ЭКОНОМИКА И МАТЕМАТИЧЕСКИЕ МЕТОДЫ, 2009, том 45, № 4, с. 105–120

МЕТОДЫ ОПТИМИЗАЦИИ

СЕМЕЙСТВО РЕЛАКСАЦИОННЫХ СУБГРАДИЕНТНЫХ МЕТОДОВ С ДВУХРАНГОВОЙ КОРРЕКЦИЕЙ МАТРИЦ МЕТРИКИ

© 2009 г. В.Н. Крутиков, Т.А. Горская

(Кемерово)

Предложено семейство релаксационных субградиентных методов (РСМ) с двухранговой коррекцией матриц метрики для решения задач негладкой безусловной оптимизации. Доказана сходимость его алгоритмов на строго выпуклых функциях. Приводятся результаты численного исследования одного алгоритма семейства.

1. ВВЕДЕНИЕ

Излагаемое в работе семейство методов минимизации с двухранговои коррекцией матриц метрики (обозначим его $R_{\alpha\beta}$) принадлежит классу релаксационных методов ε-субградиентного типа (РСМ) (Демьянов, Васильев, 1981) с растяжением пространства (R_{PCM}) (Крутиков, Петрова, 2003; Крутиков, 2003), к числу которых, как доказано в (Крутиков, 2003), относится и r-алгоритм (Шор, 1979). Актуальность исследования и совершенствования подобного рода алгоритмов обусловлена тем, что при относительно невысоких вычислительных затратах на итерации они имеют высокую скорость сходимости и работоспособны в условиях отсутствия выпуклости и гладкости, в силу чего дополняют по области применимости и вычислительной эффективности известные методы минимизации, основанные на квадратичной либо кусочно-линейной моделях функции, в частности квазиньютоновские (Денис, Шнабель, 1988), и алгоритмы, определяемые идеологией метода уровней (Lemarechal, Nemirovskii, Nesterov, 1995; Гольштейн, Немировский, Нестеров, 1995; Бэр, Гольштейн, Соколов, 2000; Гольштейн 2006).

Пусть решается задача минимизации выпуклой на R^n функции $f(x)$. В РСМ последовательные приближения строятся по формулам

$$x_{i+1} = x_i - \gamma_i s_i, \quad \gamma_i = \arg\min_{\gamma} f(x_i - \gamma s_i),$$

где направление спуска s_i выбирается как решение неравенств (Демьянов, Васильев, 1981):

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

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

Идея построения быстросходящихся алгоритмов обучения для методов класса R_{PCM} состоит в распространении итерационного метода наименьших квадратов (ИМНК) на решение систем неравенств типа (1) на отделимых множествах, близких к плоским множествам (содержащихся в некоторой гиперплоскости) (Крутиков, Петрова, 2003). Подобным образом получены алгоритм с растяжением пространства в направлении субградиента (МРП) (Крутиков, Петрова, 2003) и одноранговое семейство методов с растяжением пространства R_α (Крутиков, 2003). В (Крутиков, Арышев, 2005) предложен алгоритм решения неравенств с растяжением/сжатием пространства на плоских отделимых множествах, не выводимый посредством техники ИМНК.

В настоящей работе дано обоснование сходимости этого алгоритма, сформулировано семейство алгоритмов решения неравенств с двухранговой коррекцией матриц метрики, включающее в себя алгоритмы обучения семейства R_α, обоснованы ограничения на характеристики отделимых множеств и параметры алгоритмов обучения, обеспечивающие их сходимость. На базе предложенного семейства алгоритмов обучения разработано двухранговое семейство методов минимизации $R_{\alpha\beta}$ и на строго выпуклых функциях (при соответствующих ограничениях на субградиентные множества) обоснована сходимость его алгоритмов. Приводятся результаты численного тестирования одного варианта алгоритма, сходного по форме с квазиньютоновскими методами.

2. СЕМЕЙСТВО МЕТОДОВ РЕШЕНИЯ НЕРАВЕНСТВ

Обозначим через $\eta_G \equiv \eta(G)$ ближайший к началу координат вектор множества *G:*

$$\rho_G \equiv \rho(G) = \| \eta(G) \|, \quad \mu_G = \eta(G)/\| \eta(G) \|, \quad s^* = \mu_G/\rho_G, \quad R_G \equiv R(G) = \max_{g \in G} \| g \|,$$

$$R_S \equiv R_S(G) = \max_{g \in G}(\mu_G, g), \quad r_G \equiv r(G) = \rho_G/R_S, \quad v(G) = \rho_G/R_G.$$

Будем полагать, что относительно множества *G* выполняется следующее предположение.

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 (1). Параметры ρ_G и R_S характеризуют толщину множества *G* в направлении μ_G, выражаемую в виде двухстороннего неравенства

$$\rho_G \le (\mu_G, g) \le R_S \quad \forall g \in G. \tag{2}$$

Из (2) с учетом вида вектора s^* получим

$$1 \le (s^*, g) \le R_S/\rho_G \quad \forall g \in G. \tag{3}$$

Граничное значение R_S/ρ_G в (3) существенно влияет на скорость сходимости рассматриваемых в работе методов решения неравенств. Величина R_S согласно ее определению и (2) удовлетворяет ограничениям

$$\rho_G \le R_S \le \| \mu_G \| \max_{g \in G} \| g \| \le R_G. \tag{4}$$

Для произвольной симметричной строго положительно определенной матрицы *H* размера $n \times n$ будем использовать обозначение $H > 0$. В предлагаемом алгоритме строятся последовательные приближения решения системы (1) в виде $s_i = H_g g_i$, где g_i – произвольный вектор из *G,* а матрицы $H_i > O$, $i = 0, 1, ...$, корректируются так, что при определенных ограничениях на характеристики множества *G* через конечное число итераций будет получено решение системы (1).

Алгоритм решения неравенств А(α, β).

1. Пусть $i = 0$. Зададим параметры α, β, удовлетворяющие условиям

$$\alpha > 1, \quad 0 < \beta \le 1, \quad \alpha\beta > 1, \tag{5}$$

начальную матрицу $H_0 = I$ и выберем $g_0 \in G$.

2. Найдем вектор $u_i \in G$, такой, что

$$(H_i u_i, g_i) \le 0. \tag{6}$$

Если такого вектора не существует, то $H_i g_i \in S(G)$ и алгоритм заканчивает свою работу.

3. Вычисляем вектор $y_i = g_i - u_i$ и ортогональный ему вектор $v_i = H_i p_i$, где

$$p_i = u_i + t_i y_i = (1 - t_i)u_i + t_i g_i, \quad t_i = -(H_i y_i, u_i)/(H_i y_i, y_i). \tag{7}$$

Затем получаем новое приближение матрицы метрики:

$$H_{i+1} = H_i - (1 - 1/\alpha^2)H_i y_i y_i^T H_i^T/(y_i, H_i y_i) - (1 - 1/\beta^2)H_i p_i p_i^T H_i^T/(p_i, H_i p_i). \tag{8}$$

4. Выбираем произвольный вектор $g_{i+1} \in G$.

Увеличиваем *i* на 1 и переходим на п. 2 алгоритма.

В работе (Крутиков, 2003) было доказано присутствие в *r*-алгоритме метода решения неравенств, дано его описание и предложен более общий алгоритм, отличающийся от алгоритма А(α, β) лишь формулой преобразования матриц (8). Этот метод положен в качестве основы одноранговогo семейства методов с растяжением пространства R_α. Для уменьшения числа извлечений векторов из множества (в алгоритме минимизации извлечение означает вычисление субградиента) рассмотрим предложенные в (Крутиков, 2003) способы вычисления вектора g_{i+1} (п. 4 алгоритма A(α, β)) с учетом того, чтобы новое приближение $s_{i+1} = H_{i+1}g_{i+1}$ удовлетворяло неравенству $(H_{i+1}g_{i+1}, u_i) > 0$ из (1) для последнего извлеченного вектора $u_i \in G$.

1. Выбираем вектор g_{i+1}, как в методе решения неравенств r-алгоритма, полученном в (Крутиков, 2003):

$$g_{i+1} = u_i. \tag{9}$$

2. Из оболочки векторов u_i, g_i в текущей метрике выбираем вектор минимальной длины:

$$g_{i+1} = p_i. \tag{10}$$

3. Получаем семейство алгоритмов решения неравенств, комбинируя (9) и (10):

$$g_{i+1} = \Delta p_i + (1-\Delta) u_i, \quad 0 \le \Delta \le 1, \tag{11}$$

где Δ – задаваемый параметр.

Обозначим через A(α, β, Δ) алгоритм Δ(α, β) с реализацией его п. 4 по формуле (11). Частные случаи алгоритма с использованием (9) или (10) вместо п. 4 будем обозначать Δ(α, β, 0) и Δ(α, β, 1), соответственно. Алгоритм A(α, β, 1) при фиксированных $H_i = I$, $i = 0, 1, ...$, реализует алгоритм решения неравенств методом сопряженных субградиентов (Wolfe, 1974). Алгоритм Δ(α, β, 0) при β = 1 является методом решения неравенств r-алгоритма. Алгоритм A(α, β, Δ) при β = 1 является алгоритмом обучения одноранговогo семейства R_α.

Пусть $A_i = H_i^{-1}$. Обозначим через Sp 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 $A^{1/2}A^{1/2} = A$.

Лемма 1. *Пусть* $H_i > 0$, *матрица* H_{i+1} *получена в результате* (8), *где параметры* α, β *удовлетворяют условию* (5), *а для произвольных векторов* $y_i \ne 0$, $p_i \ne 0$ *выполняется равенство*

$$g_{i+1} = p_i. \tag{12}$$

Тогда $H_{i+1} > 0$ *и*

$$A_{i+1} = A_i + (\alpha^2 - 1)\frac{y_i y_i^T}{(y_i, H_i y_i)} + (\beta^2 - 1)\frac{p_i p_i^T}{(p_i, H_i p_i)}, \tag{13}$$

$$\mathrm{Sp} A_{i+1} = \mathrm{Sp} A_i + (\alpha^2 - 1)\frac{(y_i, y_i)}{(y_i, H_i y_i)} + (\beta^2 - 1)\frac{(p_i, p_i)}{(p_i, H_i p_i)}, \tag{14}$$

$$\det H_{i+1} = \det H_i/\alpha^2\beta^2, \quad \det A_{i+1} = \alpha^2\beta^2 \det A_i. \tag{15}$$

Д о к а з а т е л ь с т в о. Равенство (12) позволяет представить (8) в виде двух последовательных преобразований:

$$H_{i+} = H_i - \left(1 - \frac{1}{\alpha^2}\right)\frac{H_i y_i y_i^T H_i^T}{(y_i, H_i y_i)}, \tag{16}$$

$$H_{i+1} = H_{i+} - \left(1 - \frac{1}{\beta^2}\right)\frac{H_{i+} p_i p_i^T H_{i+}^T}{(p_i, H_{i+} p_i)}. \tag{17}$$

Действительно, в силу (12) из (16) следует, что $H_{i+}p_i = H_i p_i$. Используя полученное равенство и подставляя формулу (16) в (17), получим (8).

Дважды применяя к (16) и (17) формулу Шермана–Мориссона, с последующим учетом равенства $H_{i+}p_i = H_i p_i$ получим (13). Из (13) следует (14). Для преобразования (16) известна формула преобразования определителя (Шор, 1979): $\det H_{i+} = \det H_i/\alpha^2$. Применяя ее дважды для преобразований (16) и (17), получим (15). Из (15) с учетом $H_i > 0$ и (5) следует, что $H_{i+1} > 0$.

В следующей лемме изучаются свойства последовательностей, генерируемых введенными алгоритмами.

Лемма 2. *Пусть множество G удовлетворяет предположению* 1. *Тогда для генерируемых последовательностей при* $i = 0, 1,...$ *одним из алгоритмов* $A(\alpha, \beta)$ *или* $A(\alpha, \beta, \Delta)$ *выполняются соотношения:*

1) *значение* t_i *в* (7) *доставляет минимум величинам* (p_i, H_ip_i) *и* $(p_i, H_{i+1}p_i)$;

2) $\| y_i \| > 0, t_i \in (0,1), p_i \in G, g_{i+1} \in G, H_{i+1} > 0, (H_{i+1}p_i, u_i) > 0$;

а для последовательностей алгоритма $A(\alpha, \beta, \Delta)$ *дополнительно выполняется неравенство* $(H_{i+1}g_{i+1}, u_i) > 0$.

Доказательство. Проведем доказательство леммы методом индукции. Пусть на некоторой итерации i: $H_i > 0$, $g_i \in C$. При $i = 0$ это допущение выполнено для обоих алгоритмов. Покажем выполнимость соотношений леммы для заданного i.

Так как $(H_iy_i, y_i) = (H_iu_i, u_i) + (H_ig_i, g_i) - 2(H_iu_i, g_i)$, $H_i > 0$, u_i, $g_i \in G$, и выполняется неравенство (6), то $(H_iy_i, y_i) > 0$, что определяет корректность формул (7) и вместе с $H_i > 0$ доказывает неравенство $\| y_i \| > 0$.

Если $H_i > 0$, то выполняется первое утверждение леммы. Действительно, решая задачу минимизации $(p_i, H_{i+1}p_i)$ по параметру t_i, где p_i определен в (7), найдем

$$t_i = -\frac{(H_{i+1}y_i, u_i)}{(H_{i+1}y_i, y_i)} = -\frac{(H_iy_i, u_i)}{(H_iy_i, y_i)} = \frac{(H_iu_i, u_i) - (H_iu_i, g_i)}{(H_iu_i, u_i) + (H_ig_i, g_i) - 2(H_iu_i, g_i)}. \tag{18}$$

В силу справедливости (12), которое устанавливается непосредственной проверкой, из (8) следует равенство $H_{i+1}y_i = H_iy_i/\alpha^2$, которое обосновывает замену $H_{i+1} \to H_i$ в (18) и доказывает высказанное утверждение.

Из (6), свойства $H_i > 0$ и u_i, $g_i \in G$ следует, что числитель и знаменатель в (18) строго положительны и числитель строго больше знаменателя. Поэтому $t_i \in (0, 1)$. Используя (7) и учитывая выпуклость множества G, u_i, $g_i \in G$ и $t_i \in (0, 1)$, получим $p_i \in G$. Поскольку в (11) $0 \le \Delta \le 1$ и u_i, $p_i \in G$, то для алгоритма $A(\alpha, \beta, \Delta)$ аналогично докажем $g_{i+1} \in G$.

Так как $p_i \in G$, то $p_i \neq 0$. Следовательно, все условия леммы 1 выполнены. Опираясь на ее результаты, имеем $H_{i+1} > 0$.

В силу (11) для обоснования неравенства $(H_{i+1}g_{i+1}, u_i) > 0$ нам достаточно доказать неравенство $(H_{i+1}p_i, u_i) > 0$, которое выполняется при $(H_ip_i, u_i) > 0$, что следует из (8) и свойства (12). Покажем выполнимость последнего неравенства при условии $H_i > 0$. Из разложений $g_i = u_i + y_i$, $u_i = p_i - t_iy_i$, и (12) следуют равенства:

$$(H_ip_i, g_i) = (H_ip_i, u_i) = (H_ip_i, p_i) > 0. \tag{19}$$

В силу $H_i > 0$ и $p_i \in G$ в (19) будет выполнено и требуемое неравенство, что завершает доказательство утверждений леммы для заданного i. Продолжая процесс индукции, получим доказательство леммы.

Поскольку в п. 4 алгоритма $A(\alpha, \beta, \Delta)$ генерируются векторы $g_{i+1} \in G$, то он является частным случаем алгоритма $A(\alpha, \beta)$. Изложим идею обоснования последнего. На основании (3) и неравенства Шварца для произвольных $g_i \in G$ и $H_i > 0$ получим:

$$1 \le (s^*, g_i)^2 \le (s, A_i^{1/2}H_i^{1/2}g_i)^2 \le (s^*, A_is^*)(g_i, H_ig_i). \tag{20}$$

Покажем, что существует область параметров α, β, при которых правая часть (20) для последовательностей $\{g_i\}$ и $\{H_i\}$, генерируемых алгоритмом, при условии возможности выполнения п. 2 убывает и через конечное число итераций станет меньше единицы. В силу (20) этого не может быть. Следовательно, через конечное число итераций в п. 2 невозможно будет найти вектор $u_i \in G$, удовлетворяющий (6), т.е. будет найдено решение системы (1).

Изучим поведение последовательности $\{\tau_i = \min_{0 \le j \le i-1} [(y_j, H_jy_j)/(y_j, y_j)]\}$.

Теорема 1. *Пусть множество G удовлетворяет предположению 1, последовательность* $\{H_i\}$ – *результат преобразования (8), где* y_i, $p_i \in R^n$, $y_i \neq 0$, $p_i \neq 0$, $i = 0, 1, ...,$ – *произвольные векторы, связанные соотношением (12), параметры* α, β *удовлетворяют (5) и* $H_0 = I$. *Тогда*

$$\tau_i \leq \frac{i(\alpha^2-1)}{n[(\alpha^2\beta^2)^{i/n}-1]}, \quad i \geq 1. \tag{21}$$

Д о к а з а т е л ь с т в о. Из (15) и начального условия $H_0 = I$ следует, что

$$\det A_i = \det A_0(\alpha^2\beta^2)^i = (\alpha^2\beta^2)^i. \tag{22}$$

Согласно неравенству $\mathrm{Sp}\, A_i/n \geq (\det A_i)^{1/n}$ для среднеарифметического и среднегеометрического собственных значений матрицы $A_i > 0$ с учетом (14) и (22) получим:

$$1+\frac{\alpha^2-1}{n}\sum_{l=0}^{i-1}\frac{(y_l, y_l)}{(y_l, H_l y_l)}+\frac{\beta^2-1}{n}\sum_{l=0}^{i-1}\frac{(p_l, p_l)}{(p_l, H_l p_l)} \geq (\alpha^2\beta^2)^{i/n}.$$

Преобразуем его с учетом неравенства (5) для β:

$$\sum_{l=0}^{i-1}\frac{(y_l, y_l)}{(y_l, H_l y_l)} \geq \frac{n}{\alpha^2-1}\left[(\alpha^2\beta^2)^{i/n}-1-\frac{\beta^2-1}{n}\sum_{l=0}^{i-1}\frac{(p_l, p_l)}{(p_l, H_l p_l)}\right] \leq \frac{n[(\alpha\beta)^{2i/n}-1]}{\alpha^2-1}.$$

Отсюда, поскольку хотя бы одно слагаемое суммы при некотором j будет не меньше среднего, то

$$\frac{(y_l, y_l)}{(y_l, H_l y_l)} \geq \frac{1}{i}\sum_{l=0}^{i-1}\frac{(y_l, y_l)}{(y_l, H_l y_l)} \geq \frac{n[(\alpha\beta)^{2i/n}-1]}{\alpha^2-1}, \quad 0 \leq j \leq i-1,$$

что доказывает (21).

Лемма 3. *Пусть множество G удовлетворяет предположению* 1, *а последовательность* $\{\pi_i = \min\limits_{0 \leq j \leq i-1} (g_j, H_j g_j)\}$ *вычисляется на основе характеристик алгоритма A(*α, β*). Тогда*

$$\pi_i \leq \frac{4iR_G^2(\alpha^2-1)}{n[(\alpha^2\beta^2)^{i/n}-1]}, \quad i \geq 1. \tag{23}$$

Доказательство. Исходя из (6) можно записать:

$$(y_j, H_j y_j) = (g_j, H_j g_j)+(u_j, H_j u_j)-2(g_j, H_j u_j) \geq (g_j, H_j g_j)+(u_j, H_j u_j).$$

Отсюда следует

$$\frac{(y_j, H_j y_j)}{(y_j, y_j)} \geq \frac{(g_j, H_j g_j)}{(\|g_j\|+\|u_j\|)^2} \geq \frac{(g_j, H_j g_j)}{4R_G^2},$$

что вместе с (21) доказывает (23).

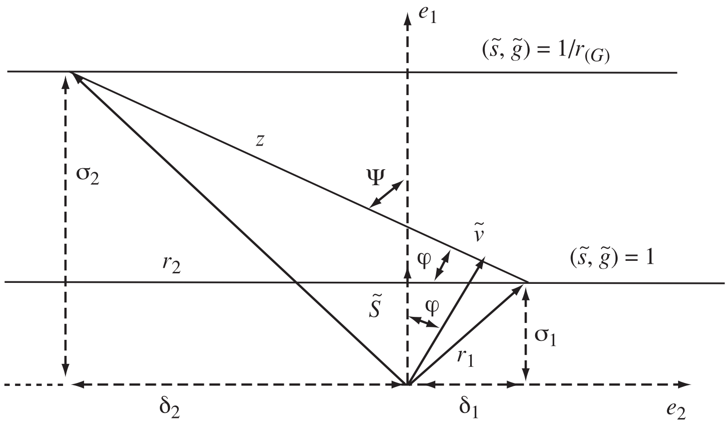
Исследуем поведение сомножителя $(s^*, A_i s^*)$ в (20). Из (13) получим:

$$(s^*, A_{i+1}s^*) = (s^*, A_i s^*)+(\alpha^2-1)\frac{(s^*, y_i)^2}{(y_i, H_i y_i)}+(\beta^2-1)\frac{(s^*, p_i)^2}{(p_i, H_i p_i)}. \tag{24}$$

Для некоторого фиксированного i обозначим: $\tilde{s}^* = A_i^{1/2}s^*$, $\tilde{v} = A_i^{1/2}v_i = H_i^{1/2}p_i$, $r_1 = H_i^{1/2}g_i$, $r_2 = H_i^{1/2}u_i$, $z = H_i^{1/2}y_i$; P – плоскость, образованную векторами r_1 и r_2; $\tilde{s}$ – проекцию вектора $\tilde{s}^*$ на плоскость P; $\tilde{G} = P \cap \{H_i^{1/2}g \mid \forall g \in G\}$; ϕ – угол, образованный векторами $\tilde{s}$, $\tilde{v}$; ψ – угол, образованный векторами $\tilde{s}$, z.

Из (12) следует ортогональность векторов z, $\tilde{v}$. Поэтому $\psi = 90° - \varphi$ и $\cos^2\psi = \sin^2\varphi$. С учетом равенства $(\tilde{s}, \tilde{g}) = (\tilde{s}^*, \tilde{g}) = (s^*, g)$, справедливого для $\forall\tilde{g} \in \tilde{G}$, неравенство (4) представим в виде

$$1 \leq (\tilde{s}, \tilde{g}) \leq R_S/\rho_G = r_G^{-1} \quad \forall\tilde{g} \in \tilde{G}. \tag{25}$$



Плоскость векторов r_1 и r_2

В плоскости P введем систему координат с ортонормированными векторами $e_1 = \tilde{s}/\| \tilde{s} \|$ и e_2. В силу задания плоскости P и вектора e_1 вектор e_2 определяется с точностью до знака, а его выражение не приводится, поскольку оно напрямую не используется. На рисунке изображены введенные характеристики и границы (25). Преобразуем величины из (24):

$$\frac{(s^*, y_i)^2}{(y_i, H_i y_i)} = \frac{(A_i^{1/2} s^*, H_i^{1/2} y_i)^2}{(y_i, H_i y_i)} = \frac{(\tilde{s}^*, z)^2}{(z, z)} = \frac{(\tilde{s}, z)^2}{(z, z)} = \frac{\| \tilde{s} \|^2 \| z \|^2 \cos^2 \psi}{\| z \|^2} =$$

$$= \frac{\| \tilde{s} \|^2 \| z \|^2 \sin^2 \varphi}{\| z \|^2 (\sin^2 \varphi + \cos^2 \varphi)} = \frac{\| \tilde{s} \|^2 \operatorname{tg}^2 \varphi}{1 + \operatorname{tg}^2 \varphi},$$

$$\frac{(s^*, p_i)^2}{(p_i, H_i p_i)} = \frac{(\tilde{s}, \tilde{v})^2}{(\tilde{v}, \tilde{v})} = \frac{\| \tilde{s} \|^2 \operatorname{ctg}^2 \varphi}{1 + \operatorname{ctg}^2 \varphi} = \frac{\| \tilde{s} \|^2}{1 + \operatorname{tg}^2 \varphi}.$$

Из (24) с учетом последних равенств следует:

$$(s^*, A_{i+1} s^*) = (s^*, A_i s^*) + \| \tilde{s} \|^2 \left[(\alpha^2 - 1) \frac{\operatorname{tg}^2 \varphi}{1 + \operatorname{tg}^2 \varphi} + (\beta^2 - 1) \frac{1}{1 + \operatorname{tg}^2 \varphi} \right]. \tag{26}$$

Для оценки сверху правой части (26) нам потребуется оценить границы изменений величины $\operatorname{tg}^2\varphi$.

Лемма 4. *Для величины* $\operatorname{tg}^2\varphi$ *справедлива оценка*

$$\operatorname{tg}^2 \varphi \le (1 - r(G))^2 / 4r(G). \tag{27}$$

Доказательство. Обозначим (σ_1, δ_1) и (σ_2, δ_2) – координатное представление векторов r_1 и r_2 соответственно во введенной ранее системе координат. На рисунке принято максимизирующее оценку $\operatorname{tg}^2\varphi$ расположение векторов r_1 и r_2 на границах (25). Действительно, удлиняя и сокращая векторы r_1 и r_2, можно их вывести на противоположные границы с одновременным увеличением острого угла φ, что только увеличит оценку $\operatorname{tg}^2\varphi$. Так как

$$(H_i u_i, g_i) = (r_1, r_2) = \sigma_1 \sigma_2 + \delta_1 \delta_2 \le 0$$

и компоненты σ_1, σ_2 положительны, то компоненты δ_1 и δ_2 должны иметь разные знаки. Из последнего неравенства и соотношения между среднеарифметическим и среднегеометрическим следует

$$(|\delta_1| + |\delta_2|)^2 / 4 \ge |\delta_1 \delta_2| \ge \sigma_1 \sigma_2.$$

Отсюда с учетом различия знаков компонент δ_1, δ_2 и равенства $\sigma_1/\sigma_2 = r_G$, которое следует из расположения векторов r_1 и r_2 на границах (25), получим оценку (27):

$$\mathrm{tg}^2\varphi = \frac{(\sigma_2 - \sigma_1)^2}{(|\delta_1| + |\delta_2|)^2} \le \frac{(\sigma_2 - \sigma_1)^2}{4\sigma_1\sigma_2} = \frac{(1 - r(G))^2}{4r(G)}.$$

Обозначим

$$\theta(q) = \left(\frac{1-q}{1+q}\right)^2, \quad \theta_G = \theta(r_G) = \frac{(R_S - \rho_G)^2}{(R_S + \rho_G)^2}, \quad Q(\theta, \alpha, \beta) = (\alpha^2 - 1)\theta + (\beta^2 - 1)(1 - \theta).$$

Параметр θ_G характеризует степень “плоскости” множества G в направлении μ_G.

Лемма 5. *Пусть для множества G выполнено предположение* 1. *Тогда для последовательности* $\{(s^*, A_i s^*)\}$ *имеет место одна из оценок:*

$$(s^*, A_{i+1}s^*) \le (s^*, A_i s^*), \quad \textit{если} \quad Q(\theta_G, \alpha, \beta) \le 0; \tag{28}$$

$$(s^*, A_{i+1}s^*) \le (s^*, A_i s^*)[(1 + Q(\theta_G, \alpha, \beta)], \quad \textit{если} \quad Q(\theta_G, \alpha, \beta) > 0. \tag{29}$$

Д о к а з а т е л ь с т в о. Оценим величину (26) с учетом (27). В силу (27) имеем

$$\frac{\mathrm{tg}^2\varphi}{1 + \mathrm{tg}^2\varphi} \le \frac{(1 - r(G))^2}{4r(G)[1 - (1 - r(G))^2/4r(G)]} = \frac{(1 - r(G))^2}{(1 + r(G))^2} = \theta_G.$$

Отсюда

$$\frac{1}{1 + \mathrm{tg}^2\varphi} = 1 - \frac{\mathrm{tg}^2\varphi}{1 + \mathrm{tg}^2\varphi} \ge 1 - \theta_G.$$

Основываясь на этих неравенствах, получим

$$\frac{(\alpha^2 - 1)\mathrm{tg}^2\varphi}{1 + \mathrm{tg}^2\varphi} + \frac{(\beta^2 - 1)}{1 + \mathrm{tg}^2\varphi} \le (\alpha^2 - 1)\theta_G + (\beta^2 - 1)(1 - \theta_G) = Q(\theta_G, \alpha, \beta).$$

Отсюда и из (26) следует (28), а с учетом неравенства $(s^*, A_i s^*) \le \| \tilde{s} \|^2$ – оценка (29).

Обозначим через $q(\theta, \alpha, \beta)$ – величину, задаваемую одной из зависимостей:

$$q(\theta, \alpha, \beta) = \frac{1 + Q(\theta, \alpha, \beta)}{(\alpha\beta)^{2/n}} \quad \text{при} \quad Q(\theta, \alpha, \beta)] > 0, \tag{30}$$

$$q(\theta, \alpha, \beta) = \frac{1}{(\alpha\beta)^{2/n}} \quad \text{при} \quad Q(\theta, \alpha, \beta)] \le 0. \tag{31}$$

При ограничениях (5) введенный показатель представляет собой неубывающую функцию по θ, т.е.

$$q(\theta_1, \alpha, \beta) \le q(\theta_2, \alpha, \beta), \quad \theta_1 < \theta_2, \tag{32}$$

поскольку при ограничениях (5) $Q(\theta, \alpha, \beta)$ является линейной монотонно возрастающей функцией от θ.

Теорема 2. *Пусть множество G удовлетворяет предположению* 1 *и при некоторых v_0, θ_0 выполняются ограничения*

$$v(G) \ge v_0 > 0, \tag{33}$$

$$\theta_G < \theta_0 < 1. \tag{34}$$

Тогда для сходимости алгоритма $A(\alpha, \beta)$ достаточно существования параметров α, β, удовлетворяющих условиям (5) *и соотношению*

$$Q(\theta_0, \alpha, \beta) < 1. \tag{35}$$

При выполнении условия (35) *алгоритм* $A(\alpha, \beta)$ *сходится за конечное число итераций, которое не превосходит* i_0 *– минимального целого числа из области значений* i*, удовлетворяющих неравенству*

$$\frac{4i(\alpha^2-1)}{n(v_0)^2(1-(\alpha\beta)^{-2i/n})}[q(\theta_0, \alpha, \beta)]^i < 1. \tag{36}$$

Доказательство. Для итерации с номером i определим индекс $j_i = \arg \min\limits_{0 \le j \le i-1} [(g_j, H_j g_j)]$. Из (20) с учетом оценок (23) и (28) или (29) получим:

$$\begin{aligned} 1 \le (s^*, A_{j_i} s^*)(g_{j_i}, H_{j_i} g_{j_i}) &\le \frac{4iR_G^2(\alpha^2-1)}{n[1-(\alpha\beta)^{-2i/n}]}\frac{(s^*, A_{j_i}s^*)}{(\alpha\beta)^{2i/n}} \le \frac{4iR_G^2(\alpha^2-1)[q(\theta_G, \alpha, \beta)]^{j_i}}{n\rho_G^2[1-(\alpha\beta)^{-2i/n}](\alpha\beta)^{2(i-j_i)/n}} \le \\ &\le \frac{4i(\alpha^2-1)[q(\theta_G, \alpha\beta)]^i}{nv_G^2[1-(\alpha\beta)^{-2i/n}]} \le \frac{4i(\alpha^2-1)}{nv_0^2[1-(\alpha\beta)^{-2i/n}]}[q(\theta_0, \alpha, \beta)]^i . \end{aligned} \tag{37}$$

Здесь первое неравенство соответствует (20), второе – получено на основе (23). Третий переход сделан с учетом применяемых рекуррентно оценок (28) или (29), равенств $H_0 = A_0 = I$, $\| s^* \| = 1/\rho_G$ и определения (30), (31). Четвертое неравенство опирается на соотношение $q(\theta, \alpha, \beta) \ge 1/(\alpha\beta)^{2/n}$, которое следует из определения (30), (31). Последний переход сделан на основе неравенств (33), (34) и свойства (32).

Согласно (35) правая часть (37) убывает с ростом i и через конечное число итераций станет меньше единицы при условии, что по-прежнему в п. 2 алгоритма удается найти вектор $u_i \in G$, удовлетворяющий (6).

В силу (37) это невозможно. Следовательно, через конечное число итераций в п. 2 не останется векторов $u_i \in G$, удовлетворяющих (6), т.е. будет найдено решение системы (1). Гарантированную оценку числа итераций, необходимых для получения решения неравенств, дает минимальное целое i, при котором перестает выполняться неравенство (37) или выполняется неравенство (36).

Для обоснования сходимости метода минимизации необходимо указать границы возмущения множества, при которых сохраняется сходимость алгоритма. Обозначим $S_\varepsilon(G) = \{z \in R^n \,|\, \| z - x \| \le \varepsilon,\ \forall x \in G\}$ – окрестность множества G. Введем зависимость $\omega(\theta) = (1-\theta^{1/2})/(1+\theta^{1/2})$, которая является обратной для функции $\theta(\omega)$, т.е. $\omega(\theta(r)) = r$.

Лемма 6. *Пусть множество* G *удовлетворяет предположению* 1, *при некоторых* v_0, θ_0 *выполнены ограничения* (33), (34) *и существуют параметры* α, β *алгоритма* $A(\alpha, \beta)$, *удовлетворяющие неравенствам* (5) *и* (35). *Тогда при* $\varepsilon = \rho_G\Delta/2$, *где* $\Delta = r_G - r^*$, $r^* = \omega(\theta^*)$, *а величина* θ^* *удовлетворяет ограничениям*

$$\theta_G \le \theta^* \le \theta_0, \tag{38}$$

алгоритм $A(\alpha, \beta)$ *сходится на множестве* $S_\varepsilon(G)$ *за конечное число итераций, которое не превосходит* i_0 *– минимального целого числа из области значений* i, *удовлетворяющих неравенству*

$$\frac{36i(\alpha^2-1)}{n(v_0)^2(1-(\alpha\beta)^{-2i/n})}[q(\theta_0, \alpha, \beta)]^i < 1. \tag{39}$$

Доказательство. В работе (Крутиков, Петрова, 2003, лемма 3) показано, что множество $S_\varepsilon(G)$ удовлетворяет предположению 1 и $r(S_\varepsilon(G)) \ge r((G) - 2\varepsilon/\rho_G$. Подставляя в правую часть значение ε, получим $r(S_\varepsilon(G)) \ge r((G) - \Delta = r^*$. Так как функция $\theta(r)$ монотонно убывает, то отсюда и из (38) следует, что

$$\theta(r(G)) \le \theta(r(S_\varepsilon(G))) \le \theta^* \le \theta_0. \tag{40}$$

В (Крутиков, Петрова, 2003, лемма 3) были получены неравенства $R(S_\varepsilon(G)) \le R((G) + \varepsilon$, $\rho(S_\varepsilon(G)) \ge \rho((G) - \varepsilon$. Отсюда, используя заданное значение $\varepsilon = 0.5\rho_G\Delta$ и соотношение $\Delta \le 1$ (оно является следствием неравенств $0 < r_G \le 1$ и $0 < r^* < 1$), а также учитывая выполнение неравенства (33), можем записать

$$v(S_\varepsilon(G)) \geq \frac{\rho_G(1-\Delta/2)}{R_G(1+\Delta/2)} \geq \frac{\rho_G(1-1/2)}{R_G(1+1/2)} \geq v(G)/3 \geq v_0/3 > 0. \tag{41}$$

Согласно (40), (41) и (35) для алгоритма $A(\alpha, \beta)$ выполнены условия сходимости теоремы 2 на множестве $S_\varepsilon(G)$. Подставляя в (36) оценки из (40), (41), придем к соотношению (39) для оценки числа шагов алгоритма.

3. СПОСОБЫ ВЫБОРА ПАРАМЕТРОВ АЛГОРИТМА РЕШЕНИЯ НЕРАВЕНСТВ

Согласно результатам теоремы 2 для обеспечения сходимости алгоритма решения неравенств достаточно указать границу θ_0 характеристики $\theta(G)$ отделимого множества, при которой существуют параметры α, β, обеспечивающие выполнимость (35), и способ выбора этих параметров. Приведем некоторые частные случаи решения поставленной задачи.

В лемме 7 дается обоснование выбора параметров α, β, обеспечивающих минимум показателя в случае (31), т.е. при ограничении $Q(\theta, \alpha, \beta) \leq 0$.

Лемма 7. *При* θ, *удовлетворяющем неравенствам*

$$0 < \theta < 0.5, \tag{42}$$

параметры

$$\alpha = \alpha(\theta) = \sqrt{1/2\theta}, \quad \beta = \beta(\theta) = \sqrt{1/2(1-\theta)} \tag{43}$$

являются решением задачи

$$\min_{\alpha,\, \beta} \{q(\theta, \alpha, \beta) \,|\, Q(\theta, \alpha, \beta) \leq 0, \quad \alpha > 1, 0 < \beta \leq 1, \quad \alpha\beta > 1\}, \tag{44}$$

причем

$$q(\theta, \alpha(\theta), \beta(\theta)) = 1/[\alpha(\theta)\beta(\theta)]^{2/n} = [4\theta(1-\theta)]^{1/n} < 1. \tag{45}$$

Д о к а з а т е л ь с т в о. Задачу (44) в соответствии с определением (31) заменим на эквивалентную:

$$\max_{\alpha,\, \beta} \{(\alpha\beta)^2 \,|\, Q(\theta, \alpha, \beta) \leq 0, \quad \alpha > 1, 0 < \beta \leq 1, \quad \alpha\beta > 1\}. \tag{46}$$

Найдем решение задачи максимизации (46) по параметру β при фиксированном $\alpha > 1$. Поскольку целевая функция в (46) монотонно возрастает, то максимум достигается на границе допустимого множества $Q(\theta, \alpha, \beta) = 0$ при максимальном значении β: $\beta_\alpha(\alpha, \theta) = \sqrt{(1-\alpha^2\theta)/(1-\theta)}$.

Определим из $\max\limits_{\alpha} [\alpha\beta_\alpha(\alpha, \theta)]^2$ оптимальное значение $\alpha(\theta)$, приведенное в (43), и подставим его в ранее найденную зависимость $\beta_\alpha(\alpha, \theta)$. Тем самым нами будет получено оптимальное значение $\beta(\theta) = \beta_\alpha(\alpha(\theta), \theta)$, задаваемое формулой из (43).

Проверим выполнение ограничений задачи (46). При ограничениях (42) справедливы неравенства $\alpha > 1$, $0 < \beta \leq 1$ из (44). Максимальное значение квадратичной функции $\theta(1-\theta)$ достигается при $\theta = 0.5$. Поэтому при $\theta \neq 0.5$ имеет место неравенство $\theta(1-\theta) < 1/4$. Отсюда следует последнее из неравенств (44): $[\alpha(\theta)\beta(\theta)]^2 = 1/[4\theta(1-\theta)] > 1$, что доказывает неравенство (45).

В следующей лемме дается обоснование выбора параметра α одноранговоro алгоритма обучения семейства R_a.

Лемма 8. *При* θ, *удовлетворяющем неравенствам*

$$0 < \theta < 1/n, \tag{47}$$

и фиксированном $\beta = 1$ *значение*

$$\alpha = \alpha_1(\theta) = ((1-\theta)/[(n-1)\theta])^{1/2} \tag{48}$$

является решением задачи

$$\min_{\alpha}\{q(\theta,\alpha,\beta=1)\,\big|\,\alpha>1\},\tag{49}$$

где оптимальное значение показателя удовлетворяет ограничениям:

$$\left(\frac{n^2}{4(n-1)}\right)^{1/n} q(\theta,\alpha(\theta),\beta(\theta)) \le q(\theta,\alpha_1(\theta),1) = \frac{1+\theta(\alpha_2(\theta)^2-1)}{\alpha_1(\theta)^{2/n}} < 1.\tag{50}$$

Доказательство. При $\beta = 1$ и $\alpha^2 > 1$ величина $Q(\theta, \alpha, \beta) > 0$, что определяет форму показателя (31): $q(\theta, \alpha, 1) = [1 + (\alpha^2 - 1)\theta]/\alpha^{2/n}$. Найдем его производную

$$q'_\alpha(\theta,\alpha,1) = \frac{2\theta\alpha^{1+2/n}-[1+(\alpha^2-1)\theta]2\alpha^{2/n-1}/n}{\alpha^{4/n}} = \frac{2}{n\alpha^{1+2/n}}[\alpha^2\theta(n-1)-(1-\theta)].$$

Корень уравнения $q'_\alpha(\theta,\alpha,1) = 0$ задается выражением (48), причем $\alpha_1(\theta) > 1$ при ограничении (47). Если $1 < \alpha < \alpha_1(\theta)$, то $q'_\alpha(\theta,\alpha,1) < 0$, а если $\alpha > \alpha_1(\theta)$, то $q'_\alpha(\theta,\alpha,1) > 0$. Следовательно, в точке $\alpha_1(\theta)$ достигается минимум задачи (49). Поскольку $q(\theta, 1, 1) = 1$, то при $1 < \alpha \le \alpha_1(\theta)$ заведомо будет выполняться неравенство $q(\theta, \alpha, 1) < 0$, что доказывает правое неравенство в (50).

Преобразуем выражение показателя (31):

$$q(\theta,\alpha_1(\theta),1) = \frac{1+\theta(\alpha_1(\theta)^2-1)}{\alpha_1(\theta)^{2/n}} = \frac{1+\theta((1-\theta)/((n-1)\theta)-1)}{[(1-\theta)/((n-1)\theta)]^{1/n}} =$$

$$= \frac{(1-\theta)+(1-\theta)/(n-1)}{[(1-\theta)/((n-1)\theta)]^{1/n}} = \frac{n(n-1)^{1/n}(1-\theta)\theta^{1/n}}{(n-1)(1-\theta)^{1/n}} = n(n-1)^{(1-n)/n}(1-\theta)^{(n-1)/n}\theta^{1/n}.$$

Учитывая неравенства (47) и обозначения (45), получим левое неравенство в (50)

$$q(\theta,\alpha_1(\theta),1) = n(n-1)^{(1-n)/n}(1-\theta)^{(1-n)/n}\theta^{1/n} = \frac{n(n-1)^{(1-n)/n}}{4^{1/n}}((1-\theta)\ )^{(n-2)/n}[4\theta(1-\theta)\ ]^{1/n} \ge$$

$$\ge \frac{n(n-1)^{(1-n)/n}}{4^{1/n}}\left(\frac{n-1}{n}\right)^{(n-2)/n}[4\theta(1-\theta)]^{1/n} = \left(\frac{n^2}{4(n-1)}\right)^{1/n} q(\theta,\alpha(\theta),\beta(\theta)).\quad\blacksquare$$

Согласно результатам теоремы 2 и лемм 7, 8 использование дополнительного параметра β в алгоритме $A(\alpha, \beta)$ позволяет расширить область его сходимости, задаваемой ограничениями на параметр θ множества G, что следует из сравнения (42) с (47), и увеличить его скорость сходимости, о чем свидетельствует левое неравенство в (50).

4. СЕМЕЙСТВО РЕЛАКСАЦИОННЫХ СУБГРАДИЕНТНЫХ МЕТОДОВ

Дадим описание метода минимизации со встроенным алгоритмом решения неравенств $A(\alpha, \beta, \Delta)$.

Алгоритм минимизации $R_{\alpha\beta}$.

1. Зададим начальное приближение матрицы $H_0 = I$; текущее приближение минимума $x_0 \in R^n$; целые $i = 0$, $m_0 = 0$; N – период обновления; параметры α, β, удовлетворяющие условиям (5) и $0 \le \Delta \le 1$. Вычислим $g_0 \in \partial f(x_0)$. Если $g_0 = 0$, то x_0 является искомой точкой минимума и можно закончить вычисления. В противном случае переходим к п. 2.

2. Находим новое приближение для x:

$$x_{i+1} = x_i - \gamma_i s_i,\quad s_i = H_i g_i,\quad \gamma_i = \arg\min_{\gamma} f(x_i - \gamma s_i).$$

3. Вычисляем субградиент $u_i \in \partial f(x_{i+1})$ исходя из условия $(u_i, s_i) \le 0$. Если $u_i = 0$, то u_{i+1} – искомая точка минимума и можно закончить вычисления.

4. Если $i - m_i \ge N$, то необходимо произвести обновление $m_{i+1} = i$, $g_{i+1} = u_i$, $H_{i+1} = I$ и перейти к п. 7.

5. Рассчитываем $y_i = g_i - u_i$ и вектор p_i по формуле (7). Если $p_i = 0$, то обновляем значения $m_{i+1} = i$, $g_{i+1} = u_i$, $H_{i+1} = I$ и переходим к п. 7.

6. Находим H_{i+1} по формуле (8), $g_{i+1} = \Delta p_i + (1 - \Delta)u_i$ и $m_{i+1} = m_i$.

7. Увеличиваем i на единицу и переходим к п. 2.

Отметим, что в п. 3 в силу условия точного одномерного спуска вдоль направления $(-s_i)$ в новой точке x_{i+1} вектор $u_i \in \partial f(x_0)$, удовлетворяющий условию $(u_i, s_i) \le 0$, всегда существует согласно необходимому условию минимума одномерной функции. Ограничение п. 3 $(u_i, s_i) \le 0$ обеспечивает условие (6) алгоритма решения неравенств, встроенного в п. 3, 5, 6 метода минимизации.

Обозначим $D(z) = \{x \in R^n \mid f(x) \le f(z)\}$. Доказательство сходимости метода опирается на следующую лемму.

Лемма 9 (Демьянов, Васильев, 1981). *Пусть функция $f(x)$ строго выпукла на R^n, множество $D(x_0)$ ограничено, а последовательность $\{x_k\}_{k=0}^{\infty}$ такова, что $f(x_{k+1}) = \min\limits_{\alpha \in [0,1]} f(x_k + \alpha(x_{k+1} - x_k))$. Тогда $\lim\limits_{k \to \infty} \| x_{k+1} - x_k \| = 0$.*

Обозначим: i_k – индексы i, при которых происходит обновление в п. 4 или 5, $k = 0, 1, \ldots$; z_k – точки x_{i_k}; x_* – точку минимума функции; x^* – предельные точки последовательности $\{z_k\}_{k=1}^{\infty}$. Существование предельных точек последовательности $\{z_k\}$ при ограниченности множества $D(x_0)$ следует из $z_k \in D(x_0)$.

Теорема 3. *Пусть множество $D(x_0)$ ограничено, функция $f(x)$ строго выпукла на R^n, на $D(x_0)$ при $x \ne x_*$ и некоторых v_0, θ_0 выполнены ограничения:*

$$\theta(r(\partial f(x))) < \theta_0 < 1, \tag{51}$$

$$v(\partial f(x)) \ge v_0 > 0, \tag{52}$$

и пусть существуют параметры α, β алгоритма $R_{\alpha\beta}$, удовлетворяющие неравенствам (5) *и* (35). *Тогда при периоде обновления алгоритма $R_{\alpha\beta}$, $N \ge N_0$, где N_0 – удвоенное минимальное целое из области значений i, удовлетворяющих неравенству* (39), *любая предельная точка последовательности $\{z_k\}$ является точкой минимума на R^n.*

Доказательство. Обозначим $S^* = \partial f(x^*)$, $S_\varepsilon^* = S_\varepsilon(\partial f(x^*))$. Допустим, что утверждение теоремы неверно: предположим, что подпоследовательность $z_{k_s} \to x^*$, но $\rho(S^*) > 0$. Положим:

$$\varepsilon = 0.5\rho(S^*)\Delta, \quad \Delta = r(S^*) - r^*, \quad r^* = \omega(\theta_0). \tag{53}$$

Выберем $\delta > 0$, для которого $\partial f(x) \subset S_\varepsilon^* \quad \forall x \in S_\delta(x^*)$. Такой выбор возможен в силу полунепрерывности сверху точечно-множественного отображения $\partial f(x)$ (Демьянов, Васильев, 198 1 , с. 289).

Для множеств S^* и S_ε^* при выборе ε согласно ограничению (53) с учетом условий (51), (52) будут выполнены условия леммы 6 (сходимость за конечное число итераций алгоритма решения неравенств, который реализован в п. 3, 5, 6 метода минимизации), причем число итераций согласно условию теоремы не превосходит $N_0/2$ шагов. Поэтому число шагов N алгоритма решения неравенств в методе минимизации превосходит в 2 раза величину $N_0/2$ и является достаточным для получения решения на множестве S_ε^*, прежде чем произойдет обновление в п. 4. Вектор p_i, получаемый в п. 5, согласно формулам п. 3 и 5 принадлежит некоторой оболочке субградиентов из множества S_ε^*. Следовательно, пока последовательность $x_i \in S_\delta(x^*)$, вектор p_i не может обратиться в нуль и не произойдет преждевременного обновления в п. 5. Из сказанного следует, что до обновления в п. 4 или 5 будет получено решение системы неравенств на множестве S_ε^*, что означает выход последовательности x_i из окрестности $S_\delta(x^*)$.

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

$$z_{k_s} \in S_{\delta/2}(x^*), \quad x_i \in S_\delta(x^*), \quad i_{k_s} \le i \le i_{k_s} + N_0,$$

т.е. такой номер K, что точки x_i остаются в окрестности $S_\delta(x^*)$ в течение, по крайней мере, N_0 шагов алгоритма. Такой выбор возможен в силу предположения сходимости $z_{k_s} \to x^*$ и выполнения на каждом шаге алгоритма условий леммы 9. Последнее утверждение противоречит полученному ранее выводу, что последовательность x_i выйдет из окрестности $S_\delta(x^*)$ через $N_0/2$ шагов алгоритма. Полученное противоречие доказывает теорему.

5. РЕАЛИЗАЦИЯ ОДНОГО ВАРИАНТА АЛГОРИТМА СЕМЕЙСТВА

Алгоритм $R_{\alpha\beta}$ реализован при фиксированном параметре $\Delta = 0$ с процедурой одномерной минимизации, предложенной в (Крутиков, Петрова, 2003). Обозначим обращение к процедуре через $OM(\{x, s, g_x, f_x, h_0\}; \{\gamma_m, f_m, g_m, \gamma_1, u, h_1\})$. Входными параметрами являются: x – точка текущего приближения минимума; s – направление спуска; h_0 – начальный шаг; $f_x = f(x)$; $g_x \in \partial f(x)$; причем $(g_x, s) > 0$. Выходные параметры: 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, при котором в точке $w = x - \gamma_1 s$ для субградиента $u \in \partial f(w)$ справедливо неравенство $(u, s) \le 0$; h_1 – начальный шаг спуска для следующей итерации. В алгоритме минимизации точка x^+ используется в качестве нового приближения минимума, субградиент g_m – для формирования направления спуска, а субградиент u – для проведения итерации метода решения неравенств.

В OM ищется приближение минимума одномерной функции $\varphi(\beta) = f(x - \beta s)$. Для последовательности $\beta_0 = 0$, $\beta_i = h_0 q_M^{i-1}$, $i \ge 1$, вычисляются $z_i = x - \beta_i s$, $r_i \in \partial f(z_i)$, $i = 0, 1, \ldots$ Пусть l – номер i, при котором впервые выполнится соотношение $(r_i, s) \le 0$. На основе информации $\{\gamma_0 = \beta_{l-1}, f_0 = f(z_{l-1}), g_0 = r_{l-1}, \gamma_1 = \beta_l, f_1 = f(z_l), g_1 = r_l\}$, заданной на отрезке локализации $[\gamma_0, \gamma_1]$ одномерного минимума, определяется γ^* – точка минимума одномерной кубической аппроксимации функции $\varphi(\beta)$ на отрезке локализации. Затем вычисляются огрубленное приближение минимума

$$\gamma_m = \begin{cases} 0.1\gamma_1, \text{ если } l = 1, \ \gamma^* \le 0.1\gamma_1, \\ \gamma_1, \text{ если } \gamma_1 - \gamma^* \le 0.2(\gamma_1 - \gamma_0), \\ \gamma_0, \text{ если } l > 1, \ \gamma^* - \gamma_0 \le 0.2(\gamma_1 - \gamma_0), \\ \gamma^* \text{ в остальных случаях,} \end{cases}$$

и начальный шаг спуска для следующей итерации $h_1 = q_m(h_0\gamma_1)^{1/2}$. В OM был использован следующий набор параметров $\{q_m = 0.8, q_M = 3\}$.

Алгоритм минимизации $r_{OM}(\alpha, \beta)$. Здесь обновление матрицы заменено на коррекцию диагональных элементов, а точный одномерный спуск – на приближенный.

1. Зададим: $H_0 = I$, $x_0 \in R^n$, $I = 0$; параметры α, β, удовлетворяющие (5); параметр $\varepsilon_0 = 10^{-8}$; параметры останова: N – максимально допустимое число итераций; ε_x – точность минимизации по аргументу; ε_g – точность минимизации по градиенту. Затем вычислим $g_0 \in \partial f(x_0)$. Если $g_0 = 0$, то x_0 – искомая точка минимума и можно закончить вычисления.

2. Положим $\pi_i = \max\limits_{1 \le j \le n} H_{i,j}$. Если $\pi_i \le \varepsilon_0^{1/2}$, то $H_i = H_i/\pi_i$, $h_i = h_i\pi_i^{1/2}$, $\pi_i = 1$. Если $(g_i, H_ig_i)/(g_i, g_i) \le \varepsilon_0\pi_i$, то $H_i = H_i + 10\varepsilon_0\pi_i I$.

3. Произведем одномерный спуск вдоль $s_i = H_ig_i/(g_i, H_ig_i)^{1/2}$: $OM(\{x_i, s_i, g_i, f_i, h_i\}; \{\gamma_i, f_{i+1}, g_{i+1}, \tilde{\gamma}_i, u_i, h_{i+1}\})$ и найдем новое приближение минимума $x_{i+1} = x_i - \gamma_i s_i$.

4. Если выполняется одно из условий окончания процесса минимизации $i > N$, $\|x_{i+1} - x_i\| \le \varepsilon_x$, $\|g_{i+1}\| \le \varepsilon_g$, то необходимо закончить вычисления.

5. Найдем $y_i = u_i - g_i$ и вектор p_i по формуле (7). Если $(p_i, H_ip_i) \le \varepsilon_0(y_i, H_iy_i)$, то необходимо произвести преобразование матрицы по формуле (16), в противном случае – по формуле (8).

6. Увеличить i на единицу и перейти к следующей итерации на п. 2.

При $\beta = 1$ алгоритм $r_{OM}(\alpha, \beta)$ реализует предложенный в (Крутиков, Арышев, 2004) алгоритм $r_{OM}(\alpha)$ (r-алгоритм с процедурой одномерной минимизации OM).

Таблица 1

Размерность (число задач)	Методы						
	число вычислений *nfg* (число решенных задач в группе)						
	МУ $\times 10^{-4}$	МШ $\times 10^{-4}$	МОС $\times 10^{-4}$	МЭ $\times 10^{-4}$	МРП $\times 10^{-4}$	$r_{OM}(\alpha) \times 10^{-4}$	$r_{OM}(\alpha, \beta) \times 10^{-4}$
5(8)	258 (8)	1090 (8)	202 (7)	4156 (8)	402 (8)	376 (8)	334 (8)
10(8)	346 (8)	1720 (7)	326 (8)	2361 (8)	805 (8)	746 (8)	688 (8)
15(8)	443 (7)	3448 (8)	599 (8)	3009 (4)	1113(8)	1245 (8)	1008 (8)
50(8)	955 (8)	5435 (6)	2456 (8)		3212(8)	5020 (8)	2959 (8)

В п. 5 алгоритма $r_{OM}(\alpha, \beta)$ производится итерация алгоритма по поиску решения неравенств с использованием субградиента u_i. Субградиент g_{i+1} необходим для формирования направления спуска. Формула (16) применяется в ситуациях, когда субградиенты u_i, g_i почти параллельны.

На итерации алгоритма $r_{OM}(\alpha, \beta)$ производятся в среднем два вычисления функции и субградиента, три умножения матрицы на вектор и затрачивается примерно $n(n + 1)$ операций умножения на преобразование симметричной матрицы.

6. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ИССЛЕДОВАНИЯ

Тестовые функции. В качестве первой группы тестов использовались известные задачи оптимизации, на которых в (Нестерова, Скоков, 1994) проведено подробное численное исследование программ негладкой оптимизации, а в (Скоков, 1997) – вариантов МУ (обозначения, описания и начальные точки для функций соответствуют (Нестерова, Скоков, 1994)):

1) (MAXQ) – максимум из квадрата координаты по всем координатам текущей точки;

2) (MAXL) – максимум модуля координаты по всем координатам текущей точки;

3) (GOFN) – разность максимума координаты по всем координатам текущей точки, умноженного на n, и суммы всех координат этой точки;

4) (HILB) – квадратичная функция с матрицей Гильберта;

5) (MXC) – максимум модуля координат с номером k, умноженных на k^3;

6) (SMD) – сумма модулей координат в текущей точке в степени k, умноженных на k^3;

7) (MXN) – максимум линейной комбинации норм векторов;

8) (PLN) – задача приближения данных полиномом степени $(n – 1)$.

Функции 1)–3) приведены в (Lemarecnal, 1982), а 4)–8) – в (Нестеров, Пурмаль, 1984).

Во второй группе представлены невыпуклые функции малой размерности, известные под названиями: 9) ROS 2; 10) WOOD 4; 11) POW 4.

Третью группу тестов составляют функции с высокой степенью вытянутости поверхностей уровня:

12) $f_{12}(x) = \sum_{i=1}^{n} x_i^2\, i, \quad x_0 = (10, 10, ..., 10);$

13) $f_{13}(x) = \sum_{i=1}^{n} x_i^2\, i^6, \quad x_0 = (10/1, 10/2, ..., 10/n);$

14) $f_{14}(x) = \sum_{i=1}^{n} x_i^2 (n/i)^6, \quad x_0 = (10, 10, ..., 10);$

15) (Скоков, 1997): $f_{15}(x) = \sum_{i=1}^{n-1} [1000(x_i - x_{i+1})^2 + (1 - x_{i+1})^2];$

Таблица 2

Параметры функции			Название метода число вычислений функции и субградиента *nfg*			
номер	ε	n	МРП	$r_{OM}(\alpha)$	$r_{OM}(\alpha, \beta)$	КНМ
15	10^{-5}	5–50	57–172	45–275	30–106	11–101
...	...	...	...	...	...	...
9	10^{-10}	2	47	65	59	38
10	10^{-10}	4	58	202	87	96
11	10^{-10}	4	53	61	60	51
...	...	...	...	...	...	...
12	10^{-10}	(100) 1000	(291) 1553	(249) 2072	(132)286	(106*) 350
13	10^{-10}	(100) 1000	(2127)27269	(2333) 34702	(859) 8285	(214*) 2238*
14	10^{-10}	(100) 1000	(612) 3572	(521) 3644	(351) 1823	(348*) 3920*
15	10^{-5}	(100) 1000	(262) 541	(448)2179	(175) 298	(201) 364*
16	10^{-10}	(100) 1000	(231) 4460	(172) 1094	(109)213	(520*) 2332*
17	10^{-4}	(100) 1000	(1558) 18088	(3098) 40738	(1873)27370	–
18	10^{-4}	(100) 1000	(3006) –	(3817)57336	(2084)28105	–

16) (Brodlie, 1972): $f_{16}(x) = \left(\sum_{i=1}^{n} x_i^2 i\right)^r, \quad x_0 = (1, 1, ..., 1), \ r = 2;$

17) MXC: $f_{17}(x) = \max_{1 \le i \le n} (\left| x_i \right| i^3), \quad x_0 = (10/1, 10/2, ..., 10/n);$

18) $f_{18}(x) = \sum_{i=1}^{n} \left| x_i \right| i^3, \quad x_0 = (10/1, 10/2, ..., 10/n).$

Они использовались в качестве большеразмерных тестов при $n = 100$ и $n = 1000$.

Результаты тестирования. В табл. 1 приведены результаты счета на первых восьми функциях для метода эллипсоидов (МЭ) (Немировский, Юдин, 1979), r-алгоритма (МШ) (Шор, 1979), метода ортогонального спуска (МОС) (Скоков, Щепакин, 1994), метода уровней (МУ) (Гольштейн, Немировский, Нестеров, 1995), взятые. в (Нестерова, Скоков, 1994), и алгоритмов с растяжением пространства $r_{OM}(\alpha^2 = 30, \beta^{\ 2} = 0.2)$ (здесь $(\alpha\beta)^2 = 6$ – результирующий коэффициентом растяжения), $r_{OM}(\alpha^2 = 6)$, МРП (Крутиков, Петрова, 2003) при одинаковых коэффициентах растяжения, с процедурой одномерной минимизации OM при $q_m = 0.8$ и $q_M = 3$. Во всех методах в каждой точке функция и субградиент вычислялись одновременно. В ячейках табл. 1 первым стоит *nfg* – число вычислений функции и субфадиента по каждому методу, оно равно моменту, когда выполнилось условие $f_k - f^* \le \varepsilon$, в скобках указано число решенных задач. Величина е приведена в таблице под названием метода.

На этой системе тестов МУ по критерию *nfg* значительно превосходит исследуемые методы класса R_{PCM}. По критерию времени счета этот вывод будет справедлив при достаточно высокой трудоемкости вычисления функции и субградиента (Нестерова, Скоков, 1994). В табл. 2 для функции 15 при $n = 5, ..., 50$ приведено минимальное (M_{MIN}) и максимальное (M_{MAX}) число *nfg* для достижения заданной точности. Для МУ при той же точности в работе (Скоков, 1997) получено: M_{MIN} =П8, M_{MAX} = 1000, что существенно выше затрат алгоритма $r_{OM}(\alpha, \beta)$. Последнее вместе с результатами табл. 1 и результатами счета на невыпуклых функциях 9–11, приведенными в табл. 2, свидетельствует о различии областей эффективного применения методов из R_{PCM} и МУ.

Результаты счета по второй и третьей группам функций представлены в табл. 2, где в круглых скобках указаны результаты при $n = 100$. На гладких функциях 9–16 приведены результаты работы квазиньютоновского метода с формулой BPO8 (КНМ), который был реализован с точным и неточным одномерными спусками, использующими кубическую интерполяцию (Денис, Шнабель, 1988). КНМ с неточным спуском на функциях 9–11 показывает результаты, аналогичные полученным в (Brodlie, 1972). При малой сложности и размерности квадратичной функции его число *nfg* незначительно превышает число итераций при точном спуске, но при росте сложности и размерности (функции 13–15) его эффективность снижается. Поэтому в таблице приведено наименьшее из двух чисел: числа *nfg* при неточном одномерном спуске и удвоенного числа итераций при точном спуске, которое в случае использования помечено знаком “*”. Удвоение сделано с учетом минимальных затрат на одномерный спуск. Например, на функции 14 при *n,* равном 100 и 1000, затраты *nfg* для КНМ при неточном спуске составляют соответственно 1704 и 14107, что существенно выше затрат, приведенных в табл. 2.

На гладких невыпуклых функциях алгоритмы 9–11 из R_{PCM} работоспособны и практически эквивалентны методу КНМ. На функции 16, характеризующейся высокой степенью изменчивости гессиана, затраты *nfg* алгоритма $r_{OM}(\alpha, \beta)$ значительно ниже затрат КНМ. На квадратичных функциях 12–15 при $n \le 100$ затраты КНМ метода меньше затрат алгоритма $r_{OM}(\alpha, \beta)$, причем это преимущество уменьшается с ростом размерности. При $n = 1000$ алгоритм $r_{OM}(\alpha, \beta)$ уступает методу КНМ только на функции 13. Функции 13 и 14 имеют одинаковые границы спектра собственных значений, но у функции 14 плотность собственных значений гессиана смещена в область малых значений. Подобное смещение увеличивает затраты для метода КНМ и уменьшает – для методов из R_{PCM}. Замедление скорости сходимости КНМ на функции 14 и его медленная сходимость на функции 16 закономерны. Оно объясняется быстрым ростом матрицы метрики в обследованном подпространстве и связанным с этим замедлением выхода в новое подпространство, обусловленным понижением скорости сходимости и точности процесса построения матриц метрики и необходимостью повышения точности минимизации в обследованном подпространстве в обратной пропорции к степени роста матрицы метрики.

Как следует из данных, приведенных в таблицах, затраты *nfg* алгоритма $r_{OM}(\alpha, \beta)$ ниже, чем у алгоритмов $r_{OM}(\alpha)$ и МРП, причем с ростом размерности его преимущества более ощутимы. Например, метод МРП на функции 18 перестает сходиться уже при $n = 400$.

СПИСОК ЛИТЕРАТУРЫ

Бэр К., Гольштейн Е.Г., Соколов Н.А. (2000): Об использовании метода уровней для минимизации выпуклых функций, не все значения которых конечны // *Экономика и мат. методы.* Т. 36. № 4.

Гольштейн Е.Г., Немировский А.С., Нестеров Ю.Е. (1995): Метод уровней, его обобщения и приложения // *Экономика и мат. методы.* Т. 31. Вып. 3.

Гольштейн Е.Г. (2006): Метод минимизации негладких квазивыпуклых функций, использующий неточные данные // *Экономика и мат. методы.* Т. 42. № 2.

Демьянов В.Ф., Васильев Л.В. (1981): Недифференцируемая оптимизация. М.: Наука.

Денис Дж., Шнабель Р. (1988): Численные методы безусловной оптимизации и решения нелинейных уравнений. М.: Мир.

Крутиков В.Н., Петрова Т.В. (2003): Релаксационный метод минимизации с растяжением пространства в направлении субградиента // *Экономика и мат. методы.* Т. 39. № 1.

Крутиков В.Н. (2003): Одноранговое семейство релаксационных субградиентных методов с растяжением пространства // *Электронный журнал “Исследовано в России”.* № 209.

Крутиков В.Н., Арышев Д.В. (2004): Реализация алгоритмов однорангового семейства субградиентных методов // *Электронный журнал “Исследовано в России”.* № 43.

Крутиков В.Н., Арышев Д.В. (2005): Субградиентный метод с растяжением–сжатием пространства // *Вестник КемГУ.* Вып. 4 (22).

Немировский А.С., Юдин Д.Б. (1979): Сложность задач и эффективность методов оптимизации. М.: Наука.

Нестерова С.И., Скоков В.А. (1994): Численный анализ программ негладкой безусловной оптимизации // *Экономика и мат. методы.* Т. 30. Вып. 2.

Нестеров Ю.Е., ПурмальЕ.И. (1984): Анализ эффективности методов негладкой оптимизации. М.: ЦЭМИ АН СССР.

Скоков В.А. (1997): Варианты метода уровней для минимизации негладких выпуклых функций и их численное исследование // *Экономика и мат. методы.* Т. 33. Вып. 1.

Скоков В.А., Щепакин М.Б. (1994): Численный анализ одного варианта метода ортогонального спуска *Кибернетика и системный анализ.* № 2.

Шор Н.З. (1979): Методы минимизации недифференцируемых функций и их приложения. Киев: Наукова думка.

Brodlie K. (1972): An Assement of Two Aapproaches to Variable Metric Methods // *Math. Program.* Vol. 7. № 12.

Lemarechal C., Nemirovskii A., Nesterov Yu. (1995): New Variants of Bundle Methods // *Math. Program. Seria B.* Vol. 69. № 1.

Lemarechal C. (1974): An Algorithm for Minimizing Convex Functions. In: *"Proceedings of IFIP Congress"*. North-Holland, Amsterdam.

Lemarechal C. (1982): Numerical Experiments in Nonsmoth Optimization. In: *"Progress in Nondifferentiable Optimization".* Luxenburg: IIASA-report.

Wolfe P. (1974): Note on a Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions // *Math. Program.* Vol. 7. № 3.

Поступила в редакцию
04.04.2008 г.

The Family of Relaxation Sub-Gradient Methods with Double-Rank Correction of Matrixes of the Metrics

V. N. Kroutikov, T. A. Gorskaya

The family of relaxation sub-gradient methods with double-rank correction of matrixes of the metrics for solving non-smooth absolute optimization is suggested. Convergence of its algorithms has been proven with strictly convex functions. Results of numerical research are presented of one algorithm of family.