

УДК 519.6

АЛГОРИТМЫ ОБУЧЕНИЯ НА ОСНОВЕ ОРТОГОНАЛИЗАЦИИ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ВЕКТОРОВ

Я. Н. Вершинин, В. Н. Крутиков

ORTHOGONALIZATION OF THE SEQUENCE VECTORS LEARNING ALGORITHMS

Y. N. Vershinin, V. N. Krutikov

Эффективные релаксационные субградиентные методы оптимизации негладких функций основываются на алгоритмах решения систем неравенств, целью которых является поиск направления спуска на очередной итерации основного алгоритма. Для нахождения решения указанных систем неравенств могут быть использованы итеративные алгоритмы, называемые алгоритмами обучения. В этой работе предлагаются и изучаются многошаговые методы, в которых для ускорения сходимости используется идея ортогонализации последовательных направлений процесса обучения.

Effective relaxation subgradient optimization methods for nonsmooth functions are based on algorithms for solving systems of inequalities, which are focused on computing the descent direction for the next iteration of the main algorithm. Iterative algorithms, called learning algorithms, can be used to find the solution of such inequality systems. In this paper we propose and study multi-step methods which use orthogonalization of sequential directions to accelerate learning process convergence.

Ключевые слова: алгоритм Качмажа, алгоритм с попарной ортогонализацией векторов, многошаговый алгоритм, скорость сходимости.

Keywords: Kaczmarz algorithm, pairwise orthogonalization of sequential vectors algorithm, multistep algorithm, convergence rate.

1. Постановка задачи

Пусть требуется решить задачу оценивания неизвестных параметров линейной модели.

Задача 1. По данным наблюдения $g_k \in R^n$, $y_k \in R^1$, $k = 0, 1, 2, \dots$, процесса

$$y_k = (s^*, g_k) + \xi_k, \quad (1)$$

где ξ_k – ошибка наблюдения, необходимо найти вектор неизвестных параметров $s^* \in R^n$.

Здесь и далее n – размерность пространства, а $(\cdot, \cdot)$ – оператор скалярного произведения. Соотношение (1) будем называть обучающим соотношением. Можно сказать, что целью решения задачи 1 является нахождение $s^* \in R^n$, в определенной мере удовлетворяющего обучающим соотношениям.

Алгоритм Качмажа [1] (см., например, [2 – 3]) для оценки s^* в задаче 1 имеет вид [4–6]:

$$s_{k+1} = s_k + \frac{y_k - (s_k, g_k)}{\|g_k\|^2} g_k, \quad (2)$$

где $\|\cdot\|$ – Евклидова норма вектора.

Итерацию (2) при $\xi_k = 0$ можно интерпретировать как шаг минимизации функции

$$\phi(s) = (s - s^*, s - s^*)$$

из точки s_k в направлении g_k , т. е.:

$$s_{k+1} = s_k - t^* g_k, \quad t^* = \frac{(g_k, s_k) - (g_k, s^*)}{(g_k, g_k)},$$

где t^* – оптимальный шаг. Обозначим вектор невязки, как $\Delta_k = s_k - s^*$. Мера качества направле-

ния спуска g_k минимизации квадратичной функции $\phi(s)$ с равными собственными значениями матрицы вторых производных определяется выражением

$$q_k^2 = (\Delta_k, g_k)^2 / \|\Delta_k\|^2 \|g_k\|^2. \quad (3)$$

Вместо (g_k, s^*) мы имеем наблюдения y_k . Поэтому реальный шаг в (2)

$$t = [(g_k, s_k) - y_k] / (g_k, g_k).$$

В качестве меры ошибки наблюдений y_k примем относительную ошибку

$$\lambda_k = (t - t^*) / t^* = [(g_k, s_k) - y_k] / [(g_k, s_k) - (g_k, s^*)] - 1. \quad (4)$$

Сделаем следующее предположение о данных задачи 1.

Предположение 1. Пусть данные в задаче 1 удовлетворяют следующим условиям:

$$q_k^2 \geq q^2 > 0. \quad (5)$$

$$0 \leq -\lambda_k \leq \lambda < 1. \quad (6)$$

$$(g_k, \Delta_k) \geq 0. \quad (7)$$

Ограничение (6) на λ_k обеспечивает гарантированную точность одномерного спуска алгоритма (2) вдоль g_k и одновременно соответствует неполной релаксации одномерного спуска, т. е. отсутствию проскакивания точки одномерного минимума.

Свойство (7) говорит о согласованности векторов невязки Δ_k и обучения g_k . Вектор g_k в алгоритме обучения (2) играет роль, аналогичную градиенту в градиентном методе. При условии (7) справедлив более сильный аналог (5):

$$1 \geq q_k = \frac{(\Delta_k, g_k)}{\|\Delta_k\| \|g_k\|} \geq q > 0, q_k^2 \geq q^2 > 0. \quad (8)$$

Далее мы будем использовать равенство

$$(g_k, s_k) - y_k = (1 + \lambda_k)(g_k, \Delta_k), \quad (9)$$

которое следует из (4).

Исследуемые в данной работе алгоритмы строятся на принципе ортогонализации последовательных векторов процесса обучения. Выбор этого принципа обусловлен тем, что процесс минимизации квадратичной функции $\phi(s)$ с одинаковыми собственными значениями матрицы вторых производных гарантированно сходится к стационарной точке не более чем за n итераций при условии взаимной ортогональности направлений движения.

2. Оценка скорости сходимости алгоритма Качмажа

На каждом шаге алгоритма Качмажа оценка новых параметров производится так, что выполняются обучающие соотношения (1).

Лемма 1. Пусть последовательность $\{s_k\}$ – результат процесса (2) решения задачи 1, $\|g_k\| \neq 0$ для $k = 0, 1, 2, \dots$, $s_0 \in R^n$. Тогда выполнены равенства

$$(s_{k+1}, g_k) = y_k. \quad (10)$$

Доказательство. Умножив (2) на g_k , получим:

$$(s_{k+1}, g_k) = (s_k, g_k) + \frac{y_k - (s_k, g_k)}{\|g_k\|^2} (g_k, g_k) = y_k.$$

Теорема 1. Пусть последовательность $\{s_k\}$ – результат процесса (2) решения задачи 1, $\|g_k\| \neq 0$ для $k = 0, 1, 2, \dots$, $s_0 \in R^n$. Тогда, если выполнено предположение 1, имеет место оценка:

$$(\Delta_k, \Delta_k) \leq (\Delta_0, \Delta_0) [1 - q^2 (1 - \lambda^2)]^k. \quad (11)$$

Доказательство. Вычитая из обеих частей (2) вектор s^* , с учетом (9), получим:

$$\Delta_{k+1} = \Delta_k - \frac{(1 + \lambda_k)(g_k, \Delta_k)}{(g_k, g_k)} g_k. \quad (12)$$

Отсюда найдем:

$$(\Delta_{k+1}, \Delta_{k+1}) = (\Delta_k, \Delta_k) - \frac{(g_k, \Delta_k)^2}{(g_k, g_k)} (1 - \lambda_k^2). \quad (13)$$

Из (13), с учетом (3), получим:

$$(\Delta_{k+1}, \Delta_{k+1}) = (\Delta_k, \Delta_k) [1 - q^2 (1 - \lambda_k^2)]. \quad (14)$$

При условиях (5) и (6) из (14) следует оценка:

$$(\Delta_{k+1}, \Delta_{k+1}) \leq (\Delta_k, \Delta_k) [1 - q^2 (1 - \lambda^2)]. \quad (15)$$

Для обоснования (11) используем рекуррентно оценку (15).

3. Алгоритм обучения на основе попарной ортогонализации последовательных векторов

Если в (2) используется система из n взаимно ортогональных векторов g_k , и отсутствует ошибка,

т. е. $y_k = (s^*, g_k)$, g_k тогда решение $s^* \in R^n$ будет получено за цикл итераций (2) вдоль этих векторов. Мы, используя ортогонализацию, построим алгоритм, который решает задачу точно в пространстве двух последовательных векторов.

Введем итерационный процесс с ортогонализацией последовательных векторов:

$$s_{k+1} = s_k + \frac{y_k - (s_k, g_k)}{(p_k, g_k)} p_k, \quad (16)$$

где

$$p_k = \begin{cases} g_k, \\ g_k - \frac{(g_k, g_{k-1})}{\|g_{k-1}\|^2} g_{k-1}, \end{cases} \quad (17)$$

если $(g_k, g_{k-1}) \geq 0$,
если $(g_k, g_{k-1}) < 0$.

На каждом шаге алгоритма (16), (17) оценка новых параметров производится так, что выполняются обучающие соотношения (1).

Лемма 2. Пусть последовательность $\{s_k\}$ – результат процесса (16), (17) решения задачи 1, $\|g_k\| \neq 0$ для $k = 0, 1, 2, \dots$, $s_0 \in R^n$. Тогда выполнены равенства

$$(s_{k+1}, g_k) = y_k, \quad (18)$$

$$(s_{k+1}, g_{k-1}) = y_{k-1}, \text{ если } (g_k, g_{k-1}) < 0. \quad (19)$$

Доказательство. Умножив (16) на g_k , получим:

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

И если $(g_k, g_{k-1}) < 0$, то с учетом $(p_k, g_{k-1}) = 0$, что следует из (17), получим:

$$(s_{k+1}, g_{k-1}) = (s_{k-1}, g_{k-1}) + \frac{y_{k-1} - (s_{k-1}, g_{k-1})}{(p_{k-1}, g_{k-1})} (p_{k-1}, g_{k-1}) = y_{k-1}.$$

Согласно (18), (19), в том случае, когда в (17) срабатывает ортогонализация, выполняются два последних обучающих соотношения.

Теорема 2. Пусть последовательность $\{s_k\}$ – результат процесса (16), (17) решения задачи 1, $\|g_k\| \neq 0$ для $k = 0, 1, 2, \dots$, $s_0 \in R^n$. Тогда, если выполнено предположение 1, имеет место оценка

$$(\Delta_k, \Delta_k) \leq (\Delta_0, \Delta_0) [1 - q^2 (1 - \lambda^2)]^k. \quad (20)$$

При этом на отдельной итерации оценка скорости сходимости процесса (16), (17) не хуже, чем

у алгоритма Качмажа при условии шага каждого из алгоритмов из одной и той же точки s_k .

Доказательство. Исследование случая $(g_k, g_{k-1}) \geq 0$ в (17) нами проведено при оценке скорости сходимости алгоритма Качмажа. Поэтому мы исследуем случай ортогонализации в (17). Вы-

читая из обеих частей (16) вектор s^* , с учетом (9), получим:

$$\Delta_{k+1} = \Delta_k - \frac{(1 + \lambda_k)(g_k, \Delta_k)}{(p_k, g_k)} p_k. \quad (21)$$

Отсюда найдем:

$$\begin{aligned} (\Delta_{k+1}, \Delta_{k+1}) &= (\Delta_k, \Delta_k) - 2 \frac{(1 + \lambda_k)(p_k, \Delta_k)(g_k, \Delta_k)}{(p_k, g_k)} + \frac{(1 + \lambda_k)^2 (g_k, \Delta_k)^2 (p_k, p_k)}{(p_k, g_k)^2} = \\ &= (\Delta_k, \Delta_k) - \frac{(1 + \lambda_k)(g_k, \Delta_k)}{(p_k, g_k)} \left[2(p_k, \Delta_k) - \frac{(1 + \lambda_k)(g_k, \Delta_k)(p_k, p_k)}{(p_k, g_k)} \right]. \end{aligned} \quad (22)$$

В силу $(p_k, g_{k-1}) = 0$, что следует из (17), выполняется равенство $(p_k, p_k) = (p_k, g_k)$. Поэтому (22) приводится к виду:

$$(\Delta_{k+1}, \Delta_{k+1}) = (\Delta_k, \Delta_k) - \frac{(1 + \lambda_k)(g_k, \Delta_k)}{(p_k, g_k)} [2(p_k, \Delta_k) - (1 + \lambda_k)(g_k, \Delta_k)]. \quad (23)$$

Умножив (21) на g_k , с учетом (7), получим:

$$(g_k, \Delta_{k+1}) = -\lambda_k (g_k, \Delta_k) \geq 0. \quad (24)$$

Умножив (17) на Δ_k , с учетом $(g_k, g_{k-1}) < 0$ и неравенства (24) для индекса $k \rightarrow k-1$, получим оценку:

$$\begin{aligned} (p_k, \Delta_k) &= (g_k, \Delta_k) - \\ &- \frac{(g_k, g_{k-1})(g_{k-1}, \Delta_k)}{(g_{k-1}, g_{k-1})} \geq (g_k, \Delta_k). \end{aligned} \quad (25)$$

При условии, что $(p_k, g_k) > 0$ из (23), с учетом (25) и (7) получим:

$$\begin{aligned} (\Delta_{k+1}, \Delta_{k+1}) &\leq (\Delta_k, \Delta_k) - \\ &- \frac{(1 - \lambda_k^2)(g_k, \Delta_k)^2}{(p_k, g_k)}. \end{aligned} \quad (26)$$

Умножив (17) на g_k , получим:

$$\begin{aligned} (p_k, g_k) &= (g_k, g_k) - \\ &- \frac{(g_k, g_{k-1})^2}{(g_{k-1}, g_{k-1})} \leq (g_k, g_k). \end{aligned} \quad (27)$$

Рассмотрим положение векторов g_k и $\hat{\Delta}_k$ – проекции вектора Δ_k в плоскости ортогональных векторов g_{k-1} , p_k . Согласно (17), вектор g_k имеет отрицательную компоненту вдоль g_{k-1} . Согласно неравенству (24) для индекса k равного $k-1$, вектор $\hat{\Delta}_k$ имеет положительную компоненту вдоль g_{k-1} . Поэтому угол между g_k , $\hat{\Delta}_k$ больше угла между векторами g_k , p_k . Следовательно, для косинусов этих углов можно записать цепочку неравенств:

$$\begin{aligned} \frac{(p_k, g_k)}{\|p_k\| \|g_k\|} &\geq \frac{(\hat{\Delta}_k, g_k)}{\|\hat{\Delta}_k\| \|g_k\|} \geq \\ &\geq \frac{(\Delta_k, g_k)}{\|\Delta_k\| \|g_k\|} \geq q_k. \end{aligned} \quad (28)$$

Отсюда и равенства

$$\|p_k\| = \sqrt{(p_k, p_k)} = \sqrt{(g_k, p_k)} \text{ следует } \min_l \quad (29)$$

Из (29), вместе с (24) имеем:

$$\begin{aligned} (g_k, g_k) &\geq (p_k, g_k) \geq \\ &\geq q_k^2 (g_k, g_k) > 0. \end{aligned} \quad (30)$$

Таким образом, в силу (30), оценка (26) получена корректно, а на основании (24) может быть преобразована к неравенству:

$$(\Delta_{k+1}, \Delta_{k+1}) \leq (\Delta_k, \Delta_k) - \frac{(1 - \lambda_k^2)(g_k, \Delta_k)^2}{(g_k, g_k)}. \quad (31)$$

При условиях (5) и (6) из (31) следует оценка:

$$(\Delta_{k+1}, \Delta_{k+1}) \leq (\Delta_k, \Delta_k) [1 - q^2 (1 - \lambda^2)]. \quad (32)$$

Для обоснования (20) используем рекуррентно оценку (32).

Сравнивая оценки (26) и (13) приходим к выводу, что на отдельной итерации при условии, что оба алгоритма ((2) и (16), (17)) делают шаг из одной точки, в силу неравенства в (26) и равенства в (13), следует последнее утверждение теоремы о том, что на отдельной итерации оценка скорости сходимости процесса (16), (17) не хуже, чем у алгоритма Качмажа.

4. Многошаговый алгоритм обучения

Развивая идею алгоритма с парной ортогонализацией последовательных векторов обучения, введем итерационный процесс с ортогонализацией,

следующих друг за другом направлений смещения, т. е. $(p_k, p_{k-1}) = 0$.

$$s_{k+1} = s_k + \frac{y_k - (s_k, g_k)}{(p_k, g_k)} p_k, \quad (33)$$

где

$$p_k = \begin{cases} g_k, & \text{если } (g_k, p_{k-1}) \geq 0, \\ g_k - \frac{(g_k, p_{k-1})}{\|p_{k-1}\|^2} p_{k-1}, & \text{если } (g_k, p_{k-1}) < 0. \end{cases} \quad (34)$$

На каждом шаге алгоритма оценка новых параметров производится так, что выполняются обучающие соотношения (1). Поведение исследуемого алгоритма представлено на рисунке 1.

Лемма 3. Пусть последовательность $\{s_k\}$ – результат процесса (33), (34) решения задачи 1, $\|g_k\| \neq 0$ для $k = 0, 1, 2, \dots$, $s_o \in R^n$. Тогда выполнены равенства

$$(s_{k+1}, g_k) = y_k. \quad (35)$$

Доказательство. Используя (33) и (34), получим:

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

Для оценки скорости сходимости нам потребуется следующая лемма, которая исследует характеристики процесса (33), (34) в случае $(g_k, p_{k-1}) < 0$.

Лемма 4. Пусть последовательность $\{s_k\}$ – результат процесса (33), (34) решения задачи 1, $\|g_k\| \neq 0$ для $k = 0, 1, 2, \dots$, $s_o \in R^n$. Тогда, если выполнено предположение 1, то имеют место оценки:

$$(p_k, \Delta_k) \geq (g_k, \Delta_k). \quad (36)$$

$$(p_k, \Delta_{k+1}) \geq 0. \quad (37)$$

Доказательство. Будем проводить доказательство по индукции. При $k = 0$ выполняются равенства $p_k = g_k$, $p_{k-1} = 0$. Поэтому неравенство (36)

выполнено. Вычитая из обеих частей (33) вектор s^* , с учетом (9), получим:

$$k = 1, \dots, q \quad (38)$$

Умножив (38) на p_k и произведя преобразования, получим:

$$(p_k, \Delta_{k+1}) = (p_k, \Delta_k) - \frac{(1 + \lambda_k)(g_k, \Delta_k)}{(p_k, g_k)} (p_k, p_k) = (p_k, \Delta_k) - (1 + \lambda_k)(g_k, \Delta_k) \geq -\lambda_k(g_k, \Delta_k) \geq 0. \quad (39)$$

Переходы в (39) сделаны на основе равенства $(p_k, p_k) = (p_k, g_k)$, следующего из (34) в случае $(g_k, p_{k-1}) < 0$, и неравенства (36), позволившего сделать замену $(p_k, \Delta_k) \rightarrow (g_k, \Delta_k)$.

В предположении, что неравенства выполнены на предыдущей итерации $k - 1$, докажем их справедливость на текущей итерации k следующим образом. Для случая, $(g_k, p_{k-1}) < 0$ умножив (34) на

$$\Delta_k \text{ и учитывая, что } (p_{k-1}, \Delta_k) \geq 0, \text{ которое справедливо в силу (37), получим:} \\ (p_k, \Delta_k) = (g_k, \Delta_k) - \frac{(g_k, p_{k-1})}{\|p_{k-1}\|^2} (p_{k-1}, \Delta_k) \geq (g_k, \Delta_k). \quad (40)$$

В результате мы доказали (36). Справедливость (37) следует из (39).

Относительно сходимости нового алгоритма обучения справедливо следующее.

Теорема 3. Пусть последовательность $\{s_k\}$ – результат процесса (33), (34) решения задачи 1, $\|g_k\| \neq 0$ для $k = 0, 1, 2, \dots$, $s_o \in R^n$. Тогда, если выполнено предположение 1, имеет место оценка:

$$(\Delta_k, \Delta_k) \leq (\Delta_0, \Delta_0)[1 - q^2(1 - \lambda^2)]^k. \quad (41)$$

При этом на отдельной итерации оценка скорости сходимости процесса (33), (34) не хуже, чем у алгоритма Качмажа, при условии шага каждого из алгоритмов из одной и той же точки s_k .

Доказательство. Из (38) найдем:

$$\begin{aligned} (\Delta_{k+1}, \Delta_{k+1}) &= (\Delta_k, \Delta_k) - 2 \frac{(1 + \lambda_k)(p_k, \Delta_k)(g_k, \Delta_k)}{(p_k, g_k)} + \frac{(1 + \lambda_k)^2 (g_k, \Delta_k)^2 (p_k, p_k)}{(p_k, g_k)^2} = \\ &= (\Delta_k, \Delta_k) - \frac{(1 + \lambda_k)(g_k, \Delta_k)}{(p_k, g_k)} \left[2(p_k, \Delta_k) - \frac{(1 + \lambda_k)(g_k, \Delta_k)(p_k, p_k)}{(p_k, g_k)} \right]. \end{aligned} \quad (42)$$

В силу ортогональности $(p_k, g_{k-1}) = 0$, что следует из (37), выполняется равенство $(p_k, p_k) = (p_k, g_k)$. Поэтому (42) приводится к виду

$$\min_j \quad (43)$$

При условии, что $(p_k, g_k) > 0$ из (43), с учетом (7) и (36), получим:

$$(\Delta_{k+1}, \Delta_{k+1}) \leq (\Delta_k, \Delta_k) - \frac{(1 - \lambda_k^2)(g_k, \Delta_k)^2}{(p_k, g_k)}. \quad (44)$$

Для случая, $(g_k, p_{k-1}) < 0$, умножив (34) на $a_{i_i}^{k\bar{l}}$, получим:

$$(p_k, g_k) = (g_k, g_k) - \frac{(g_k, p_{k-1})^2}{\|p_{k-1}\|^2} \leq (g_k, g_k). \quad (45)$$

Рассмотрим положение вектора g_k и $\hat{\Delta}_k$ – проекции вектора Δ_k в плоскости ортогональных векторов p_{k-1} , p_k . Согласно (34), вектор g_k имеет отрицательную компоненту вдоль p_{k-1} . Согласно неравенству (37) для индекса k равного $k-1$, вектор $\hat{\Delta}_k$ имеет положительную компоненту вдоль p_{k-1} . Поэтому угол между g_k , $\hat{\Delta}_k$ больше угла между векторами g_k , p_k . Следовательно, для ко-

синусов этих углов можно записать цепочку неравенств:

$$\begin{aligned} \frac{(p_k, g_k)}{\|p_k\| \|g_k\|} &\geq \frac{(\hat{\Delta}_k, g_k)}{\|\hat{\Delta}_k\| \|g_k\|} \geq \\ &\geq \frac{(\Delta_k, g_k)}{\|\Delta_k\| \|g_k\|} \geq q_k. \end{aligned} \quad (46)$$

Отсюда и равенства:

$$\begin{aligned} ||p_k|| &= \sqrt{(p_k, p_k)} = \sqrt{(g_k, p_k)} \text{ следует} \\ (p_k, g_k) &\geq q_k ||p_k|| ||g_k|| = \\ &= q_k ||g_k|| \sqrt{(g_k, p_k)}. \end{aligned} \quad (47)$$

Из (44), вместе с (45) имеем:

$$\begin{aligned} (g_k, g_k) &\geq (p_k, g_k) \geq \\ &\geq q_k^2 (g_k, g_k) > 0. \end{aligned} \quad (48)$$

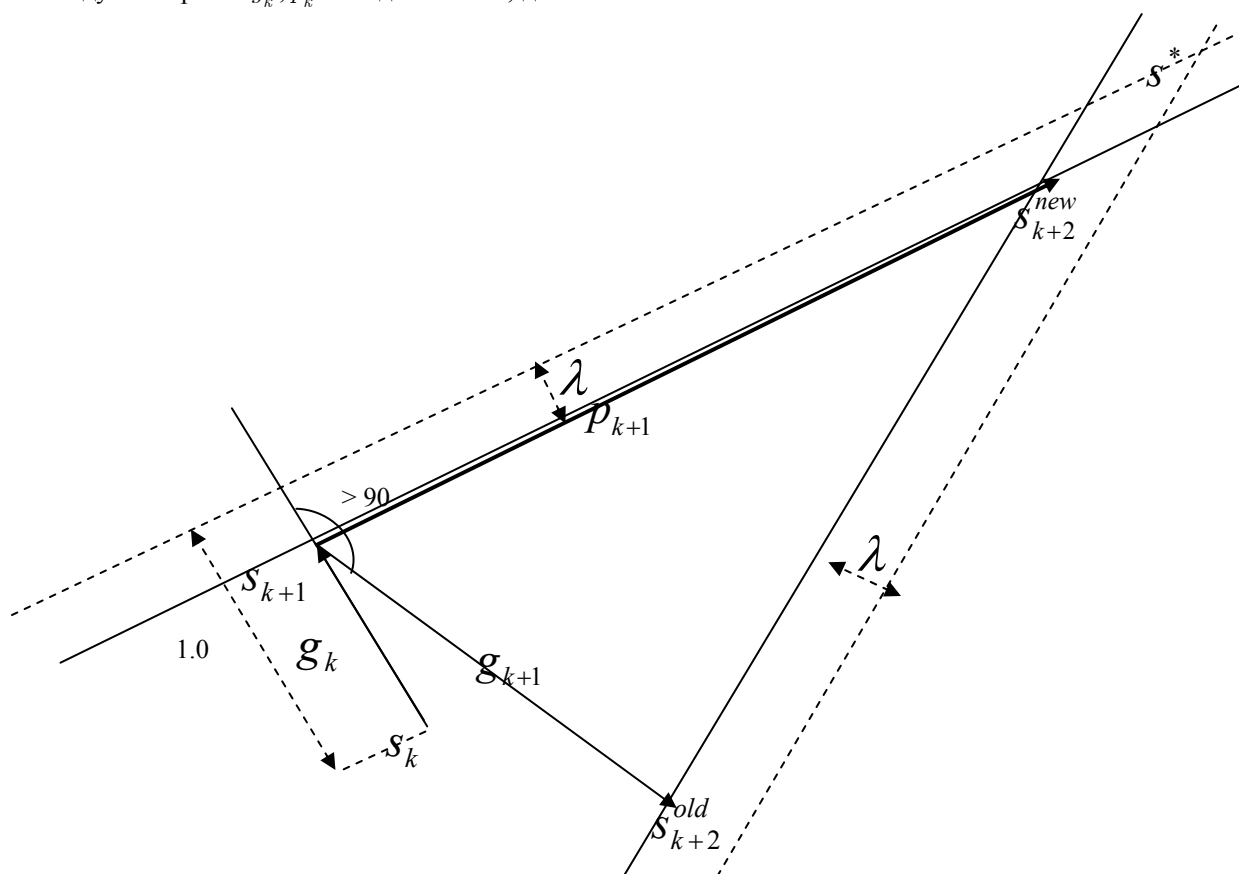


Рис. 1. Поведение многошагового алгоритма обучения на k -ой итерации, $t = 1, \dots, T$ – приближение по алгоритму Качмажа, $x^0, x^*(1), \dots, x^*(T)$ – приближение в случае многошагового алгоритма, S^* – искомый вектор спуска

Таким образом, в силу (48), оценка (44) получена корректно, а на основании (45) приводится к виду:

$$(\Delta_{k+1}, \Delta_{k+1}) \leq (\Delta_k, \Delta_k) - \frac{(1 - \lambda_k^2)(g_k, \Delta_k)^2}{(g_k, g_k)}. \quad (49)$$

При условиях (5) и (6) из (49) следует оценка:

$$x^*(t-1), \dots, x^*(1). \quad (50)$$

Для обоснования (41) используем рекуррентно оценку (50).

Сравнивая оценки (49) и (13), приходим к выводу, что на отдельной итерации, при условии, что оба алгоритма ((2) и (33), (34)) делают шаг из одной точки, в силу неравенства в (49) и равенства в (13), следует последнее утверждение теоремы о том, что на отдельной итерации оценка скорости сходимости процесса (33), (34) не хуже, чем у алгоритма Качмажа.

Заключение

В работе предложены и исследованы теоретически многошаговые алгоритмы обучения при условиях, аналогичных условиям работы алгоритмов обучения в субградиентных методах минимизации. Ускоряющие свойства предлагаемых методов определяются использованием идеи ортогонализации последовательных направлений процесса обучения. Теоретически показано, что свойства сходимости подобных алгоритмов не хуже, чем у алгоритма Качмажа. Предложенные алгоритмы в первую очередь предназначены для ускорения сходимости

обучающихся субградиентных методов минимизации негладких функций [6].

Литература

1. Kaczmarz, S. Approximate solution of systems of linear equations / S. Kaczmarz // *Internat. J. Control.* – 1993. – V. 54. – № 3.
2. Цыпкин, Я. З. Основы теории обучающихся систем / Я. З. Цыпкин. – М.: Наука, 1981. – 251 с.
3. Аведьян, Э. Д. Обобщенный алгоритм Качмажа / Э. Д. Аведьян, Я. З. Цыпкин // *Автоматика и телемеханика.* – 1949. – № 1.
4. Крутиков, В. Н. Релаксационный метод минимизации с растяжением пространства в направлении субградиента / В. Н. Крутиков, Т. В. Петрова // *Экономика и мат. методы.* – 2003. – Т. 39. – Вып. 1.
5. Крутиков, В. Н. Семейство релаксационных субградиентных методов с двухранговой коррекцией матриц метрики / В. Н. Крутиков, Т. А. Горская // *Экономика и мат. методы.* – 2009. – Т. 45. – № 4.
6. Крутиков, В. Н. Обучающие методы безусловной оптимизации и их применение / В. Н. Крутиков. – Томск: Изд-во ТГПУ, 2008. – 234 с.