если множество G удовлетворяет предположению 1 и при этом $\theta(M_G) \leqslant \theta_A$, то для последовательности $\{A_k\}$, генерируемой алгоритмом $\mathrm{A}(\alpha_k, \beta_k)$, имеют место неравенства*

$$(s^*, A_{k+1} s^*) \leqslant (s^*, A_k s^*) \leqslant \cdots \leqslant (s^*, A_0 s^*) \leqslant \frac{1}{\rho_G^2}. \quad (28)$$

Окончательно приходим к обоснованию скорости сходимости алгоритма.

Теорема 1. *Пусть множество G удовлетворяет предположению 1 и при некоторых V_0 и θ_A выполняются ограничения (10) и*

$$0 < V_0 \leqslant V(G). \quad (29)$$

Тогда алгоритм $\mathrm{A}(\alpha_k, \beta_k)$ сходится за конечное число итераций, число которых не превосходит минимального целого k_0 из области значений k, удовлетворяющих неравенству

$$\frac{4k(q^2 \alpha^2 - 1)}{n V_0^2 [(\alpha^2 \beta^2)^{k/n} - 1]} < 1. \quad (30)$$

Обозначим $a = \alpha_k^2$, $b = \beta_k^2$, $\theta = \sin^2\varphi$ и представим выражение в квадратных скобках из (25) в виде $Q(\theta, a, b) = a\theta + b(1-\theta)$. Согласно (25) невозрастание $(s^*, A_{k+1}s^*) \leqslant (s^*, A_k s^*)$ будет достигнуто при $Q(\theta, a, b) \leqslant 0$. В этом случае ввиду (24) и (27) скорость сходимости алгоритма будет тем выше, чем больше произведение $ab = \alpha_k^2\beta_k^2$. Таким образом, пришли к задаче максимизации (16).

Для обоснования сходимости метода минимизации необходимо указать границы возмущения множества, задать параметры метода с учётом этих границ, при которых сохраняется сходимость алгоритма решения неравенств. Обозначим окрестность множества G через $S_\varepsilon(G) = \{z \in \mathbb{R}^n \mid \|z - x\| \leqslant \varepsilon$ для всех $x \in G\}$. В следующей теореме указаны ограничения на параметры отделимого множества и границы его возмущений, при которых сохраняется сходимость метода решения неравенств.

Теорема 2. *Пусть множество G удовлетворяет предположению 1 и при некоторых V_0 и θ_A выполняются ограничения (10) и (29). Тогда на множестве $S_\varepsilon(G)$ при $\varepsilon \leqslant \rho_G\Delta/2$, где $\Delta = \theta_A - \theta_G$, алгоритм $\mathrm{A}(\alpha_k, \beta_k)$ сходится за конечное число итераций, число которых не превосходит минимального целого k_0 из области значений k, удовлетворяющих неравенству*

$$\frac{25k(q^2\alpha^2 - 1)}{nV_0^2[(\alpha^2\beta^2)^{k/n} - 1]} < 1. \tag{31}$$

3. Субградиентный метод минимизации

Дадим описание метода минимизации со встроенным алгоритмом решения систем неравенств $\mathrm{RA}(\alpha_k, \beta_k)$.

Введём обозначения: $D(z) = \{x \in \mathbb{R}^n \mid f(x) \leqslant f(z)\}$, k_i — индексы k, при которых происходит обновление на шагах 6 или 8, $i = 1, 2, \ldots$, z_i — точки x_{k_i}, x_* — точка минимума функции, x^* — предельные точки последовательности $\{z_i\}_{i=1}^{\infty}$.

Теорема 3. *Пусть множество $D(x_0)$ ограничено, функция $f(x)$ строго выпукла на $\mathbb{R}^n$ и выполнены следующие ограничения:*

$$\theta(M(\partial f(x))) < \theta_A < \frac{1}{2}, \tag{32}$$

$$V(\partial f(x)) \geqslant V_0 > 0 \tag{33}$$

на $D(x_0)$ при $x \neq x_*$ и некоторых V_0, θ_A. Тогда при периоде обновления $N \geqslant N_0$ алгоритма $\mathrm{RA}(\alpha_k, \beta_k)$, где N_0 — удвоенное минимальное целое из области значений k, удовлетворяющих неравенству (31), любая предельная точка последовательности $\{z_i\}$ является точкой минимума на $\mathbb{R}^n$.

Алгоритм 2. $\mathrm{RA}(\alpha_k, \beta_k)$

Вход: начальное приближение матрицы $H_0 = I$, текущее приближение минимума $x_0 \in \mathbb{R}^n$, целые $k = 0$, $m_0 = 0$, период обновления N, параметр $\theta_A < 1/2$.

Выход: множество точек x_k, $k = 1, 2, \dots$.

1: Вычислить $g_0 \in \partial f(x_0)$.
2: Если $g_0 = 0$, то x_0 — точка минимума. Закончить вычисления.
3: Вычислить новое приближение $x_{k+1} = x_k - \gamma_k s_k$, где $s_k = H_k g_k$ и γ_k находится линейным поиском из [20].
4: Вычислить субградиент $u_k \in \partial f(x_{k+1})$, исходя из условия $(u_k, s_k) \leqslant 0$.
5: Если $u_k = 0$, то x_{k+1} — точка минимума. Закончить вычисления.
6: Если $k - m_k \geqslant N$, то положить $m_{k+1} = k$, $g_{k+1} = u_k$, $H_{k+1} = I$. Перейти на шаг 12.
7: Вычислить значения векторов y_k, p_k согласно (12).
8: Если $p_k = 0$, то положить $m_{k+1} = k$, $g_{k+1} = u_k$, $H_{k+1} = I$. Перейти на шаг 12.
9: Вычислить $m_{k+1} = m_k$, $C_k = \min\{|(y_k, H_k g_k)|, |(y_k, H_k u_k)|\}$, $\theta_{gk}(m(\theta_A))$ согласно (13) и параметр $\theta_k = \max\{\theta_A/q^2, \min\{\theta_{gk}(m(\theta_A)), \theta_A\}\}$.
10: Найти параметры $\alpha_k^2 = a(\theta_k)$, $\beta_k^2 = b(\theta_k)$ и рассчитать $H_{k+1} = (H_k, \alpha_k, \beta_k, y_k, p_k)$ из (6).
11: Положить $g_{k+1} = u_k$.
12: Положить $k := k + 1$ и перейти на шаг 3.

4. Результаты численного исследования

Мы реализовали алгоритм $\mathrm{RA}(\alpha_k, \beta_k)$ и сравнили полученные результаты со следующими методами:

1) РСМ с растяжением пространства в направлении субградиента (SD) [20];

2) r-алгоритм [1], реализованный в [20,21] ($\mathrm{r_{OM}}(\alpha)$) с параметром растяжения пространства $\alpha^2 = 6$.

Алгоритм $\mathrm{RA}(\alpha_k, \beta_k)$ был реализован с фиксированными значениями $\alpha^2 = 1/(2\theta_A)$, $\beta^2 = 1/(2(1 - \theta_A))$ при значении θ_A, обеспечивающем суммарное растяжение $\alpha^2\beta^2 = 1/(4\theta_A(1 - \theta_A)) = 6$. Обозначим его через $\mathrm{RA}(\alpha, \beta)$. Значение θ_A использовалось в алгоритме $\mathrm{RA}(\alpha_k, \beta_k)$ с динамическим способом выбора параметра растяжения пространства. Все алгоритмы реализованы с грубым одномерным поиском из [20]. Во всех методах функция и градиент вычислялись одновременно.

В качестве тестовых брались функции с высокой степенью вытянутости поверхностей уровня, возрастающей по мере роста размерности:

1) $f_1(x) = \sum\limits_{i=1}^{n} x_i^2 i^6$, $x_0 = (10/1, 10/2, \dots, 10/n)$, $\varepsilon = 10^{-10}$,

2) $f_2(x) = \sum_{i=1}^{n} x_i^2 (n/i)^6$, $x_0 = (10/1, 10/2, \ldots, 10/n)$, $\varepsilon = 10^{-10}$,

3) $f_3(x) = (\sum_{i=1}^{n} x_i^2 \cdot i)^r$, $x_0 = (1, 1, \ldots, 1)$, $r = 2$, $\varepsilon = 10^{-10}$,

4) $f_4(x) = \sum_{i=1}^{n} |x_i| i^3$, $x_0 = (10/1, 10/2, \ldots, 10/n)$, $\varepsilon = 10^{-4}$,

5) $f_5(x) = \sum_{i=1}^{n} x_i^2 (n/i)^4$, $x_0 = (10/1, 10/2, \ldots, 10/n)$, $\varepsilon = 10^{-10}$,

6) $f_6(x) = \sum_{i=1}^{n} |x_i| (n/i)^2$, $x_0 = (10/1, 10/2, \ldots, 10/n)$, $\varepsilon = 10^{-4}$.

Таблица 1

Результаты минимизации функций
$f_1(x),\ f_2(x),\ f_3(x),\ \varepsilon = 10^{-10}$

Метод / n	RA(α_k, β_k)	RA(α, β)	SD	$\mathrm{r_{OM}}(\alpha)$
$f_1(x)$, $x_0 = (10/1, 10/2, \ldots, 10/n)$				
100	865	999	2127	2333
200	1698	1953	4585	5244
300	2535	2919	7117	8480
400	3370	3911	9791	11773
500	4179	4882	12366	15281
600	4947	5846	15537	19073
700	5804	6821	18450	22500
800	6645	7735	21387	26096
900	7429	8657	24671	30233
1000	8291	9666	27447	34702
$f_2(x)$, $x_0 = (10/1, 10/2, \ldots, 10/n)$				
100	284	355	480	482
200	460	590	788	852
300	628	824	1063	1223
400	781	979	1307	1587
500	907	1157	1497	1900
600	1044	1360	1742	2188
700	1167	1497	1898	2512
800	1291	1637	2095	2829
900	1409	1777	2293	3101
1000	1493	1944	2555	3300
$f_3(x)$, $x_0 = (1, 1, \ldots, 1)$, $r = 2$				
200	133	112	365	295
400	164	141	395	505
600	196	164	409	702
800	215	190	421	900
1000	228	207	433	1094

Таблица 2

Результаты минимизации функции $f_4(x)$,
$x_0 = (10/1, 10/2, \ldots, 10/n)$, $\varepsilon = 10^{-4}$

Метод / n	$\mathrm{RA}(\alpha_k, \beta_k)$	$\mathrm{RA}(\alpha, \beta)$	SD	$\mathrm{r_{OM}}(\alpha)$
100	2019	2667	4214	3505
200	4307	5971	9087	8826
300	6615	9571	11144	14018
400	8940	13344	23687	19549
500	11296	16835	28037	24865
600	13678	20468	39703	31502
700	16071	25022	44573	38796
800	18604	28139	52380	44200
900	21184	32332	61631	43502
1000	23629	36094	72175	49050

В наборе тестов представлены квадратичные и кусочно линейные функции. Функции f_1 и f_2 квадратичные, где отношение минимального собственного значения к максимальному равно $1/n^6$. Отношение размаха поверхности уровня по осям координат минимального к максимальному равно $1/n^3$. У функции f_2 по сравнению с функцией f_1 плотность собственных значений выше в области малых значений. Функция f_3 гладкая с невысокой степенью разброса вытянутости поверхностей уровня. Её сложность обусловлена степенью выше квадратичной. Функция f_4 кусочно линейная. У этой функции отношение размаха поверхности уровня по осям координат минимального к максимальному равно $1/n^3$, т. е. то же самое, что и для квадратичных функций f_1 и f_2.

Представляет интерес сравнение сложности минимизации гладких и негладких функций методами негладкой оптимизации при условии идентичности их отношений размаха поверхности.

В табл. 1–3 приведено число вычислений значений функции и субградиента, затраченных на достижение необходимой точности по функции $f(x_k) - f^* \leqslant \varepsilon$.

Согласно результатам, приведённым в табл. 1, алгоритмы $\mathrm{RA}(\alpha, \beta)$ и $\mathrm{RA}(\alpha_k, \beta_k)$ существенно превосходят методы SD и $\mathrm{r_{OM}}(\alpha)$ на гладких функциях. У функции f_2 собственные значения гессиана смещены в область малых значений, что положительно сказалось на скорости сходимости субградиентных методов.

По соотношению пропорций размаха по координатным осям поверхности уровня функции f_1, f_2 и f_4 сходны. Функция f_4 является сложной для минимизации субградиентными методами. Сравнивая результаты табл. 1 и 4, можно отметить, что нет кардинальных отличий в скорости

Таблица 3

Результаты минимизации функции $f_4(x)$ с искажением субградиента

Метод / n	$\mathrm{RA}(\alpha_k, \beta_k)$	$\mathrm{RA}(\alpha, \beta)$	SD	$\mathrm{r_{OM}}(\alpha)$
100	2123	3179	5739	4777
200	4495	6732	13364	10665
300	6877	10400	20589	16889
400	9232	14325	30132	23397
500	11787	18275	35544	30015
600	14264	22474	47664	36749
700	16787	26724	54768	43589
800	19395	31030	68944	50737
900	22035	35555	78697	57817
1000	24615	39992	82490	64777

сходимости субградиентных методов SD и $\mathrm{r_{OM}}(\alpha)$ на этих функциях. Метод $\mathrm{RA}(\alpha_k, \beta_k)$ здесь существенно превосходит другие методы в скорости сходимости.

Для имитации наличия толщины субградиентного множества при минимизации функции f_4 субградиенты $g(x) \in \partial f(x)$ в процессе минимизации генерировались с помехой по формуле $g(x) \in (1 + \xi)\partial f(x)$, где $\xi \in [0, 1]$ — равномерно распределённое случайное число. Помеха отрицательно влияет как на качество одномерного поиска, так и на качество направления спуска. Результаты приведены в табл. 3. Здесь, как и на функции f_4, метод $\mathrm{RA}(\alpha_k, \beta_k)$ существенно превосходит другие методы в скорости сходимости.

Преимущество нового алгоритма сохраняется и при гораздо большей размерности и степени вырожденности функций. Табл. 4 демонстрирует сравнительные результаты для двух задач с 10000 переменных. Результаты минимизации гладкой функции f_5 даны также в сравнении с квазиньютоновским (QN) методом.

Таблица 4

Результаты минимизации для высокой размерности

Метод / n	$\mathrm{RA}(\alpha_k, \beta_k)$	$\mathrm{RA}(\alpha, \beta)$	SD	$\mathrm{r_{OM}}(\alpha)$	QN
$f_5(x)$, $x_0 = (10/1, 10/2, \ldots, 10/n)$, $\varepsilon = 10^{-10}$					
10000	3184	4065	5669	9907	4836
$f_6(x)$, $x_0 = (10/1, 10/2, \ldots, 10/n)$, $\varepsilon = 10^{-4}$					
10000	3532	5804	10858	10640	—

Стоит отметить, что преимущество нашего нового алгоритма заметно в том числе на гладких функциях (см. функцию f_5 в табл. 4). Для самой сложной негладкой функции f_6 последний столбец пуст, так как метод QN не дал результата даже после продолжительных вычислений.

Относительно скорости сходимости представленных методов можно сделать ряд выводов.

На основании результатов минимизации гладких функций f_1, f_2, f_3 можно сделать вывод о том, что субградиентные методы с растяжением пространства могут быть полезными и при минимизации гладких функций. При этом алгоритмы $\mathrm{RA}(\alpha,\beta)$ и $\mathrm{RA}(\alpha_k,\beta_k)$ показывают на гладких функциях существенно лучшие результаты, нежели другие субградиентные методы SD и $\mathrm{r_{OM}}(\alpha)$.

Метод $\mathrm{RA}(\alpha_k,\beta_k)$ существенно превосходит методы SD и $\mathrm{r_{OM}}(\alpha)$ при минимизации негладких функций.

Заключение

Задача построения эффективного РСМ для минимизации негладких функций заключалась в создании алгоритма формирования направления спуска, обеспечивающего выход из окрестности текущего минимума посредством одномерной минимизации вдоль него. В результате разработан алгоритм формирования матриц метрики с оптимальным набором параметров, превращающий широкий пучок субградиентов текущей окрестности минимума в узкий пучок направлений, образующих острый угол со всеми субградиентами окрестности текущего приближения минимума. Применение этого алгоритма позволило создать эффективный РСМ с коррекцией матриц метрики ранга 2, аналогичный по структуре квазиньютоновским методам. Задача построения матрицы метрики была поставлена как задача решения системы неравенств. Для её решения использовались формализованная модель субградиентных множеств, понятия и алгоритмы теории обучения.

Вычислительный эксперимент на сложных большеразмерных функциях подтверждает эффективность предложенного алгоритма. Алгоритмы подобного типа имеют важное прикладное значение. Возможность применения РСМ при решении негладких невыпуклых задач оптимизации даёт возможность его использования в задачах обучения математических моделей с использованием технологий подавления неинформативных переменных, подобных «лассо Тибширани» [29–31].

ЛИТЕРАТУРА

1. **Шор Н. З.** Применение метода градиентного спуска для решения сетевой транспортной задачи // Мат. науч. семинара по теор. и прикл. вопросам

кибернетики и исследования операций. Вып. 1. К.: Науч. совет по кибернетике АН УССР, 1962. 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. **Гольштейн Е. Г., Немировский А. С., Нестеров Ю. Е.** Метод уровней, его обобщения и приложения // Экономика и мат. методы. 1983. Т. 31, вып. 3. С. 164–180.
6. **Nesterov Yu. E.** Universal gradient methods for convex optimization problems // Math. Program. Ser. A. 2015. V. 152. P. 381–404.
7. **Gasnikov A. V., Nesterov Yu. E.** Universal method for stochastic composite optimization. Ithaca, NY: Cornell Univ., 2016. (Cornell Univ. Libr. e-Print Archive; arXiv:1604.05275).
8. **Ouyang H., Gray A.** Stochastic smoothing for nonsmooth minimizations: Accelerating SGD by exploiting structure // Proc. 29th Int. Conf. Machine Learning (Edinburgh, Scotland, June 26–July 1, 2012). Madison, WI: Omnipress, 2012. P. 33–40.
9. **Boob D., Deng Q., Lan G.** Stochastic first-order methods for convex and nonconvex functional constrained optimization // Math. Program. 2022. [in print]. Available at `doi.org/10.1007/s10107-021-01742-y` (accessed June 17, 2022).
10. **Lan G.** First-order and stochastic optimization methods for machine learning. Cham: Springer, 2020.
11. **Ghadimi S., Lan G.** Accelerated gradient methods for nonconvex nonlinear and stochastic programming // Math. Program. 2016. V. 156, No. 1–2. P. 59–99.
12. **Fang C., Li C. J., Lin Z., Zhang T.** Spider: Near-optimal non-convex optimization via stochastic path-integrated differential estimator // Advances in Neural Information Processing Systems 31. 32nd Annual Conf. (Montréal, Canada, Dec. 3–8, 2018). Red Hook, NY: Curran Associates, 2018. P. 687–697.
13. **Немировский А. С., Юдин Д. Б.** Сложность задач и эффективность методов оптимизации. М.: Наука, 1979.
14. **Шор Н. З.** Методы минимизации недифференцируемых функций и их приложения. К.: Наук. думка, 1979.
15. **Cao H., Song Y., Khan K.** Convergence of subtangent-based relaxations of non-linear programs // Processes. 2019. V. 7, No. 4. P. 221.
16. **Поляк Б. Т.** Минимизация негладких функционалов // Журн. вычисл. математики и мат. физики. 1969. Т. 9, вып. 3. С. 509–521.
17. **Крутиков В. Н., Самойленко Н. С., Мешечкин В. В.** О свойствах метода минимизации выпуклых функций, релаксационного по расстоянию до экстремума // Автоматика и телемеханика. 2019. Т. 80, вып. 1. С. 126–137.
18. **Демьянов В. Ф., Васильев Л. В.** Недифференцируемая оптимизация. М.: Наука, 1981.

19. Lemarechal C. An extension of Davidon methods to non-differentiable problems // Math. Program. Study. 1975. V. 3. P. 95–109.

20. 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.

21. 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.

22. Скоков В. А. Замечание к методам оптимизации, использующим операцию растяжения пространства // Кибернетика и систем. анализ. 1974. № 4. С. 115–117.

23. Крутиков В. Н., Самойленко Н. С. О скорости сходимости субградиентного метода с изменением метрики и его приложения в схемах нейросетевых приближений // Вестн. Томск. гос. ун-та. Сер. Математика и механика. 2018. № 55. С. 22–37.

24. Nocedal J., Wright S. J. Numerical optimization. New York: Springer, 2006.

25. Avriel M. Nonlinear programming: Analysis and methods. Mineola: Dover Publ., 2003.

26. Нурминский Е. А., Тьен Д. Метод сопряжённых субградиентов с ограниченной памятью // Автоматика и телемеханика. 2014. Т. 75, вып. 4. С. 646–656.

27. Цыпкин Я. З. Основы теории обучающихся систем. М.: Наука, 1970.

28. Жуковский Е. Л., Липцер Р. Ш. О рекуррентном способе вычисления нормальных решений линейных алгебраических уравнений // Журн. вычисл. математики и мат. физики. 1972. Т. 12, вып. 4. С. 843–857.

29. Krutikov V. N., Kazakovtsev L. A., Kazakovtsev V. L. Non-smooth regularization in radial artificial neural networks // IOP Conf. Ser.: Materials Science and Engineering. 2018. V. 450, No. 4, ID 042010. 7 p.

30. Krutikov V. N., Kazakovtsev L. A., Shkaberina G. Sh., Kazakovtsev V. L. New method of training two-layer sigmoid neural networks using regularization // IOP Conf. Ser.: Materials Science and Engineering. 2019. V. 537, No. 4, ID 042055. 6 p.

31. Tibshirani R. J. Regression shrinkage and selection via the Lasso // J. Royal Stat. Soc. Ser. B (Methodological). 1996. V. 58. P. 267–288.

32. Frostig R., Ge R., Kakade S. M., Sidford A. Un-regularizing: Approximate proximal point and faster stochastic algorithms for empirical risk minimization // Proc. Mach. Learn. Res. 2015. V. 37. P. 2540–2548.

Крутиков Владимир Николаевич
Станимирович Предраг
Инденко Оксана Николаевна
Товбис Елена Михайловна
Казаковцев Лев Александрович

Статья поступила
10 мая 2022 г.
После доработки —
10 мая 2022 г.
Принята к публикации
12 мая 2022 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
/DISCRETE ANALYSIS AND OPERATIONS RESEARCH/
July–September 2022. Vol. 29, No. 3. P. 24–44

UDC 519.8 DOI 10.33048/daio.2022.29.739

OPTIMIZATION OF SUBGRADIENT METHOD PARAMETERS ON THE BASE OF RANK-TWO CORRECTION OF METRIC MATRICES

V. N. Krutikov[1, a], *P. S. Stanimirović*[2, b], *O. N. Indenko*[1, c], *E. M. Tovbis*[3, d], and *L. A. Kazakovtsev*[3, e]

[1] Kemerovo State University,
6 Krasnaya Street, 650043 Kemerovo, Russia

[2] Faculty of Sciences and Mathematics, University of Niš,
33 Višegradska Street, 18000 Niš, Serbia

[3] Reshetnev Siberian State University of Science and Technology,
31 Krasnoyarskiy Rabochiy Avenue, 660031 Krasnoyarsk, Russia

E-mail: [a] `krutikovvn@rambler.ru`, [b] `pecko@pmf.ni.ac.rs`,
[c] `oksana230805@mail.ru`, [d] `sibstu2006@rambler.ru`, [e] `levk@bk.ru`

Abstract. We establish a relaxation subgradient method (RSM) that includes parameter optimization utilizing metric rank-two correction matrices with a structure analogous to quasi-Newtonian (QN) methods. The metric matrix transformation consists of suppressing orthogonal and amplifying collinear components of the minimal length subgradient vector. The problem of constructing a metric matrix is formulated as a problem of solving an involved system of inequalities. Solving such system is based on a new learning algorithm. An estimate for its convergence rate is obtained depending on the parameters of the subgradient set. A new RSM has been developed and investigated on this basis. Computational experiments on complex large-scale functions confirm the effectiveness of the proposed algorithm. Tab. 4, bibliogr. 32.

Keywords: convex optimization, nonsmooth optimization, relaxation subgradient method.

This research is supported by the Ministry of Science and Higher Education of Russia (State Contract FEFE–2020–0013). The work of the second author is supported by the Science Foundation of the Republic of Serbia (Grant 7750185) and the Ministry of Education, Science and Technological Development of the Republic of Serbia (Contract 451–03–68/2020–14/200124).

English version: Journal of Applied and Industrial Mathematics **16** (3) (2022).

REFERENCES

1. **N. Z. Shor,** Application of the gradient descent method for solving network transportation problems, in *Proc. Scientific Seminar on Theoretic and Applied Problems of Cybernetics and Operations Research,* Issue 1 (Nauch. sovet po kibernetike AN USSR, Kyev, 1962), pp. 9–17 [Russian].
2. **B. T. Polyak,** A general method for solving extremal problems, *Dokl. Akad. Nauk SSSR* **174** (1), 33–36 (1967) [Russian].
3. **B. T. Polyak,** *Introduction to Optimization* (Nauka, Moscow, 1983) [Russian].
4. **P. Wolfe,** Note on a method of conjugate subgradients for minimizing nondifferentiable functions, *Math. Program.* **7** (1), 380–383 (1974).
5. **E. G. Gol'shtein, A. S. Nemirovskii,** and **Yu. E. Nesterov,** Level method, its generalizations and applications, *Ekon. Mat. Metody* **31** (3), 164–180 (1983) [Russian].
6. **Yu. E. Nesterov,** Universal gradient methods for convex optimization problems, *Math. Program., Ser. A,* **152**, 381–404 (2015).
7. **A. V. Gasnikov** and **Yu. E. Nesterov,** *Universal method for stochastic composite optimization* (Cornell Univ., Ithaca, NY, 2016) (Cornell Univ. Libr. e-Print Archive, arXiv:1604.05275).
8. **H. Ouyang** and **A. Gray,** Stochastic smoothing for nonsmooth minimizations: Accelerating SGD by exploiting structure, in *Proc. 29th Int. Conf. Machine Learning, Edinburgh, Scotland, June 26–July 1, 2012* (Omnipress, Madison, WI, 2012), pp. 33–40.
9. **D. Boob, Q. Deng,** and **G. Lan,** Stochastic first-order methods for convex and nonconvex functional constrained optimization // Math. Program. 2022. [in print]. Available at `doi.org/10.1007/s10107-021-01742-y` (accessed June 17, 2022).
10. **G. Lan,** *First-Order and Stochastic Optimization Methods for Machine Learning* (Springer, Cham, 2020).
11. **S. Ghadimi** and **G. Lan,** Accelerated gradient methods for nonconvex nonlinear and stochastic programming, *Math. Program.* **156** (1–2), 59–99 (2016).
12. **C. Fang, C. J. Li, Z. Lin,** and **T. Zhang,** Spider: Near-optimal non-convex optimization via stochastic path-integrated differential estimator, in *Advances in Neural Information Processing Systems 31* (32nd Annual Conf., Montréal, Canada, Dec. 3–8, 2018) (Curran Associates, Red Hook, NY, 2018), pp. 687–697.
13. **A. S. Nemirovskii** and **D. B. Yudin,** *Complexity of Problems and Efficiency of Methods in Optimization* (Nauka, Moscow, 1979) [Russian].
14. **N. Z. Shor,** *Minimization Methods for Non-differentiable Functions and Their Applications* (Nauk. Dumka, Kyev, 1979) [Russian].
15. **H. Cao, Y. Song,** and **K. Khan,** Convergence of subtangent-based relaxations of non-linear programs, *Processes* **7** (4), 221 (2019).
16. **B. T. Polyak,** Minimization of nonsmooth functionals, *Zh. Vychisl. Mat. Mat. Fiz.* **9** (3), 509–521 (1969) [Russian] [*Comput. Math. Math. Phys.* **9** (3), 14–29 (1969)].

17. **V. N. Krutikov, N. S. Samoilenko,** and **V. V. Meshechkin,** On the properties of the method of minimization for convex functions with relaxation on the distance to extremum, *Avtom. Telemekh.,* No. 1, 126–137 (2019) [Russian] [*Autom. Remote Control* **80** (1), 102–111 (2019)].
18. **V. F. Demyanov** and **L. V. Vasilyev,** *Non-differentiable Optimization* (Nauka, Moscow, 1981) [Russian].
19. **C. Lemarechal,** An extension of Davidon methods to non-differentiable problems, *Math. Program. Study* **3**, 95–109 (1975).
20. **V. N. Krutikov** and **T. V. Petrova,** Relaxation method of minimization with space extension in the subgradient direction, *Ekon. Mat. Metody* **39** (1), 106–119 (2003) [Russian].
21. **V. N. Krutikov** and **T. A. Gorskaya,** A family of subgradient relaxation methods with rank 2 correction of metric matrices, *Ekon. Mat. Metody* **45** (4), 37–80 (2009) [Russian].
22. **V. A. Skokov,** Note on minimization methods employing space stretching, *Kibern. Sist. Anal.,* No. 4, 115–117 (1974) [Russian] [*Cybern. Syst. Anal.* **10** (4), 689–692 (1974)].
23. **V. N. Krutikov** and **N. S. Samoilenko,** On the convergence rate of the subgradient method with metric variation and its applications in neural network approximation schemes, *Vestn. Tomsk. Gos. Univ., Ser. Mat. Mekh.,* No. 55, 22–37 (2018) [Russian].
24. **J. Nocedal** and **S. J. Wright,** *Numerical Optimization* (Springer, New York, 2006).
25. **M. Avriel,** *Nonlinear Programming: Analysis and Methods* (Dover Publ., Mineola, 2003).
26. **E. A. Nurminskii** and **D. Tien,** Method of conjugate subgradients with constrained memory, *Avtom. Telemekh.,* No. 4, 67–80 (2014) [Russian] [*Autom. Remote Control* **75** (4), 646–656 (2014)].
27. **Ya. Z. Tsypkin,** *Basics of Theory of Learning Systems* (Nauka, Moscow, 1970) [Russian].
28. **E. L. Zhukovskii** and **R. Sh. Liptser,** A recurrence method for computing the normal solutions of linear algebraic equations, *Zh. Vychisl. Mat. Mat. Fiz.* **12** (4), 843–857 (1972) [Russian] [*Comput. Math. Math. Phys.* **12** (4), 1–18 (1972)].
29. **V. N. Krutikov, L. A. Kazakovtsev,** and **V. L. Kazakovtsev,** Nonsmooth regularization in radial artificial neural networks, *IOP Conf. Ser.: Materials Science and Engineering* **450** (4), ID 042010. 7 p. (2018).
30. **V. N. Krutikov, L. A. Kazakovtsev, G. Sh. Shkaberina,** and **V. L. Kazakovtsev,** New method of training two-layer sigmoid neural networks using regularization, *IOP Conf. Ser.: Materials Science and Engineering* **537** (4), ID 042055. 6 p. (2019).
31. **R. J. Tibshirani,** Regression shrinkage and selection via the Lasso, *J. Royal Stat. Soc., Ser. B (Methodological)* **58**, 267–288 (1996).

32. **R. Frostig, R. Ge, S. M. Kakade,** and **A. Sidford,** Un-regularizing: Approximate proximal point and faster stochastic algorithms for empirical risk minimization, *Proc. Mach. Learn. Res.* **37**, 2540–2548 (2015).

Vladimir N. Krutikov
Predrag S. Stanimirović
Oksana N. Indenko
Elena M. Tovbis
Lev A. Kazakovtsev

Received May 10, 2022
Revised May 10, 2022
Accepted May 12, 2022