ИЗВЕСТИЯ ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЙ

2000 МАТЕМАТИКА № 12 (463)

УДК 519.853

И.Я. ЗАБОТИН

ОБ УСТОЙЧИВОСТИ АЛГОРИТМОВ БЕЗУСЛОВНОЙ МИНИМИЗАЦИИ ПСЕВДОВЫПУКЛЫХ ФУНКЦИЙ

Статья посвящена исследованию устойчивости известных и новых алгоритмов безусловной минимизации псевдовыпуклых функций к ошибкам вычисления градиентов. Вопросы устойчивости методов выпуклого программирования обсуждаются во многих работах (напр., [1]–[5]). Используемые в них подходы, как правило, существенно опираются на то, что исследуемые методы применяются для минимизации выпуклых, а не псевдовыпуклых функций. В статье предлагается один способ исследования устойчивости довольно большой группы алгоритмов, пригодных для минимизации псевдовыпуклых функций. Способ основан на разработанной общей схеме построения алгоритмов минимизации псевдовыпуклых функций, в котором заложена возможность неточного вычисления градиента в итерационных точках. В этой схеме переход от одной точки к другой происходит в некоторых линейных многообразиях произвольной размерности. За счет большой свободы в выборе многообразий схема допускает различные реализации, среди которых есть как известные, так и новые алгоритмы. Поскольку сходимость и оценки скорости сходимости общей схемы доказываются с учетом неточного вычисления градиентов, то тем самым исследуется влияние помех и на сходимость алгоритмов, которые вкладываются в общую схему. Приводятся примеры использования предлагаемого подхода к исследованию устойчивости известных и новых одношаговых и многошаговых алгоритмов минимизации гладких псевдовыпуклых функций.

Решается задача

$$\min_{x\in R_n} f(x), \tag{1}$$

где $f(x)$ — достигающая своего минимального значения в n-мерном евклидовом пространстве R_n непрерывно дифференцируемая псевдовыпуклая функция.

Положим $f^* = \min\limits_{x\in R_n} f(x)$, $X^* = \{x \in R_n : f(x) = f^*\}$, $K = \{0, 1, \dots\}$, $H = \{1, \dots, n\}$, $f'(x)$ — градиент функции $f(x)$ в точке $x = (\xi_1, \dots, \xi_n)$. Напомним, что функция $f(x)$ называется псевдовыпуклой в R_n, если для любых $x, y \in R_n$ выполняется импликация $\langle f'(x), y - x\rangle \ge 0 \Rightarrow f(y) \ge f(x)$ (напр., [6], с. 522).

Опишем сначала общую схему решения задачи (1), которая вырабатывает последовательность приближений x_k, $k \in K$. При этом будем считать, что градиент в точке x_k отыскивается с некоторой погрешностью $r_k \in R_n$, т. е. на k-м шаге вычисляется вектор $\widetilde{f}'(x_k) = f'(x_k) + r_k$ вместо $f'(x_k)$. Векторы r_k названы в ([5], с. 96) детерминированными помехами в вычислении градиента, причем абсолютными, если $\|r_k\| \le \varepsilon$, $\varepsilon \ge 0$, и относительными, если $\|r_k\| \le \varepsilon\|f'(x_k)\|$. Там же и, например, в ([7], с. 92) обсуждаются причины возникновения этих видов помех.

Предлагаемая схема решения задачи (1) заключается в следующем. Выбирается точка $x_0 \in R_n$ начального приближения и числа $0 < \lambda' \le \lambda'' < \infty$. Пусть построена точка $x_k \notin X^*$, $k \in K$.

1. Вычисляется $\widetilde{f}'(x_k)$. Выбираются число

$$\lambda' \le \lambda_k \le \lambda'' \tag{2}$$

и вектор $p_k \in R_n$ так, чтобы $g_k \triangleq \lambda_k \widetilde{f}'(x_k) + p_k \ne 0$.

2. Задается множество индексов $H_k \subset H$, выбираются векторы h_i^k, $i \in H_k$, так, чтобы вектор g_k был их линейной комбинацией, и строится линейное многообразие

$$M(x_k) = \Big\{ x \in R_n : x = x_k + \sum_{i \in H_k} \alpha_i h_i^k,\ \alpha_i \in R_1 \Big\}.$$

3. Отыскивается такая точка $y_k \in M(x_k)$, что

$$f(y_k) \le f_k^* + \delta_k, \quad \delta_k \ge 0, \tag{3}$$

либо

$$f(y_k) \le (1 - \nu_k) f(x_k) + \nu_k f_k^*, \quad 0 < \nu \le \nu_k \le 1, \tag{4}$$

где $f_k^* = \min\limits_{x \in U(x_k)} f(x)$, $U(x_k) = \{x \in R_n : x = x_k + \alpha g_k,\ \alpha \in R_1\}$.

4. Приближение $x_{k+1} \in R_n$ выбирается из условия

$$f(x_{k+1}) \le f(y_k). \tag{5}$$

Ниже для обеспечения сходимости будут введены дополнительные условия на выбор чисел δ_k, вектора p_k и величину погрешности $\|r_k\|$.

В случае $M(x_k) = R_n$ переход от точки x_k к точке x_{k+1} можно осуществить следующим образом. Любым известным алгоритмом строится минимизирующая последовательность $\{z_s\}$ при $z_0 = x_k$. Полагается $y_k = z_s$, если для z_s выполняется условие (3) или (4). Это замечание позволяет на основе общей схемы при условии ее сходимости строить смешанные алгоритмы, в которых можно от шага к шагу менять способы построения направлений и шаговых множителей.

Обратим внимание на то, что в (3), (4) использовано значение f_k^* минимума функции $f(x)$ на прямой $U(x_k)$, а не на многообразии $M(x_k)$. Значит, схема позволяет одновременно находить приближенное минимальное значение функции $f(x)$ на прямой и любым известным методом проводить спуск из точки x_k в многообразии $M(x_k)$ до выполнения неравенств (3) или (4). Подчеркнем, что точка x_{k+1} может и не принадлежать многообразию $M(x_k)$.

Ниже, если специально не оговаривается, будем считать, что последовательность $\{x_k\}$ построена по предложенной схеме, и $f'(x_k) \ne 0$ $\forall k \in K$.

Лемма 1. *Пусть при некотором $k \in K$ вектор r_k удовлетворяет неравенству*

$$\langle f'(x_k), r_k \rangle \ge -\gamma_k \|f'(x_k)\|^2, \tag{6}$$

где $0 \le \gamma_k \le \gamma < 1$, а вектор p_k — неравенству

$$\langle f'(x_k), p_k \rangle \ge -\sigma_k \|f'(x_k)\|^2, \tag{7}$$

где $0 \le \sigma_k \le \sigma < \lambda'(1-\gamma)$. Тогда

$$\langle f'(x_k), g_k \rangle \ge (\lambda'(1-\gamma) - \sigma) \|f'(x_k)\|^2.$$

Доказательство непосредственно следует из равенства $\langle f'(x_k), g_k \rangle = \lambda_k \|f'(x_k)\|^2 + \lambda_k \langle f'(x_k), r_k \rangle + \langle f'(x_k), p_k \rangle$ и условий (6), (7), (2).

Заметим, что если в лемме 1 условие (7) заменить, например, условием

$$\langle \widetilde{f}'(x_k), p_k \rangle \ge \|r_k\|\,\|p_k\| - \sigma_k \|f'(x_k)\|^2,$$

то утверждение ее останется справедливым, поскольку из этого неравенства следует $\langle \widetilde{f}'(x_k), p_k \rangle + \langle r_k, p_k \rangle \ge \|r_k\|\,\|p_k\| - \sigma_k \|f'(x_k)\|^2$, а значит, и (7). Утверждение леммы также останется в силе, если (7) заменить условием $\langle \widetilde{f}'(x_k), p_k \rangle \ge \theta_k \|r_k\|\,\|p_k\|$, $\theta_k \ge 1$.

Лемма 2. *Пусть выполняются условия леммы 1. Тогда точку $y_k \in M(x_k)$ можно выбрать так, что вместе с условием (3) или (4) выполнится неравенство*

$$f(y_k) < f(x_k). \tag{8}$$

Доказательство. Пусть точка $x_k^* \in M(x_k)$ такова, что $f(x_k^*) = m_k = \min\limits_{x \in M(x_k)} f(x)$. Значит, $f(x_k^*) \le f(x_k)$. Допустим, что последнее соотношение выполняется как равенство. Тогда в силу выбора x_k^*

$$f(x_k) \le f\Bigl(x_k + \sum_{i \in H_k} \alpha_i h_i^k\Bigr) \quad \forall \alpha_i \in R_1, \;\; i \in H_k. \tag{9}$$

Поскольку $f'(x_k) \ne 0$, то по лемме 1 вектор $-g_k$ является направлением спуска для функции $f(x)$ в точке x_k, и найдется такое число $\tau < 0$, что $f(x_k + \tau g_k) < f(x_k)$. По условию выбора векторов h_i^k существуют такие числа τ_i^k, $i \in H_k$, что $g_k = \sum\limits_{i \in H_k} \tau_i^k h_i^k$. Следовательно, $f(x_k + \sum\limits_{i \in H_k} \alpha_i^k h_i^k) < f(x_k)$, где $\alpha_i^k = \tau \tau_i^k$. Это неравенство противоречит утверждению (9). Таким образом, $f(x_k^*) < f(x_k)$, а поскольку $m_k \le f_k^*$, то в качестве точки y_k, удовлетворяющей условиям (3), (8) или (4), (8), можно взять точку x_k^*. $\square$

Как показано в лемме 2, последовательность $\{f(x_k)\}$ может быть построена с условием релаксации. Исследуем сначала сходимость релаксационного варианта схемы.

Теорема 1. *Пусть множество $X^0 = \{x \in R_n : f(x) \le f(x_0)\}$ ограничено, векторы r_k, p_k удовлетворяют неравенствам* (6), (7),

$$\|r_k\| \le c_1 < \infty, \quad \|p_k\| \le c_2 < \infty \quad \forall k \in K, \tag{10}$$

$$\delta_k \to 0, \quad k \to \infty, \tag{11}$$

и точки y_k для всех $k \in K$ выбраны согласно (3), (8) *или* (4), (8). *Тогда для последовательности $\{x_k\}$ справедливо равенство*

$$\lim_{k \to \infty} f(x_k) = f^*. \tag{12}$$

Доказательство. Так как согласно (5), (8) последовательность $\{f(x_k)\}$ монотонно убывает и по условию $f^* > -\infty$, то существует

$$\lim_{k \to \infty} f(x_k) \ge f^*, \tag{13}$$

и, значит,

$$\lim_{k \to \infty} (f(x_{k+1}) - f(x_k)) = 0. \tag{14}$$

Допустим, что неравенство (13) выполняется как строгое, т. е. утверждение теоремы неверно. Согласно неравенствам (2) и условиям теоремы последовательности $\{x_k\}$, $\{r_k\}$, $\{p_k\}$, $\{\lambda_k\}$, $k \in K$, ограничены. Следовательно, можно выделить такую подпоследовательность $K_1 \subset K$, что подпоследовательности $\{x_k\}$, $\{r_k\}$, $\{p_k\}$, $\{\lambda_k\}$, $k \in K_1$, будут сходящимися. Пусть $\overline{x}$, $\overline{r}$, $\overline{p}$, $\overline{\lambda}$ — соответственно их предельные точки. Положим $\overline{g} = \overline{\lambda}(f'(\overline{x}) + \overline{r}) + \overline{p}$. Поскольку $f'(\overline{x}) \ne 0$ и функция $f(x)$ непрерывно дифференцируема, то по лемме 1 $\langle f'(\overline{x}), \overline{g} \rangle \ge (\lambda'(1-\gamma) - \sigma)\|f'(\overline{x})\|^2 > 0$, и $-\overline{g}$ является направлением спуска для функции $f(x)$ в точке $\overline{x}$. Значит, существует такое число $\tau < 0$, что $f(\overline{x} + \tau \overline{g}) - f(\overline{x}) = \eta < 0$. Выберем такие окрестности ω_1, ω_2, ω_3, ω_4 точек $\overline{x}$, $\overline{r}$, $\overline{p}$ и числа $\overline{\lambda}$ соответственно, что

$$f(x + \tau(\lambda(f'(x) + r) + p)) - f(x) \le \eta/2 \quad \forall x \in \omega_1, \;\; r \in \omega_2, \;\; p \in \omega_3, \;\; \lambda \in \omega_4.$$

Тогда найдется такой номер N, что $f(x_k + \tau(\lambda_k(f'(x_k) + r_k) + p_k)) - f(x_k) \le \eta/2 \;\forall k \ge N$, $k \in K_1$, а поскольку $f_k^* \le f(x_k + \tau g_k)$, то $f_k^* - f(x_k) \le \eta/2 \;\forall k \ge N$, $k \in K_1$. Согласно (3)–(5) $f(x_{k+1}) - \delta_k \le f_k^*$ или $(f(x_{k+1}) - f(x_k))/\nu_k + f(x_k) \le f_k^* \;\forall k \in K$. Следовательно, для всех $k \ge N$, $k \in K_1$, выполняются неравенства $f(x_{k+1}) - f(x_k) - \delta_k \le \eta/2$ или $f(x_{k+1}) - f(x_k) \le \nu\eta/2$, которые в силу (14), (11) и условия $\eta < 0$ при достаточно больших значениях $k \in K_1$ становятся противоречивыми. $\square$

Исследуем сходимость нерелаксационного варианта схемы с выбором точек y_k согласно условию (3).

Лемма 3. *Пусть $f'(x)$ удовлетворяет в R_n условию Липшица с константой L, подмножества $K' \subset K$ и $K'' \subset K$ таковы, что точки y_k при построении последовательности $\{x_k\}$ выбираются для всех $k \in K'$ согласно* (3), *а для всех $k \in K''$ — согласно* (4). *Тогда*

$$f(x_k) - f(x_{k+1}) \geq \alpha\langle f'(x_k), g_k\rangle - (L\alpha^2/2)\|g_k\|^2 - \delta_k \quad \forall \alpha \in R_1, \quad k \in K', \tag{15}$$

$$f(x_k) - f(x_{k+1}) \geq \nu_k\alpha\langle f'(x_k), g_k\rangle - (L\alpha^2/2)\|g_k\|^2 \qquad \forall \alpha \in R_1, \quad k \in K''. \tag{16}$$

Утверждения леммы 3 не связаны с условиями (6), (7) задания векторов r_k, p_k в схеме, поэтому ее доказательство полностью повторяет обоснование аналогичной леммы из [8].

Теорема 2. *Пусть для всех $k \in K$ векторы r_k, p_k удовлетворяют условиям* (6), (7), (10), *точки y_k при построении последовательности $\{x_k\}$ выбраны согласно* (3), *а числа δ_k, $k \in K$, таковы, что $\sum\limits_{k=0}^{\infty} \delta_k = \delta < \infty$. Если множество*

$$X_\delta^0 = \{x \in R_n : f(x) \leq f(x_0) + \delta\}$$

ограничено, $f'(x)$ удовлетворяет в R_n условию Липшица с константой L, то справедливо равенство (12).

Доказательство. Из (15) и леммы 1 следует неравенство $f(x_k) - f(x_{k+1}) \geq \alpha(\lambda'(1-\gamma) - \sigma)\|f'(x_k)\|^2 - (L\alpha^2/2)\|g_k\|^2 - \delta_k$ $\forall \alpha \geq 0$, $k \in K$. Значит,

$$f(x_k) - f(x_{k+1}) \geq \max_{\alpha\geq 0}(\alpha(\lambda'(1-\gamma) - \sigma)\|f'(x_k)\|^2 - (L\alpha^2/2)\|g_k\|^2) - \delta_k =$$

$$= (\lambda'(1-\gamma) - \sigma)^2\|f'(x_k)\|^4/(2L\|g_k\|^2) - \delta_k, \quad k \in K. \tag{17}$$

Далее, как и в [8], используя методику Ф.П. Васильева ([9], сс. 93, 265), нетрудно показать существование предела (13) и равенство (14), а также ограниченность последовательности $\{\|g_k\|\}$. Тогда из (17), (14) с учетом выбора чисел δ_k следует равенство $\lim\limits_{k\to\infty} f'(x_k) = 0$, а значит, и (12). $\square$

Отметим, что сходимость схемы обоснована в теоремах 1, 2 без предположения о стремлении к нулю последовательности $\{\|r_k\|\}$.

Получим оценки скорости сходимости схемы. Пусть погрешности r_k вычисления градиента удовлетворяют условию

$$\|r_k\| \leq \varepsilon\|f'(x_k)\|, \tag{18}$$

где $0 \leq \varepsilon < 1$, а векторы p_k — условию

$$\|p_k\| \leq \Delta\|f'(x_k)\|, \tag{19}$$

где $0 \leq \Delta < \lambda'(1-\varepsilon)$, либо

$$\|p_k\| \leq \Delta'\|\widetilde{f}'(x_k)\|, \quad 0 \leq \Delta' < \lambda'(1-\varepsilon)/(1+\varepsilon). \tag{20}$$

Заметим, что условие (18) можно заменить и другими аналогичными требованиями, например,

$$\|r_k\| \leq \varepsilon'\|\widetilde{f}'(x_k)\|, \quad 0 \leq \varepsilon' < 1/2. \tag{21}$$

Однако, в таком случае $\|\widetilde{f}'(x_k)\| \leq \|f'(x_k)\| + \varepsilon'\|\widetilde{f}'(x_k)\|$ или $\|\widetilde{f}'(x_k)\| \leq (1/(1-\varepsilon'))\|f'(x_k)\|$, и из (21) следует (18) при $\varepsilon = \varepsilon'/(1-\varepsilon')$. Если для векторов r_k, p_k выполняются (18), (20), то $\|p_k\| \leq \Delta'(1+\varepsilon)\|f'(x_k)\|$, а значит, справедливо (19), и условия (18), (19) являются более общими, чем (18), (20).

Положим $(f(x_k) - f^*)/\|f'(x_k)\| = \theta(x)$ $\forall x \notin X^*$.

Теорема 3. *Пусть функция $f(x)$ такова, что*

$$\theta(x) \le \theta < \infty \quad \forall x \in X_\delta^0 \setminus X^*, \tag{22}$$

$f'(x)$ удовлетворяет в R_n условию Липшица с константой L, при построении последовательности $\{x_k\}$ точки y_k находятся для всех $k \in K$ только из условия (3), *причем $0 \le \delta_k \le \eta k^{-2}$, $\eta > 0$, $k = 1, 2, \dots$, либо только из условия* (4), *а векторы r_k, p_k выбираются согласно* (18), (19). *Тогда справедлива оценка*

$$0 \le a_k = f(x_k) - f^* \le c/k, \quad c > 0, \quad k = 1, 2, \dots \tag{23}$$

Доказательство. Так как $\langle f'(x_k), g_k\rangle = \lambda_k \langle f'(x_k), f'(x_k) + r_k\rangle + \langle f'(x_k), p_k\rangle \ge \lambda_k \|f'(x_k)\|^2 - \lambda_k \|f'(x_k)\| \, \|r_k\| - \|f'(x_k)\| \, \|p_k\|$, $k \in K$, то из (18), (19), (2) следует, что

$$\langle f'(x_k), g_k\rangle \ge (\lambda'(1-\varepsilon) - \Delta)\|f'(x_k)\|^2 > 0 \quad \forall k \in K. \tag{24}$$

Кроме того, из (18), (19), (2) и неравенства $\|g_k\| \le \lambda_k \|\widetilde{f}'(x_k)\| + \|p_k\|$ легко получается оценка

$$\|g_k\| \le (\lambda''(1+\varepsilon) + \Delta)\|f'(x_k)\| \quad \forall k \in K. \tag{25}$$

Докажем (23), считая сначала, что точки y_k находятся при всех $k \in K$ согласно (3). Из (15), (24), (25) для всех $\alpha \ge 0$, $k \in K$ следует неравенство $f(x_k) - f(x_{k+1}) \ge \alpha(\lambda'(1-\varepsilon) - \Delta)\|f'(x_k)\|^2 - \alpha^2 (L(\lambda''(1+\varepsilon) + \Delta)^2/2)\|f'(x_k)\|^2 - \delta_k$. Отсюда

$$a_k - a_{k+1} = f(x_k) - f(x_{k+1}) \ge \max_{\alpha \ge 0} (\alpha(\lambda'(1-\varepsilon) - \Delta) - \alpha^2 L(\lambda''(1+\varepsilon) + \Delta)^2/2)\|f'(x_k)\|^2 - \\ - \delta_k = (\lambda'(1-\varepsilon) - \Delta)^2 \|f'(x_k)\|^2 / (2L(\lambda''(1+\varepsilon) + \Delta)^2) - \delta_k \quad \forall k \in K, \tag{26}$$

а значит, для всех $k \in K$ выполняются неравенства $f(x_{k+1}) \le f(x_k) + \delta_k$. Суммируя их по k от 0 до $m-1$, где $m \in K \setminus \{0\}$, получим $f(x_m) \le f(x_0) + \delta$, $m = 1, 2, \dots$ Таким образом, $x_k \in X_\delta^0$ $\forall k \in K$, и согласно (22)

$$a_k \le \theta \|f'(x_k)\| \quad \forall k \in K. \tag{27}$$

Тогда из (26), (27) имеем $a_k - a_{k+1} \ge a_k^2 (\lambda'(1-\varepsilon) - \Delta)^2 / (2L\theta^2(\lambda''(1+\varepsilon) + \Delta)^2) - \eta k^{-2}$, $k = 1, 2, \dots$ Значит, $a_{k+1} \le a_k - a_k^2/\eta' + \eta' k^{-2}$, $k = 1, 2, \dots$, где $\eta' = \max\{\eta, 2L\theta^2(\lambda''(1+\varepsilon) + \Delta)^2/(\lambda'(1-\varepsilon) - \Delta)^2\}$. Из последнего неравенства и леммы 2.3.5 ([9], с. 95) следует (23).

Пусть теперь точки y_k выбираются для всех $k \in K$ согласно (4). Из (16), (24), (25), (4) для всех $\alpha \ge 0$, $k \in K$ следует оценка $f(x_k) - f(x_{k+1}) \ge \nu\alpha(\lambda'(1-\varepsilon) - \Delta)\|f'(x_k)\|^2 - (L\alpha^2/2)(\lambda''(1+\varepsilon) + \Delta)^2 \|f'(x_k)\|^2$. Отсюда $a_k - a_{k+1} = f(x_k) - f(x_{k+1}) \ge \max_{\alpha \ge 0} (\alpha\nu(\lambda'(1-\varepsilon) - \Delta) - \alpha^2 L(\lambda''(1+\varepsilon) + \Delta)^2/2)\|f'(x_k)\|^2 = \nu^2(\lambda'(1-\varepsilon) - \Delta)^2 \|f'(x_k)\|^2 / (2L(\lambda''(1+\varepsilon) + \Delta)^2) > 0$ $\forall k \in K$. Поскольку $x_k \in X^0$ $\forall k \in K$, то справедливо (27). Тогда из последнего неравенства и леммы 2.3.4 ([9], с. 94) с учетом (27) следует оценка (23). $\square$

Отметим, что выбор константы θ в неравенстве (22) для выпуклой и сильно выпуклой функции $f(x)$ обсуждается в [8].

Теорема 4. *Пусть функция $f(x)$ дважды дифференцируема, $f'(x)$ удовлетворяет в R_n условию Липшица с константой L, существует такое число $\mu > 0$, что*

$$\langle f''(x)\xi, \xi\rangle \ge \mu \|\xi\|^2 \quad \forall x, \xi \in R_n, \tag{28}$$

где $f''(x)$ — матрица вторых производных функции $f(x)$ в точке x. Пусть векторы r_k, p_k удовлетворяют условиям (18), (19), *точки y_k выбираются согласно* (3) *или* (4), *причем $\delta_k = 0$, $\nu_k = \nu = 1$ $\forall k \in K$. Тогда для последовательности $\{x_k\}$ справедливы оценки*

$$0 \le a_k \le a_0 q^k, \quad \|x_k - x^*\|^2 \le (2/\mu) a_0 q^k \quad \forall k \in K, \tag{29}$$

где $0 < q = 1 - \mu(\lambda'(1-\varepsilon) - \Delta)^2/(L(\lambda''(1+\varepsilon) + \Delta)^2) < 1$, x^* — решение задачи (1).

Доказательство теоремы опирается на методику работы ([9], с. 267) и почти полностью повторяет обоснование аналогичных оценок из [8].

Заметим, что теорема 4 справедлива не для псевдовыпуклых, а для выпуклых и сильно выпуклых функций, удовлетворяющих (28).

Теорема 5. *Пусть последовательность* $\{x_k\}$*, построенная по предложенной схеме, релаксационна, погрешности* r_k *удовлетворяют условию* (6) *или* (18)*,* $a_k \le \overline{a}$*, а числа* θ_k *таковы, что* $0 < \theta_k \le \|\widetilde{f}'(x_k)\|/\overline{a}$ *для всех* $k \in K$*. Тогда справедлива оценка*

$$f(x_m) - f^* \le a_0 \left[1 + a_0 \sum_{k=0}^{m-1} \frac{\theta_k^2 (f(x_k) - f(x_{k+1}))}{\|\widetilde{f}'(x_k)\|^2} \right]^{-1}, \quad m = 1, 2, \dots$$

Доказательство. Согласно (6), (18) $r_k \ne -f'(x_k)$, т. е. $\|\widetilde{f}'(x_k)\| > 0$ и указанная в условиях теоремы последовательность $\{\theta_k\}$ существует. Очевидно, $\theta_k a_k \le \|\widetilde{f}'(x_k)\|$ $\forall k \in K$. Значит, $a_k - a_{k+1} = f(x_k) - f(x_{k+1}) = \frac{f(x_k) - f(x_{k+1})}{\|\widetilde{f}'(x_k)\|^2} \|\widetilde{f}'(x_k)\|^2 \ge \frac{\theta_k^2 (f(x_k) - f(x_{k+1}))}{\|\widetilde{f}'(x_k)\|^2} a_k^2$. Отсюда и из леммы 9.2.14 ([10], с. 170) следует утверждение теоремы. $\square$

Заметим, что в условиях теоремы можно положить, например, $\overline{a} = a_0$. Некоторые оценки для алгоритмов псевдовыпуклого программирования получены с использованием методики В.Г. Карманова в [11]. При обсуждении ниже устойчивости конкретных алгоритмов понадобятся следующие вспомогательные утверждения.

Лемма 4. *Пусть вектор* r_k *удовлетворяет неравенству* (18)*, где* $\varepsilon \ge 0$*. Тогда выполняется условие* (6) *при* $\gamma_k = (1 + \varepsilon^2)/2$*.*

Доказательство. Так как $\|\widetilde{f}'(x_k)\|^2 = \|f'(x_k)\|^2 + 2\langle f'(x_k), r_k\rangle + \|r_k\|^2 \ge 0$, то $\langle f'(x_k), r_k\rangle \ge -(1/2)\|f'(x_k)\|^2 - (1/2)\|r_k\|^2$, и в силу (18) $\langle f'(x_k), r_k\rangle \ge -(1/2)(\|f'(x_k)\|^2 + \varepsilon^2\|f'(x_k)\|^2) = -((1+\varepsilon^2)/2)\|f'(x_k)\|^2$. $\square$

Лемма 5. *Пусть векторы* r_k*,* p_k *таковы, что выполняются неравенства* (18)*,* (19)*, где* $\varepsilon \ge 0$*,* $\Delta \ge 0$*, и, кроме того,*

$$\langle \widetilde{f}'(x_k), p_k\rangle \ge -\widetilde{\sigma}_k \|\widetilde{f}'(x_k)\|^2, \quad \widetilde{\sigma}_k \ge 0.$$

Тогда справедливо неравенство (7)*, где* $\sigma_k = \varepsilon\Delta + (1+\varepsilon)^2\widetilde{\sigma}_k$*.*

Доказательство. По условию $\langle f'(x_k), p_k\rangle + \langle r_k, p_k\rangle \ge -\widetilde{\sigma}_k\|f'(x_k) + r_k\|^2$. Отсюда с учетом (18), (19) следуют оценки $\langle f'(x_k), p_k\rangle \ge -\|r_k\|\,\|p_k\| - \widetilde{\sigma}_k(\|f'(x_k)\|^2 + 2\langle f'(x_k), r_k\rangle + \|r_k\|^2) \ge -\varepsilon\Delta\|f'(x_k)\|^2 - \widetilde{\sigma}_k(\|f'(x_k)\|^2 + 2\varepsilon\|f'(x_k)\|^2 + \varepsilon^2\|f'(x_k)\|^2) = -(\varepsilon\Delta + \widetilde{\sigma}_k(1 + 2\varepsilon + \varepsilon^2))\|f'(x_k)\|^2$. $\square$

Лемма 6. *Пусть* $x_k \notin X^*$*, вектор* r_k *удовлетворяет неравенству* (18)*, где* $0 \le \varepsilon < 1$*. Тогда* $-\widetilde{f}'(x_k)$ *— направление спуска для функции* $f(x)$ *в точке* x_k*.*

Доказательство утверждения следует из равенства $\langle f'(x_k), \widetilde{f}'(x_k)\rangle = \|f'(x_k)\|^2 + \langle f'(x_k), r_k\rangle$ и леммы 4.

За счет большой свободы в выборе векторов r_k, p_k, многообразий $M(x_k)$ и точек y_k, x_{k+1} предложенная схема допускает много различных реализаций. Среди них есть как известные методы, так и новые алгоритмы. Поскольку в схеме заложена возможность приближенного вычисления векторов $f'(x_k)$, то это позволяет исследовать устойчивость обсуждаемых ниже реализаций схемы к ошибкам вычисления градиента. Приведем сначала примеры исследования устойчивости известных одношаговых методов.

В [12] предложен и назван нелинейным один метод решения задачи (1) с выпуклой функцией $f(x)$, градиент которой удовлетворяет условию Липшица. На k−м шаге этого метода выбираются векторы d_1^k, d_2^k так, чтобы $\|d_i^k\| = 1$, $\langle f'(x_k), d_i^k\rangle \ge t_i^k\|f'(x_k)\|$, $t_i^k \in (0, 1]$, $i = 1, 2$, причем для коэффициентов t_i^k должно выполняться хотя бы одно из соотношений: $t_1^k \ge t_1 > 0$ или $t_2^k \ge t_2 > 0$. Затем полагается $x_{k+1} = x_k - \beta_1^k d_1^k - \beta_2^k d_2^k$, где числа β_i^k отыскиваются из условия релаксационности процесса. При выборе значений β_1^k, β_2^k в [12] используется константа Липшица, что не всегда удобно для практического применения метода. Покажем,что, во-первых, числа β_i^k можно выбирать без использования константы L, во-вторых, можно расширить возможности выбора чисел t_i^k и векторов d_i^k в нелинейном методе, в-третьих, его можно применять для минимизации псевдовыпуклых функций, и, в-четвертых, метод [12] устойчив к ошибкам вычисления градиента. Для этого приведем ниже один алгоритм общей схемы, который обобщает нелинейный метод, и докажем его сходимость.

Пусть построена точка $x_k \notin X^*$, вычислен вектор $\widetilde{f}'(x_k) = f'(x_k) + r_k$, и погрешность r_k удовлетворяет неравенству (18). Положим

$$\lambda_k = \lambda' = \lambda'' = 1, \quad p_k = \|\widetilde{f}'(x_k)\| \sum_{i \in H_k} d_i^k - \widetilde{f}'(x_k), \tag{30}$$

где H_k — конечное множество индексов, а векторы d_i^k выбраны так, что

$$\Big\| \sum_{i \in H_k} d_i^k \Big\| \le b < \infty, \quad \langle \widetilde{f}'(x_k), d_i^k \rangle \ge t_i^k \|\widetilde{f}'(x_k)\|, \quad i \in H_k, \tag{31}$$

$$0 < t \le \sum_{i \in H_k} t_i^k. \tag{32}$$

Положим далее $h_i^k = d_i^k$ $\forall i \in H_k$, множество $M(x_k)$ и точки y_k, x_{k+1} выберем согласно общей схеме. Алгоритм описан.

В силу (30)

$$g_k = \widetilde{f}'(x_k) + p_k = \|\widetilde{f}'(x_k)\| \sum_{i \in H_k} d_i^k,$$

т. е. вектор g_k является линейной комбинацией векторов h_i^k, $i \in H_k$, как это требуется в общей схеме. Покажем, что условие

$$g_k \ne 0, \tag{33}$$

заложенное в п. 1 схемы, также выполняется. Допустим, что $g_k = 0$. Тогда $\sum_{i \in H_k} d_i^k = 0$, т. к. согласно (18) $r_k \ne -f'(x_k)$ и, значит, $\|\widetilde{f}'(x_k)\| \ne 0$. Следовательно, $d_{i_0}^k = -\sum_{i \in H_k \setminus \{i_0\}} d_i^k$, где i_0 — произвольный индекс из H_k, и $\langle \widetilde{f}'(x_k), d_{i_0}^k \rangle = -\sum_{i \in H_k \setminus \{i_0\}} \langle \widetilde{f}'(x_k), d_i^k \rangle \ge t_{i_0}^k \|\widetilde{f}'(x_k)\|$. Кроме того, в силу (31) $\sum_{i \in H_k \setminus \{i_0\}} \langle \widetilde{f}'(x_k), d_i^k \rangle \ge \|\widetilde{f}'(x_k)\| \sum_{i \in H_k \setminus \{i_0\}} t_i^k$. Складывая два последних неравенства, получим $\|\widetilde{f}'(x_k)\| \sum_{i \in H_k} t_i^k \le 0$, что противоречит (32). Таким образом, условие (33) доказано, и требования к векторам g_k, h_i^k в данном алгоритме выполнены.

Докажем, что в случае ограниченности множеств X^0 или X_δ^0 для алгоритма (30)–(32) будут справедливы утверждения теорем 1, 2. Для этого проверим сначала, что при согласованном выборе значений ε и t в неравенствах (18), (32) векторы r_k, p_k удовлетворяют соответственно условиям (6), (7).

Пусть числа ε, t таковы,что

$$0 \le \varepsilon < 1, \quad 0 < t \le 1, \quad (b+1)\varepsilon + (1-t)(1+\varepsilon) < (1-\varepsilon)/2. \tag{34}$$

По лемме 4 для вектора r_k выполняется условие (6) при

$$\gamma_k = \gamma = (1+\varepsilon^2)/2. \tag{35}$$

Согласно (30), (32) $\langle \widetilde{f}'(x_k), p_k \rangle = \|\widetilde{f}'(x_k)\| \sum_{i \in H_k} \langle \widetilde{f}'(x_k), d_i^k \rangle - \|\widetilde{f}'(x_k)\|^2 \ge -\Big(1 - \sum_{i \in H_k} t_i^k\Big) \|\widetilde{f}'(x_k)\|^2 \ge -(1-t)\|\widetilde{f}'(x_k)\|^2$. Кроме того, $\|p_k\| \le (b+1)\|\widetilde{f}'(x_k)\| \le (b+1)(\|f'(x_k)\| + \|r_k\|) \le (b+1)(1+\varepsilon)\|f'(x_k)\|$, т. е. вектор p_k удовлетворяет неравенству (19) при $\Delta = b'(1+\varepsilon)$, где $b' = b+1$. Следовательно, по лемме 5 выполняется (7), где $\sigma_k = \sigma = (1+\varepsilon)(b'\varepsilon + (1-t)(1+\varepsilon))$, причем в силу (34), (35) $\sigma < \lambda'(1-\gamma) = (1-\gamma) = (1-\varepsilon^2)/2$. Итак, при условиях (34) требования (6), (7) к векторам r_k, p_k выполнены. Сразу отметим, что для выполнения неравенств (34) достаточно значения ε и t выбрать, например, из условий

$$1/2 \le t \le 1, \quad 0 \le \varepsilon < (2t-1)/(2b+5-2t)$$

или

$$0 \le \varepsilon < 1/(2b+3), \quad ((2b+5)\varepsilon + 1)/(2(1+\varepsilon)) < t \le 1.$$

Далее, поскольку $x_k \in X^0$ или $x_k \in X^0_\delta$ в зависимости от способа выбора точки y_k, то $\|f'(x_k)\| \le c_3 < \infty \ \forall k \in K$, а значит, для векторов r_k, p_k справедливы также неравенства (10). Таким образом, показано, что, если погрешность r_k удовлетворяет для всех $k \in K$ условию (18) и значения параметров ε и t согласованы, то по теоремам 1, 2 алгоритм (30)–(32) является сходящимся.

Заметим, что в алгоритме (30)–(32) некоторые из коэффициентов t_i^k могут быть неположительными, а часть векторов d_i^k — нулевыми. На каждой итерации алгоритма можно согласовывать значения $\varepsilon_k = \|r_k\|/\|f'(x_k)\|$ и $t_k = \sum\limits_{i\in H_k} t_i^k$ аналогично тому, как согласовывались выше значения ε и t. Кроме того, в алгоритме числа λ_k можно выбирать из общего условия (2). Тогда в (30) нужно положить $p_k = \|\widetilde{f}'(x_k)\| \sum\limits_{i\in H_k} d_i^k - \lambda_k \widetilde{f}'(x_k)$, а для гарантии выполнения неравенства (7) следует согласовывать значение ε не только со значением t, но и с числами λ', λ''. Так как векторы r_k, p_k удовлетворяют (6), (7), то по лемме 2 в многообразии $M(x_k)$ можно произвести переход к точке x_{k+1} с условием релаксации.

Вернемся к обсуждению описанного выше нелинейного метода из работы [12]. Очевидно, что алгоритм (30)–(32) обобщает нелинейный метод и имеет по сравнению с последним более широкие возможности для выбора на k-м шаге векторов d_i^k, чисел t_i^k и точек x_{k+1}. Следовательно, замечания, приведенные выше для метода [12] относительно возможности использования его для задачи (1) с псевдовыпуклой функцией, а также задания векторов d_1^k, d_2^k, чисел t_1^k, t_2^k и выбора шаговых множителей β_1^k, β_2^k, обоснованы. Например, для каждого $k \in K$ одно из значений t_1^k или t_2^k может быть неположительным, а один из векторов d_1^k или d_2^k может быть нулевым. Наконец, поскольку в алгоритме (30)–(32) заложена возможность неточного вычисления векторов $f'(x_k)$, то доказана устойчивость и нелинейного метода [12] по отношению к ошибкам вычисления градиентов.

Обоснуем далее устойчивость известного метода покоординатного спуска ([10], с. 206), считая, что для всех $k \in K$ вместо вектора $f'(x_k)$ вычисляется вектор

$$\widetilde{f}'(x_k) = \left(\frac{\partial \widetilde{f}(x_k)}{\partial \xi_1}, \dots, \frac{\partial \widetilde{f}(x_k)}{\partial \xi_n}\right),$$

и $\|f'(x_k) - \widetilde{f}'(x_k)\|$ удовлетворяет (18). На k-м шаге этого метода отыскивается номер

$$i_k = \arg\max_{1\le i\le n} \left|\frac{\partial \widetilde{f}(x_k)}{\partial \xi_i}\right|, \tag{36}$$

и x_{k+1} вычисляется как точка приближенного минимума функции $f(x)$ на прямой $x_k + \beta e_{i_k}$, $\beta \in R_1$. Для доказательства устойчивости метода покажем, что он является частным случаем алгоритма (30)–(32), устойчивость которого исследована выше.

Пусть в алгоритме (30)–(32) $H_k = \{1\}$, $x_{k+1} = y_k$, точка y_k находится из условия (4), и вектор $d_1^k = (\eta_1^k, \dots, \eta_n^k)$ выбран так, что

$$\eta_{i_k}^k = n\left(\frac{\partial \widetilde{f}(x_k)}{\partial \xi_{i_k}}\right)/\|\widetilde{f}'(x_k)\|, \quad \eta_i^k = 0 \ \ \forall i \ne i_k,$$

где индекс i_k определен согласно (36). Проверим, что вектор d_1^k удовлетворяет требованиям (31), (32) алгоритма. Действительно, $\|d_1^k\|^2{=}n^2\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_{i_k}}\right)^2/\|\widetilde{f}'(x_k)\|^2{\le}n^2\left(\sum\limits_{i=1}^n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2\right)/\|\widetilde{f}'(x_k)\|^2{=}n^2$, и, кроме того, поскольку $\|\widetilde{f}'(x_k)\|^2{=}\sum\limits_{i=1}^n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2{\le}n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_{i_k}}\right)^2$, то $\langle\widetilde{f}'(x_k),d_1^k\rangle{=}n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_{i_k}}\right)^2/\|\widetilde{f}'(x_k)\|\ge\|\widetilde{f}'(x_k)\|$. Значит, неравенства (31) выполняются при $b=n$, $t_1^k=t=1$. Как показано выше, при условиях (34) векторы r_k и $p_k=\|\widetilde{f}'(x_k)\|d_1^k-\widetilde{f}'(x_k)$ удовлетворяют (6), (7), а значит, по лемме 2 в многообразии $M(x_k)$ возможен релаксационный переход. Так как $M(x_k)$ — одномерное многообразие, задающееся вектором $h_1^k=d_1^k$, а x_{k+1} — точка приближенного минимума функции $f(x)$ на прямой $M(x_k)$, то приведенный алгоритм совпадает с методом покоординатного спуска. Для сходимости этого алгоритма, а значит, и метода покоординатного спуска с неточным вычислением градиентов необходимо, как и выше, согласовать задание числа ε в (18) со значением параметра t. Поскольку в алгоритме $t=1$, то неравенства (34) будут выполняться при $0\le\varepsilon<1/(2n+3)$. Таким образом, если ε в условии (18) выбрано указанным способом, то метод покоординатного спуска для минимизации псевдовыпуклых функций будет устойчив к ошибкам вычисления градиентов. Поскольку в методе циклического покоординатного спуска не требуется вычисления градиентов, то его устойчивость в указанном смысле здесь не обсуждается.

Идейно близким к методу покоординатного спуска является известный алгоритм градиентного спуска по быстрым переменным для минимизации выпуклых функций ([13], с. 179). На каждом шаге этого метода производится спуск не по всем переменным $\xi_1,\dots,\xi_n$ задачи, а лишь по так называемым быстрым переменным ξ_j, для которых

$$\left|\frac{\partial f(x_k)}{\partial\xi_j}\right|\ge\eta_k>0.$$

Исследуем устойчивость этого метода, считая, что функция $f(x)$ псевдовыпукла, а погрешности вычисления градиентов в итерационных точках удовлетворяют условию (18). Для этого покажем, что он вкладывается в схему (30)–(32).

Положим в алгоритме (30)–(32) $H_k{=}\left\{j:1{\le}j{\le}n,\left|\frac{\partial\widetilde{f}(x_k)}{\partial\xi_j}\right|{\ge}\eta_k{>}0\right\}$, $d_i^k{=}\left(n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)/\|\widetilde{f}'(x_k)\|\right)e_i$ $\forall i\in H_k$. Проверим, что выполняются условия (31), (32). Так как $\|d_i^k\|^2=n^2\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2/\|\widetilde{f}'(x_k)\|^2\le n^2\left(\sum\limits_{j=1}^n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_j}\right)^2\right)/\|\widetilde{f}'(x_k)\|^2=n^2$ $\forall i\in H_k$, то $\left\|\sum\limits_{i\in H_k}d_i^k\right\|\le\sum\limits_{i\in H_k}\|d_i^k\|\le\sum\limits_{i\in H_k}n=nl_k\le n^2$, где l_k — количество индексов в множестве H_k. Далее, пусть $t_i^k=n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2/\|\widetilde{f}'(x_k)\|^2$. Тогда $\langle\widetilde{f}'(x_k),d_i^k\rangle=n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2/\|\widetilde{f}'(x_k)\|=t_i^k\|\widetilde{f}'(x_k)\|$ $\forall i\in H_k$, причем

$$\sum_{i\in H_k}t_i^k=n\Bigg(\sum_{i\in H_k}\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2\Bigg)/\|\widetilde{f}'(x_k)\|^2=$$

$$=n\Bigg(\sum_{i\in H_k}\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2\Bigg)/\Bigg(\sum_{i=1}^n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2\Bigg)\ge n\frac{\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_{i_k}}\right)^2+\sum\limits_{i\in H_k\setminus\{i_k\}}\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_i}\right)^2}{n\left(\frac{\partial\widetilde{f}(x_k)}{\partial\xi_{i_k}}\right)^2}\ge1,$$

где индекс i_k определен в (36). Таким образом, векторы d_i^k удовлетворяют условиям (31), (32) при $b=n^2$, $t=1$. Если значения ε, t выбраны согласно (34), то выполняются (6), (7), и по лемме 2 в многообразии $M(x_k)$ возможен переход с условием (8). Так как $M(x_k)$ совпадает с множеством, в котором происходит спуск на k-м шаге метода [13], то приведенный алгоритм является методом градиентного спуска по быстрым переменным, и последний устойчив к ошибкам вычисления градиентов.

Нетрудно проверить, что частным случаем алгоритма (30)–(32) является градиентный метод при наличии помех. Для этого в алгоритме (30)–(32) достаточно положить $H_k = \{1\}$, $x_{k+1} = y_k$, $d_1^k = \widetilde{f}'(x_k)/\|\widetilde{f}'(x_k)\|$. Очевидно, что условия (31) выполняются при $b = 1$, $t_1^k = t = 1$. Значит, описанная методика применима для исследования устойчивости градиентного метода. Более того, покажем, что градиентный метод при наличии помех для псевдовыпуклых функций имеет ту же скорость сходимости, что и метод с точным вычислением градиента. Положим в общей схеме $\lambda_k = 1$, $p_k = 0$, $H_k = \{1\}$, $h_1^k = g_k$, $x_{k+1} = y_k$ $\forall k \in K$. Тогда x_{k+1} — точка приближенного минимума функции $f(x)$ на прямой $M(x_k)$, заданной вектором $g_k = \widetilde{f}'(x_k)$, и алгоритм представляет собой градиентный метод при наличии помех. Пусть погрешности r_k вычисления градиентов в этом методе удовлетворяют условию (18). Поскольку для вектора $p_k = 0$ выполняется (19), то оценки (23), (29) скорости сходимости градиентного метода при наличии помех остаются такими же, как и для метода с точным вычислением градиента (напр., [9], с. 266). Отметим, что в ([5], с. 98) при условии (18) исследована устойчивость градиентного метода с постоянным шагом для минимизации сильно выпуклых функций. В [3] обоснована его устойчивость в случае, когда $f(x)$ выпукла, шаговые множители β_k выбираются из условия $0 < \beta' \le \beta_k \le \beta'' < 2/L$, а погрешность r_k такова, что $\|r_k\| \le \varepsilon_k$, $\sum\limits_{k=1}^{\infty} \varepsilon_k < \infty$. В [4] для сильно выпуклой функции $f(x)$ исследовано влияние помех v_k на сходимость возмущенного процесса наискорейшего спуска $x_{k+1} = x_k \beta_k f'(x_k) + v_k$.

Обсудим устойчивость обобщенного метода градиентного спуска ([10], с. 189; [14], с. 61)

$$x_{k+1} = x_k - \lambda_k G_k \widetilde{f}'(x_k), \quad k \in K, \tag{37}$$

где G_k — симметричная матрица,

$$\|G_k\| \le c_4, \quad \langle G_k x, x\rangle \ge \widetilde{\mu}\|x\|^2, \ \ \widetilde{\mu} > 0, \ \ \forall x \in R_n, \tag{38}$$

а шаг $\lambda_k > 0$ выбирается близким к полному. Отметим, что в отличие от [10], [14] градиенты в (37) вычисляются с погрешностями. Покажем, что если погрешности r_k удовлетворяют (18), а значение ε согласовано с параметром $\widetilde{\mu}$, то метод (37) вкладывается в схему (30)–(32), а значит, является устойчивым в принятом смысле.

Пусть в алгоритме (30)–(32)

$$H_k = \{1\}, \quad x_{k+1} = y_k, \quad d_1^k = G_k \widetilde{f}'(x_k)/\|\widetilde{f}'(x_k)\|, \tag{39}$$

т. е. $p_k = G_k \widetilde{f}'(x_k) - \widetilde{f}'(x_k)$, $g_k = G_k \widetilde{f}'(x_k)$, $M(x_k) = U(x_k)$. Убедимся, что вектор d_1^k удовлетворяет условиям (31), (32). Действительно, первое из неравенств (31) выполняется согласно (38) при $b = c_4$. Кроме того, также в силу (38) $\langle \widetilde{f}'(x_k), d_1^k\rangle = \langle \widetilde{f}'(x_k), G_k \widetilde{f}'(x_k)\rangle/\|\widetilde{f}'(x_k)\| \ge \widetilde{\mu}\|\widetilde{f}'(x_k)\|$. Следовательно, второе из условий (31) выполняется при $t_1^k = \widetilde{\mu}$, а в неравенстве (32) можно положить $t = \mu' \le \min\{1, \widetilde{\mu}\}$. Таким образом, требования (31), (32) к вектору d_1^k выполнены. Тогда, как показано выше, при выборе ε согласно (34) векторы r_k, p_k удовлетворяют условиям (6), (7), (10), и по лемме 1 g_k — направление спуска для функции $f(x)$ в точке x_k. Следовательно, алгоритм (30)–(32), (39) имеет вид (37), и для обобщенного метода градиентного спуска выполняются утверждения теорем 1, 2. Если матрицы G_k в (39) таковы, что $\|G_k - E\| \le \Delta < 1 - \varepsilon$ $\forall k \in K$, то векторы p_k удовлетворяют (19). Тогда для алгоритма (30)–(32), (39), а значит, и для метода (37) при соответствующих предположениях теорем 3, 4 справедливы оценки скорости сходимости (23), (29).

В схему (37) вкладываются многие известные методы минимизации выпуклых функций, например, градиентный алгоритм ($G_k = E$), метод изменения масштабов ([10], с. 190), метод Ньютона с регулировкой шага ([14], с. 68), где $G_k = (f''(x_k))^{-1}$ удовлетворяет (38) при дополнительных условиях ([14], с. 27) на $f(x)$, вариант метода Ньютона, где $G_k = (f''(x_0))^{-1}$ $\forall k \in K$ ([14], с. 73), некоторые квазиградиентные алгоритмы (напр., [10], с. 191–193), в которых матрицы G_k удовлетворяют (38). Значит, если в упомянутых методах градиент вычисляется неточно, а параметры $\varepsilon, \widetilde{\mu}$ согласованы, то методы устойчивы в принятом смысле.

Положим в общей схеме

$$\lambda_k = 1, \quad p_k = G_k \widetilde{f}'(x_k), \quad H_k = \{1\}, \quad h_1^k = -g_k, \tag{40}$$

где матрицы G_k симметричны и выбраны согласно (38), а погрешности r_k вычисления градиента удовлетворяют (18), причем $0 \le \varepsilon < \min\{1, \widetilde{\mu}/c_4\}$. Тогда по лемме 4 для векторов r_k выполняется (6) при $\gamma_k = \gamma = (1+\varepsilon^2)/2 < 1$, и, кроме того, $\langle f'(x_k), p_k\rangle = \langle f'(x_k), G_k f'(x_k)\rangle + \langle f'(x_k), G_k r_k\rangle \ge \widetilde{\mu}\|f'(x_k)\|^2 - \|f'(x_k)\|\,\|G_k r_k\| \ge (\widetilde{\mu} - c_4\varepsilon)\|f'(x_k)\|^2 \ge 0$. Следовательно, по теоремам 1, 2 алгоритм (40) является сходящимся. Отметим, что если в условиях (38), (18) $c_4{<}1$, $0{\le}\varepsilon{<}(1-c_4)/(1+c_4)$, то $\|p_k\| = \|G_k \widetilde{f}'(x_k)\| \le c_4(1+\varepsilon)\|f'(x_k)\|$, $\ c_4(1+\varepsilon) < 1-\varepsilon$, а значит, векторы p_k удовлетворяют (19), и для алгоритма (40) справедливы оценки теорем 3, 4.

В ([4], с. 94) исследована следующая процедура решения задачи (1) с выпуклой функцией $f(x)$ и точным вычислением градиента. На k-м шаге ее выбирается такое подпространство V_k произвольной размерности l_k, что проекция $P_k f'(x_k)$ вектора $f'(x_k)$ на это подпространство удовлетворяет неравенству $\|P_k f'(x_k)\| \ge \tau\|f'(x_k)\|$, где $0 < \tau < 1$. Затем в качестве x_{k+1} выбирается точка приближенного минимума функции $f(x)$ на многообразии $x_k + V_k$.

Пусть в этой процедуре для всех $k \in K$ вектор $f'(x_k)$ вычисляется с погрешностью r_k, удовлетворяющей (18), условие выбора подпространства V_k имеет вид

$$\|P_k \widetilde{f}'(x_k)\| \ge \tau\|\widetilde{f}'(x_k)\|, \tag{41}$$

а точки $x_{k+1} \in x_k + V_k$ выбираются согласно (3) или (4), где $y_k = x_{k+1}$. Докажем, что тогда при согласованном задании параметров τ и ε процедура [4], примененная для минимизации псевдовыпуклой функции, будет устойчивой к ошибкам вычисления градиентов. Для этого достаточно показать, что она вкладывается в схему (30)–(32). Пусть в (30)–(32) $H_k = \{1, \dots, l_k\}$, где l_k — размерность подпространства V_k, выбираемого согласно (41). Положим $d_1^k = P_k \widetilde{f}'(x_k)/\|\widetilde{f}'(x_k)\|$, а векторы $d_2^k, \dots, d_{l_k}^k \in V_k$ зададим так, чтобы они вместе с d_1^k составляли базис подпространства V_k и, кроме того, удовлетворяли соотношениям $\|d_i^k\| = 1$, $\langle \widetilde{f}'(x_k), d_i^k\rangle \ge 0\ \forall i = 2, \dots, l_k$. Покажем, что для векторов d_i^k, $i \in H_k$, выполняются условия (31), (32). Так как $0 \in V_k$, то по известному свойству проекции $\langle P_k \widetilde{f}'(x_k), \widetilde{f}'(x_k) - P_k \widetilde{f}'(x_k)\rangle \ge 0$. Отсюда

$$\langle \widetilde{f}'(x_k), P_k \widetilde{f}'(x_k)\rangle \ge \|P_k \widetilde{f}'(x_k)\|^2, \tag{42}$$

а значит, $\|\widetilde{f}'(x_k)\| \ge \|P_k \widetilde{f}'(x_k)\|$. Тогда $\|d_1^k\| \le 1$, и первое из неравенств (31) выполняется при $b = n$. В силу (41), (42)

$$\langle \widetilde{f}'(x_k), d_1^k\rangle \ge \|P_k \widetilde{f}'(x_k)\|^2/\|\widetilde{f}'(x_k)\| \ge \tau^2\|\widetilde{f}'(x_k)\|.$$

Следовательно, второе из условий (31) справедливо при $t_1^k = \tau^2$, $t_2^k = \dots = t_{l_k}^k = 0$, а в (32) можно положить $t = \tau^2$. Таким образом, требования (31), (32) к векторам d_i^k выполняются, процедура из работы [4] действительно вкладывается в алгоритм (30)–(32), и ее можно использовать при минимизации псевдовыпуклой функции, вычисляя векторы $f'(x_k)$ приближенно. При этом согласно (34) значения ε, τ можно выбирать, например, из условий $0 \le \varepsilon < (2\tau^2 - 1)/(2(n-\tau^2)+5)$, $1/\sqrt{2} \le \tau < 1$.

В [15] предложен следующий метод решения задачи (1). На k-м шаге выбираются замкнутое множество Q_k такое, что $\|x\| \le c$, $c > 0$, $\forall x \in Q_k$, и точка z_k, содержащаяся в Q_k вместе со своей t-окрестностью ω_k. В качестве направления спуска в точке x_k задается вектор $s_k = \overline{z}_k - z_k$, где $\overline{z}_k \in Q_k$ и $\langle f'(x_k), \overline{z}_k - z_k\rangle = \min\limits_{x \in Q_k}\langle f'(x_k), x - z_k\rangle$. На луче $x_k + \alpha s_k$, $\alpha > 0$, выбирается точка $x_{k+1} = y_k$ согласно (3) или (4), где $g_k = s_k$.

Обоснуем устойчивость этого метода в принятом смысле, считая, что погрешность r_k вычисления градиента удовлетворяет (18). Для этого покажем, что метод вкладывается в схему (30)–(32). Положим в (30)–(32) $H_k = \{1\}$, $x_{k+1} = y_k$, $d_1^k = z_k - \overline{z}_k$, и проверим, что вектор d_1^k

удовлетворяет требованиям (31), (32). Очевидно, что первое из условий (31) выполняется при $b = 2c$. Далее, пусть

$$\widetilde{z}_k = z_k - \eta_k \widetilde{f}'(x_k)/\|\widetilde{f}'(x_k)\|,$$

где число $\eta_k > 0$ таково, что $\widetilde{z}_k \in Q_k - \overset{\circ}{\omega}_k$. Так как с учетом неточного вычисления градиента условие выбора направления спуска имеет вид $\langle \widetilde{f}'(x_k), s_k\rangle = \min_{x \in Q_k} \langle \widetilde{f}'(x_k), x - z_k\rangle$, то

$$\langle \widetilde{f}'(x_k), \overline{z}_k - z_k\rangle \le \langle \widetilde{f}'(x_k), \widetilde{z}_k - z_k\rangle = -\eta_k \|\widetilde{f}'(x_k)\|.$$

Следовательно, $\langle \widetilde{f}'(x_k), d_1^k\rangle \ge \eta_k \|\widetilde{f}'(x_k)\|$, и второе из условий (31) выполняется при $t_1^k = \eta_k$. Поскольку $\widetilde{z}_k \notin \overset{\circ}{\omega}_k$, то $\|z_k - \widetilde{z}_k\| = \eta_k \ge t$, а значит, $t_1^k \ge t > 0$, и условие (32) также выполняется. Таким образом, для последовательности $\{x_k\}$, построенной описанным методом с неточным вычислением градиента, при выборе величин t, ε согласно (34) справедливы теоремы сходимости 1, 2.

Перейдем к исследованию устойчивости еще одного класса методов решения задачи (1). Для этого сначала обсудим следующий алгоритм общей схемы. Пусть для всех $k \in K$ в схеме полагается

$$H_k = \{1, \dots, l_k\}, \ \ 1 \le l_k \le n, \ \ \lambda_k = 1, \ \ p_k = 0, \ \ x_{k+1} = y_k, \ \ h_1^k = -\widetilde{f}'(x_k), \tag{43}$$

а векторы $h_2^k, \dots, h_{l_k}^k$ выбираются произвольно. Поскольку l_k-мерное многообразие $M(x_k)$ содержит вектор $x_k + g_k$, а p_k удовлетворяет условиям (7), (10), (19), то для алгоритма (43) справедливы утверждения теорем 1, 2 и оценки теорем 3–5, если для погрешностей r_k выполняются требования (6) или (18), а множества X^0 или X_δ^0 ограничены. За счет большой свободы в выборе чисел l_k и векторов h_i^k алгоритм (43) допускает много реализаций.

Пусть, например, в алгоритме (43) $l_k = 2$, а векторы h_2^k, $k \in K$, таковы, что

$$h_2^0 = 0, \ \ h_2^k = x_k - x_{k-1} \ \ \forall k = 1, 2, \dots \tag{44}$$

Тогда при $k \ge 1$ переход от точки x_k к точке x_{k+1} происходит в двумерном линейном многообразии $M(x_k)$, заданном векторами $\widetilde{f}'(x_k)$ и $x_k - x_{k-1}$, и алгоритм представляет собой известный вариант метода сопряженных градиентов (напр., [5], с. 70). При этом в отличие от [5] градиент в алгоритме может вычисляться неточно и для выбора $x_{k+1} \in M(x_k)$ имеются более широкие возможности (см. (3), (4)). Как отмечено выше, алгоритм (43) устойчив к ошибкам вычисления градиента. Следовательно, упомянутый вариант метода сопряженных градиентов для минимизации псевдовыпуклых функций также является устойчивым в принятом смысле.

Пусть в (43) $l_k = l$, $2 \le l \le n$, y_k — точка приближенного минимума $f(x)$ на $M(x_k)$, и для всех $i = 2, \dots, l$ полагается $h_i^k = 0$, если $0 \le k \le l - 2$, и $h_i^k = x_{k-i+2} - x_{k-i+1}$, если $k \ge l - 1$, т. е. на начальных шагах спуск производится по антиградиенту, а затем — в l-мерных многообразиях, задаваемых векторами $\widetilde{f}'(x_k)$ и $x_{k-i+2} - x_{k-i+1}$, $i = 2, \dots, l$. Такой алгоритм для выпуклых функций с точным вычислением градиента описан в ([16], с. 116) и назван методом многопараметрического поиска. Так как алгоритм (43) устойчив к ошибкам вычисления векторов $f'(x_k)$, то устойчивым является и метод многопараметрического поиска.

Поскольку в (43) векторы h_i^k, $i = 2, \dots, l_k$, выбираются произвольно, то в алгоритм (43), (44) и метод многопараметрического поиска можно включать различные процедуры обновления, полагая на некоторых итерациях $h_i^k = 0$, $i = 2, \dots, l_k$, т. е. делая переход, как и в начальной точке, по антиградиенту. При этом сходимость алгоритмов с любыми процедурами обновления будет сохранена, и методы останутся устойчивыми. Полезность включения процедур обновления в таких алгоритмах обсуждается, например, в ([2], с. 326; [5], с. 73).

Отметим, что один из подходов к исследованию устойчивости традиционного метода сопряженных градиентов к ошибкам вычисления производных предложен в [17].

Методы сопряженных градиентов и многопараметрического поиска относятся к классу так называемых многошаговых алгоритмов, т. к. в них при построении очередного приближения используется информация с предыдущих итераций. Обычно (напр., [5], с. 68) метод называют

N-шаговым, $N \geq 1$, если в нем итерационные точки x_k для $k \geq N$ строятся с использованием N предыдущих точек $x_{k-1}, \ldots, x_{k-N}$. По-видимому, к N-шаговым следует отнести и те алгоритмы, в которых при построении k-го приближения используется информация только с $(k-1)$-го и $(k-N)$-го шагов, а не со всех N предыдущих итераций. Проведем ниже исследование устойчивости еще нескольких алгоритмов из класса многошаговых методов.

Положим в общей схеме

$$H_k = \{1\}, \quad h_1^k = g_k, \quad x_{k+1} = y_k, \quad p_k = \alpha_k(\widetilde{f}'(x_k) - \widetilde{f}'(x_{k-1})) \quad \forall k \in K, \tag{45}$$

где числа α_k таковы, что $\alpha_0 = 0$, $0 \leq |\alpha_k| \leq \Delta' \|\widetilde{f}'(x_k)\|/b_k$, $k = 1, 2, \ldots$, $b_k \geq \|\widetilde{f}'(x_k) - \widetilde{f}'(x_{k-1})\|$, а значение Δ' определено в (20). Будем считать, что r_k удовлетворяет (18). Поскольку $\|p_k\| \leq \Delta' \|\widetilde{f}'(x_k)\| \, \|\widetilde{f}'(x_k) - \widetilde{f}'(x_{k-1})\|/b_k \leq \Delta' \|\widetilde{f}'(x_k)\|$, то для векторов p_k выполняется неравенство (20), а значит, и условие (19) при $\Delta = \Delta'(1+\varepsilon)$. Следовательно, для этого алгоритма справедливы теоремы 3, 4.

Двухшаговый алгоритм (45) можно обобщить, положив

$$p_k = \alpha_k^1(\widetilde{f}'(x_k) - \widetilde{f}'(x_{k-1})) + \cdots + \alpha_k^{N-1}(\widetilde{f}'(x_{k-N+2}) - \widetilde{f}'(x_{k-N+1})),$$

$N \geq 2$, и выбрав значения α_k^i следующим образом:

$$\alpha_k^i = 0 \;\; \forall k = 0, \ldots, N-2, \quad 0 \leq |\alpha_k^i| \leq \Delta' \|\widetilde{f}'(x_k)\|/(b_k^i(N-1)) \;\; \forall k \geq N-1,$$
$$b_k^i \geq \|\widetilde{f}'(x_{k-i+1}) - \widetilde{f}'(x_{k-i})\|, \quad i = 1, \ldots, N-1.$$

Тогда, как и выше, при условии (18) легко показать, что векторы p_k удовлетворяют неравенству (19), т. е. для N-шагового варианта также справедливы теоремы 3, 4. Отметим, что этот вариант близок к многошаговому алгоритму [18] с точным вычислением градиентов, где числа λ_k, α_k^i выбираются с использованием константы Липшица.

Пусть в (45) $p_k = \alpha_k(x_k - z_k)$, где α_k таково, что $0 \leq |\alpha_k| \leq \Delta' \|\widetilde{f}'(x_k)\|/\|x_k - z_k\|$, а z_k — любая точка, отличная от x_k. Очевидно, что p_k удовлетворяет неравенству (20). Следовательно, при условии (18) выполняется (19), и для этого алгоритма справедливы теоремы 3, 4. Заметим, что при $k \geq 1$ в качестве z_k можно выбирать любую из предыдущих итерационных точек от x_0 до x_{k-1}. Тем самым будут получаться различные многошаговые реализации, устойчивые к ошибкам вычисления градиентов. Если $z_k = x_{k-1}$ и $\alpha_k \leq 0$, то алгоритм близок к методу тяжелого шарика ([5], с. 68), а если к тому же $\lambda_k = 1$, то — к традиционному методу сопряженных градиентов (напр., [5], с. 73; [9], с. 320). Описанный алгоритм можно обобщить, положив в (45)

$$p_k = \alpha_k^1(x_k - z_k^1) + \cdots + \alpha_k^{N-1}(x_{k-N+2} - z_k^{N-1}),$$

где $N \geq 2$, $\alpha_k^i = 0 \;\forall k = 0, \ldots, N-2$, $0 \leq |\alpha_k^i| \leq \Delta' \|\widetilde{f}'(x_k)\|/(\|x_{k-i+1} - z_k^i\|(N-1)) \;\forall k \geq N-1$, $i = 1, \ldots, N-1$. При условии (18) для этого варианта также справедливы утверждения теорем 3, 4, поскольку векторы p_k удовлетворяют (19). Нетрудно видеть, что при построении p_k в (45) можно одновременно использовать предыдущие итерационные точки и градиенты функции в этих точках, например, $p_k = \alpha_k^1(x_k - x_{k-1}) + \alpha_k^2(\widetilde{f}'(x_k) - \widetilde{f}'(x_{k-1}))$. При этом значения α_k^1, α_k^2, как и выше, нужно выбрать так, чтобы векторы p_k удовлетворяли условию (20).

Остановимся подробнее на двухшаговом алгоритме схемы, который отличается от упомянутого метода сопряженных градиентов способом задания коэффициентов при векторах $x_k - x_{k-1}$. Пусть в общей схеме $H_k = \{1\}$, $h_1^k = g_k$, $\lambda_k = 1$, $k \in K$, $p_0 = 0$, $p_k = \alpha_k(x_k - x_{k-1})$, $k = 1, 2, \ldots$, $0 \leq |\alpha_k| \leq \overline{\alpha}$. Если при этом положить $x_{k+1} = y_k = \arg \min\limits_{x \in U(x_k)} f(x)$, то алгоритм можно записать в виде

$$x_{k+1} = x_k - \beta_k g_k, \quad \beta_k = \arg\min_{\beta} f(x_k - \beta g_k), \quad k \in K, \tag{46}$$

где $g_0 = \widetilde{f}'(x_k)$, $g_k = \widetilde{f}'(x_k) + \alpha_k(x_k - x_{k-1})$, $k \geq 1$. Будем по-прежнему считать, что погрешности r_k вычисления градиентов удовлетворяют условию (18). Тогда по лемме 4 $\langle f'(x_k), r_k \rangle \geq -(1+\varepsilon^2)\|f'(x_k)\|^2/2$. Отсюда с учетом равенства $\langle f'(x_k), x_k - x_{k-1} \rangle = 0$, $k = 1, 2, \ldots$, следует, что $\langle f'(x_k), g_k\rangle = \langle f'(x_k), \widetilde{f}'(x_k)\rangle = \|f'(x_k)\|^2 + \langle f'(x_k), r_k\rangle \ge (1-\varepsilon^2)\|f'(x_k)\|^2/2 > 0$, и вектор $-g_k$ для всех $k \in K$, независимо от знака числа α_k, является направлением спуска для функции $f(x)$ в точке x_k. Значит, описанный процесс релаксационен, $\beta_k > 0$, и $x_k \in X^0$ $\forall k \in K$. Если $\operatorname{diam} X^0 = d_0 < \infty$, то последовательность $\{\|f'(x_k)\|\}$ ограничена, $\|p_k\| \le \overline{\alpha} d_0$, $k \in K$, и условия (10) для векторов r_k, p_k выполняются. Кроме того, как уже отмечено, для r_k справедливо неравенство (6), а поскольку $\langle f'(x_k), p_k\rangle = 0$, то выполняется и условие (7). Таким образом, для алгоритма (46) остаются в силе теоремы сходимости 1, 2, и устойчивость такого варианта метода сопряженных градиентов обоснована. Отметим, что в (46) за счет выбора значений α_k имеются более широкие возможности для построения направлений спуска по сравнению с традиционным алгоритмом. Кроме того, в (46) приближение $x_{k+1} \in U(x_k)$ не обязательно искать из условия точного минимума функции $f(x)$. Достаточно, чтобы точка $x_{k+1} = y_k$ удовлетворяла (4) и при этом

$$\alpha_{k+1}\langle f'(x_{k+1}), x_{k+1} - x_k\rangle \ge 0. \tag{47}$$

Тогда $\langle f'(x_k), g_k\rangle > 0$, $\langle f'(x_k), p_k\rangle \ge 0$, $k \in K$, и для векторов r_k, p_k справедливы неравенства (10), (6), (7), выполнение которых требуется в теоремах 1, 2. Для обеспечения условия (47) достаточно согласовывать знаки числа α_{k+1} и произведения $\langle f'(x_{k+1}), x_{k+1} - x_k\rangle$. Если к k-му шагу α_{k+1} уже задано, то для выполнения (47) шаг β_k в (46) следует выбирать большим или меньшим полного в зависимости от знака числа α_{k+1}. Заметим также, что во всех обсужденных вариантах алгоритмов (45) и (46) допускается задание $\alpha_k = 0$ ($\alpha_k^i = 0$). Тем самым в алгоритмах предусматривается включение всевозможных процедур обновления без ущерба их сходимости и устойчивости.

К многошаговым можно отнести так называемый ускоренный метод градиентного спуска порядка N ([19], с. 99), где $N \in K$, $N \ge 2$, который применяется при минимизации “овражных” функций и заключается в следующем. Пусть построена точка x_k, $k \in K$. Из нее осуществляется N шагов методом наискорейшего спуска с точной одномерной минимизацией. При этом получаются последовательно вспомогательные точки $y_k^1, \ldots, y_k^N$, а затем полагается

$$x_{k+1} = x_k + \beta_k s_k, \tag{48}$$

где $s_k = y_k^N - x_k$, β_k — полный шаг. Отметим, что если включить вспомогательные точки y_k^i, $i = 1, \ldots, N$, в число итерационных точек последовательности $\{x_k\}$, т. е. положить

$$x_{k+1} = \begin{cases} x_k - \beta_k f'(x_k), & k \in K \setminus K^0; \\ x_{k-N} + \beta_k(x_k - x_{k-N}), & k \in K^0, \end{cases}$$

где $K^0 = \{k \in K : k = l(N+1) - 1,\ l = 1, 2, \ldots\}$, то метод можно считать $(N{+}1)$-шаговым. В ([16], с. 113) алгоритм (48) описан при $N = n$.

Покажем, что метод (48), как и градиентный, является устойчивым в принятом смысле, а при построении точек y_k^i и x_{k+1} можно отказаться от условия полного шага. Пусть в (48) градиенты в основных и вспомогательных точках вычисляются приближенно. Будем считать, что погрешности r_k вычисления градиентов в точках x_k удовлетворяют неравенству (18), а погрешности r_k^i в точках y_k^i — неравенству $\|r_k^i\| \le \varepsilon\|f'(y_k^i)\|$. Покажем, что (48) вкладывается в общую схему. Пусть в схеме $\lambda_k = 1$, $p_k = 0$, $H_k = \{1\}$, $h_1^k = \widetilde{f}'(x_k)$, а точки y_k выбираются из условия (4). Так как векторы $-\widetilde{f}'(x_k)$ и $-\widetilde{f}'(y_k^i)$ являются направлениями спуска для функции $f(x)$ в точках x_k и y_k^i соответственно, то можно считать, что $y_k = y_k^1$, и для точки x_{k+1}, полученной согласно (48), справедливы соотношения

$$f(x_{k+1}) < f(y_k^N) < \cdots < f(y_k^1) = f(y_k) < f(x_k). \tag{49}$$

Поскольку условия (6), (7), (10), (18), (19) выполняются, то метод (48) с помехами является сходящимся, а оценки его скорости сходимости остаются не хуже оценок метода без помех. Отметим, что для построения точек y_k^i, $i = 2, \ldots, N$, можно применять любые релаксационные процедуры, комбинируя их между собой.

Идея, заложенная в методе (48), может быть использована и для других исследованных выше алгоритмов с неточным вычислением градиентов. Например, для метода сопряженных градиентов она реализуется следующим образом. Из точки x_k производим N шагов указанным методом, получая последовательно точки $y_k^1, \dots, y_k^N$, а затем строим точку x_{k+1} согласно (48). Поскольку $y_k^1 = \arg\min_{\beta \geq 0} f(x_k - \beta \widetilde{f}'(x_k))$, то, положив, как и выше, $y_k = y_k^1$ в общей схеме, получим (49). Значит, такой вариант метода сопряженных градиентов вкладывается в общую схему и является сходящимся при соответствующих требованиях на r_k. Таким образом, для многих алгоритмов безусловной минимизации могут быть построены аналогично (48) многошаговые модификации.

Предложим далее для решения задачи (1) такую процедуру общей схемы, в которой для построения точки x_{k+1} может использоваться вместе с x_k любая предыдущая итерационная точка. Пусть k-е приближение найдено. Выполняем пп. 1–3 общей схемы, считая $H_k = \{1\}$. Кроме $y_k \in U(x_k)$ выбираем еще две вспомогательные точки $y_k', z_k \in R_n$ так, чтобы $f(y_k') \leq f(y_k)$, $f(z_k) \geq f(y_k')$. Затем полагаем $x_{k+1} = z_k + \beta_k(y_k' - z_k)$, где число $\beta_k > 0$ таково, что $f(x_{k+1}) \leq f(y_k')$. Заметим, что шаг β_k, обеспечивающий выполнение последнего неравенства, существует, например, $\beta_k = 1$. Поскольку справедливо (5), то описанная процедура является реализацией общей схемы, и при соответствующих условиях выбора векторов p_k, r_k для нее будут выполняться утверждения теорем 1–5. Вспомогательные точки y_k', z_k могут быть выбраны, в частности, следующим образом: $y_k' = y_k$, $z_k = y_k'$, $z_k = y_k$, $z_k = x_k$. В качестве z_k можно взять любую из точек $x_0, \dots, x_{k-1}$, получив тем самым многошаговый алгоритм. Предложенная процедура может быть полезна для минимизации "овражных" функций, если на каждом шаге точки y_k и y_k' выбирать независимо друг от друга на "дне оврага", а затем полагать $z_k = y_k$. Такая реализация близка к эвристическому алгоритму, описанному в ([9], с. 269).

Литература

1. Бабич М.Д., Иванов В.В. *Исследование полной погрешности в задачах минимизации функционалов при наличии ограничений* // Укр. матем. журн. – 1969. – Т. 21. – № 1. – С. 3–14.
2. Васильев Ф.П. *Методы решения экстремальных задач. Задачи минимизации в функциональных пространствах, регуляризация, аппроксимация.* Учеб. пособие. – М.: Наука, 1981. – 400 с.
3. Гольштейн Е.Г., Третьяков Н.В. *Градиентный метод минимизации и алгоритмы выпуклого программирования, связанные с модифицированными функциями Лагранжа* // Экономика и матем. методы. – 1975. – Т. 11. – Вып. 4. – С. 730–742.
4. Любич Ю.И., Майстровский Г.Д. *Общая теория релаксационных процессов для выпуклых функционалов* // УМН. – 1970. – Т. 25. – Вып. 1. – С. 57–112.
5. Поляк Б.Т. *Введение в оптимизацию.* – М.: Наука, 1983. – 384 с.
6. Базара М., Шетти К. *Нелинейное программирование. Теория и алгоритмы.* – М.: Мир, 1982. – 583 с.
7. Бахвалов Н.С. *Численные методы. Анализ, алгебра, обыкновенные дифференциальные уравнения.* Учеб. пособие. – М.: Наука, 1973. – 631 с.
8. Заботин И.Я. *Об одном подходе к построению алгоритмов безусловной минимизации псевдовыпуклых функций* // Изв. вузов. Математика. – 1998. – № 12. – С. 29–39.
9. Васильев Ф.П. *Численные методы решения экстремальных задач.* – М.: Наука, 1988. – 552 с.
10. Карманов В.Г. *Математическое программирование.* Учеб. пособие. – 3-е изд. – М.: Наука, 1986. – 288 с.
11. Кораблев А.И. *О релаксационных методах минимизации псевдовыпуклых функций* // Исслед. по прикл. матем. – Казань, 1980. – № 8. – С. 3–8.
12. Третьяков А.А. *Две схемы нелинейного метода оптимизации в экстремальных задачах* // Журн. вычисл. матем. и матем. физ. – 1984. – Т. 24. – № 7. – С. 986–992.
13. Моисеев Н.Н. *Численные методы в теории оптимальных систем.* – М.: Наука, 1971. – 424 с.

14. Пшеничный Б.Н., Данилин Ю.М. *Численные методы в экстремальных задачах.* – М.: Наука, 1975. – 320 с.
15. Заботин И.Я. *К вопросу о выборе направлений спуска в задачах безусловной минимизации функций* // Исслед. по прикл. матем. – Казань, 1981. – № 9. – С. 37–42.
16. Химмельблау Д. *Прикладное нелинейное программирование.* – М.: Мир, 1975. – 536 с.
17. Завриев С.К. *Об устойчивости вычислительной схемы метода сопряженных градиентов* // Вопр. кибернетики. Анализ больших систем. – М., 1992. – С. 102–118.
18. Хотеев С.В. *О многошаговых градиентных методах в задачах оптимизации* // Вопр. кибернетики. Модели и методы анализа больших систем. – М., 1991. – С. 104–111.
19. Мину М. *Математическое программирование. Теория и алгоритмы.* – М.: Наука, 1990. – 488 с.

Казанский государственный университет

Поступила
20.06.2000