Вспомогательные утверждения

Рассмотрим задачу решения гладкой системы нелинейных уравнений $F : \mathbb{R}^n \to \mathbb{R}^m$:

$$F(x) = \mathbf{0}_m, \quad \mathbf{0}_m^\top = (0, \ldots, 0). \tag{1}$$

Решение системы рассматривается в ключе задачи безусловной оптимизации невязки на основе евклидовой нормы:

$$\min_{x \in \mathbb{R}^n} \left\{ f_1(x) \stackrel{def}{=} \|F(x)\| \right\}. \tag{2}$$

Предположение 1. Пусть $F(x)$ — векторно-значное отображение, удовлетворяющее условию Липшица на $\mathcal{F}$:

$$\exists L_F > 0 : \left\| F^{'}(y) - F^{'}(x) \right\| \leq L_F \|y - x\|, \quad \forall (x, y) \in \mathcal{F}^2.$$

Предположение 2. Пусть для многозначного отображения выполнено неравенство Поляка–Лоясиевича [Поляк, 1963], то есть существует такой $\mu > 0$, для которого при любых $x \in \mathcal{F}$ выполнено неравенство:

$$\sigma_{\min}(F^{'}(x)^\top) \geq \sqrt{\mu}. \tag{3}$$

Алгоритм 1 Ускоренный «метод трех квадратов» с неточным проксимальным отображением

Input:

$$\begin{cases} y_0 = x_0 \in \mathbb{R}^n, \mathcal{L}(f_1(x_0)) \subseteq \mathcal{F} \text{ — начальное приближение}, x_{-1} = x_0; \\ \mathcal{E}(\cdot) \text{ — функция погрешности вычисления } y_{k+1}; \\ N \in \mathbb{N} \text{ — количество итераций метода}; \\ L \text{ — оценка локальной постоянной Липшица}, L > 0, L_0 = L; \\ \mathcal{T}(\cdot) \text{ — функция, определяющая значение } \tau. \end{cases}$$

for $k = 0, 1, \ldots, N - 1$ do

1. определить $\tau_k = \mathcal{T}(x_k, L_k, \varepsilon_k)$, $\varepsilon_k = \mathcal{E}(k, x_k, x_{k-1})$;
2. вычислить такой $y_{k+1} \in \mathbb{R}^n$, что
   $\psi_{x_k, L_k, \tau_k}(y_{k+1}) - \psi_{x_k, L_k, \tau_k}(T_{L_k, \tau_k}(x_k)) \leq \varepsilon_k$ и
   $f_1(x_k) - \psi_{x_k, L_k, \tau_k}(y_{k+1}) \geq 0$;
3. если $f_1(y_{k+1}) > \psi_{x_k, L_k, \tau_k}(y_{k+1})$, то положить
   $L_k := 2L_k$ и вернуться к пункту 1;
4. $L_{k+1} = \max\left\{ \frac{L_k}{2}, L \right\}$;
5. определить $\varphi_k(t) = f_1(y_{k+1} + t(y_{k+1} - y_k))$;
6. найти такой $t_k \geq 0$, что $\varphi_k(t_k) \leq \varphi_k(0)$;
7. положить $x_{k+1} = y_{k+1} + t_k(y_{k+1} - y_k)$.

end for

Output: x_N.

В данной работе рассматривается более общая форма правила обновления x_{k+1}, использующая явное выражение $T_{L_k,\tau_k}(x_k)$:

$$y_{k+1} = y(\eta_k) \overset{def}{=} x_k - \eta_k \left(F^{'}(x_k)^{\top} F^{'}(x_k) + \tau_k L_k I_n\right)^{-1} F^{'}(x_k)^{\top} F(x_k), \quad \eta_k \geq 0. \tag{4}$$

В лемме 1 доказывается формула мажоранты оптимизируемой функции в задаче (2).

Лемма 1 ([Nesterov, 2020]). Пусть $(x, y) \in \mathcal{F}^2, L \geq L_F, \tau > 0$ и выполнено предположение 1. Тогда $f_1(y) \leq \psi_{x,L,\tau}(y)$.

Доказательство. Выведем неравенство для $\left\|F(y) - F(x) - F^{'}(x)(y-x)\right\|$:

$$\begin{aligned}
&\left\|F(y) - F(x) - F^{'}(x)(y-x)\right\| = \\
&= \left\{F(y) = F(x) + \int_0^1 F^{'}(x + t(y-x))(y-x)\,\mathrm{d}\,t\right\} = \\
&= \left\|\int_0^1 \left(F^{'}(x + t(y-x)) - F^{'}(x)\right)(y-x)\,\mathrm{d}\,t\right\| \leq \\
&\leq \left\{\|\cdot\| \text{ — выпукла, неравенство Йенсена}\right\} \leq \\
&\leq \int_0^1 \left\|\left(F^{'}(x + t(y-x)) - F^{'}(x)\right)(y-x)\right\| \mathrm{d}\,t \leq \\
&\leq \int_0^1 \left\|F^{'}(x + t(y-x)) - F^{'}(x)\right\| \|y-x\|\,\mathrm{d}\,t \leq \\
&\leq \{\text{предположение 1}\} \leq \int_0^1 L_F\|y-x\|^2 t\,\mathrm{d}\,t = \frac{L_F}{2}\|y-x\|^2.
\end{aligned} \tag{5}$$

Рассмотрим неравенство между арифметическим и геометрическим средним для двух величин a и b:

$$\begin{aligned}
&a \overset{def}{=} \tau; \\
&b \overset{def}{=} \frac{1}{\tau}\left\|F(x) + F^{'}(x)(y-x)\right\|^2; \\
&\frac{a+b}{2} \geq \sqrt{ab} \Rightarrow \frac{\tau}{2} + \frac{1}{2\tau}\left\|F(x) + F^{'}(x)(y-x)\right\|^2 \geq \left\|F(x) + F^{'}(x)(y-x)\right\|.
\end{aligned} \tag{6}$$

Тогда для f_1 выполнено

$$\begin{aligned}
f_1(y) = \|F(y)\| &= \left\|F(y) - F(x) - F^{'}(x)(y-x) + F(x) + F^{'}(x)(y-x)\right\| \leq \\
&\leq \left\|F(y) - F(x) - F^{'}(x)(y-x)\right\| + \\
&+ \left\|F(x) + F^{'}(x)(y-x)\right\| \leq \{\text{неравенство из (5)}\} \leq \frac{L_F}{2}\|y-x\|^2 + \\
&+ \left\|F(x) + F^{'}(x)(y-x)\right\| \leq \\
&\leq \{\text{неравенство из (6)}\} \leq \frac{\tau}{2} + \frac{L_F}{2}\|y-x\|^2 + \\
&+ \frac{1}{2\tau}\left\|F(x) + F^{'}(x)(y-x)\right\|^2 = \psi_{x,L_F,\tau}(y) \leq \psi_{x,L,\tau}(y).
\end{aligned}$$

□

Лемма 2 устанавливает важное матричное отношение частичного порядка для вывода линейной сходимости в условии Поляка–Лоясиевича.

Лемма 2 ([Nesterov, 2020]). Пусть линейный оператор $A : \mathbb{R}^n \to \mathbb{R}^m$, с $m \le n$ обладает матрицей, не вырожденной по строкам:

$$AA^\top \succeq \mu I_m$$

для определенного $\mu > 0$. Тогда для любого $\xi > 0$ выполнено

$$A\left(\xi I_n + A^\top A\right)^{-t} A^\top \succeq \frac{\mu}{(\xi+\mu)^t} I_m, \quad t \in [0,1].$$

Отношение порядка «$\succeq$» выполнено на конусе неотрицательно определенных матриц.

Доказательство. Рассмотрим сингулярное разложение матрицы оператора A:

$$A = U\Lambda V^\top, \quad U^\top U = I_m, \quad V^\top V = I_m,$$

где Λ — диагональная матрица, $\Lambda \succeq \sqrt{\mu} I_m$ (по условию). Введем матрицу W со столбцами, ортогонально дополняющими столбцы из V до полного базиса в $\mathbb{R}^n$:

$$VV^\top + WW^\top = I_n, \quad W^\top V = \mathbf{0}_{(n-m)\times m}.$$

Пользуясь блочной структурой из-за $W^\top V = \mathbf{0}_{(n-m)\times m}$, получаем:

$$\begin{aligned}
A\left(\xi I_n + A^\top A\right)^{-t} A^\top &= U\Lambda V^\top \left(\xi\left(VV^\top + WW^\top\right) + V\Lambda^2 V^\top\right)^{-t} V\Lambda U^\top = \\
&= U\Lambda V^\top \left(V\left(\xi I_m + \Lambda^2\right) V^\top + \xi WW^\top\right)^{-t} V\Lambda U^\top = \\
&= U\Lambda V^\top \left(V\left(\xi I_m + \Lambda^2\right)^{-t} V^\top + \frac{1}{\xi^t} WW^\top\right) V\Lambda U^\top = \\
&= U\Lambda\left(\xi I_m + \Lambda^2\right)^{-t} \Lambda U^\top = \\
&= U\left(\xi \Lambda^{-\frac{2}{t}} + \Lambda^{2-\frac{2}{t}}\right)^{-t} U^\top \succeq \frac{1}{\left(\xi \mu^{-\frac{1}{t}} + \mu^{1-\frac{1}{t}}\right)^t} I_m = \\
&= \frac{\mu}{(\xi+\mu)^t} I_m.
\end{aligned}$$

□

Основные утверждения

Теорема 1 использует явную формулу проксимального отображения для вывода оценки на убывание оптимизируемой функции в процессе, организованном по схеме 1.

Теорема 1. Пусть выполнены предположения 1 и 2. Рассмотрим последовательность $\{x_k\}_{k\in\mathbb{Z}_+}$, вычисляемую по схеме 1 с правилом (4), в котором $\tau_k > 0$, $\eta_k \in (0,2)$, $k \in \mathbb{Z}_+$. Тогда для $k \in \mathbb{Z}_+$:

$$f_1(x_{k+1}) \le \frac{\tau_k}{2} + \begin{cases} \frac{f_2(x_k)}{2\tau_k}\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k\tau_k+\mu}\right), \tau_k \ge \frac{\mu}{L_k}; \\ \frac{f_2(x_k)(1-\eta_k)^2}{2\tau_k} + \frac{\eta_k(2-\eta_k)L_k f_2(x_k)}{2\mu} - \frac{\eta_k(2-\eta_k)L_k^2 f_2(x_k)\tau_k}{2\mu^2(1+\xi)^3}, \text{ для} \\ \text{некоторого } \xi \in (-1,1] \text{ при } \tau_k < \frac{\mu}{L_k}. \end{cases} \tag{7}$$

При этом для $\eta_k = 1, k \in \mathbb{Z}_+$ верно:

$$f_1(x_{k+1}) \le \frac{\tau_k}{2} + \frac{L_F}{\mu} f_2(x_k). \tag{8}$$

Доказательство. По определению $\psi_{x,L,\tau}(y)$ (лемма 1):

$$\begin{aligned} f_1(x_k) - f_1(x_{k+1}) \ge f_1(x_k) - f_1(y_{k+1}) \ge f_1(x_k) - \psi_{x_k,L_k,\tau_k}(x_{k+1}) = f_1(x_k) - \\ - \frac{\tau_k}{2} - \frac{L_k}{2}\|x_{k+1}-x_k\|^2 - \frac{1}{2\tau_k}\left\|F(x_k) + F'(x_k)(x_{k+1}-x_k)\right\|^2. \end{aligned} \tag{9}$$

Подставим выражение x_{k+1} в (9):

$$\begin{aligned} & f_1(x_k) - f_1(x_{k+1}) \ge f_1(x_k) - \frac{\tau_k}{2} - \frac{1}{2\tau_k}\left\|F(x_k) + F'(x_k)(x_{k+1}-x_k)\right\|^2 - \\ & - \frac{L_k}{2}\|x_{k+1}-x_k\|^2 = f_1(x_k) - \frac{\tau_k}{2} - \\ & - \frac{1}{2\tau_k}\left\|F(x_k) - \eta_k F'(x_k)\left(L_k\tau_k I_n + F'(x_k)^\top F'(x_k)\right)^{-1} F'(x_k)^\top F(x_k)\right\|^2 - \\ & - \frac{L_k}{2}\left\|\eta_k\left(L_k\tau_k I_n + F'(x_k)^\top F'(x_k)\right)^{-1} F'(x_k)^\top F(x_k)\right\|^2 = \\ & = f_1(x_k) - \frac{\tau_k}{2} - \frac{f_2(x_k)}{2\tau_k} + \\ & + \frac{1}{2\tau_k}\left(2\left\langle \eta_k\left(L_k\tau_k I_n + F'(x_k)^\top F'(x_k)\right)^{-1} F'(x_k)^\top F(x_k), F'(x_k)^\top F(x_k)\right\rangle - \right. \\ & - \left\langle \eta_k\left(L_k\tau_k I_n + F'(x_k)^\top F'(x_k)\right)^{-1} F'(x_k)^\top F(x_k), \right. \\ & \left. F'(x_k)^\top F'(x_k)\eta_k\left(L_k\tau_k I_n + F'(x_k)^\top F'(x_k)\right)^{-1} F'(x_k)^\top F(x_k)\right\rangle - \\ & - L_k\tau_k\left\langle \eta_k\left(L_k\tau_k I_n + F'(x_k)^\top F'(x_k)\right)^{-1} F'(x_k)^\top F(x_k), \right. \end{aligned} \tag{10}$$

$$
\begin{aligned}
&\eta_k \left(L_k\tau_k I_n + F^{'}(x_k)^\top F^{'}(x_k) \right)^{-1} F^{'}(x_k)^\top F(x_k) \Big\rangle \Big) = \\
&= f_1(x_k) - \frac{\tau_k}{2} - \frac{f_2(x_k)}{2\tau_k} + \\
&+ \frac{\eta_k(2-\eta_k)}{2\tau_k} \left\langle \left(L_k\tau_k I_n + F^{'}(x_k)^\top F^{'}(x_k) \right)^{-1} F^{'}(x_k)^\top F(x_k), F^{'}(x_k)^\top F(x_k) \right\rangle = \\
&= f_1(x_k) - \frac{\tau_k}{2} - \frac{f_2(x_k)}{2\tau_k} + \\
&+ \frac{\eta_k(2-\eta_k)}{2\tau_k} \left\langle F^{'}(x_k) \left(L_k\tau_k I_n + F^{'}(x_k)^\top F^{'}(x_k) \right)^{-1} F^{'}(x_k)^\top F(x_k), F(x_k) \right\rangle .
\end{aligned} \tag{11}
$$

Тогда $f_1(x_{k+1})$ имеет верхнюю оценку по (10):

$$
\begin{aligned}
&f_1(x_{k+1}) \leq \frac{\tau_k}{2} + \frac{f_2(x_k)}{2\tau_k} - \\
&- \frac{\eta_k(2-\eta_k)}{2\tau_k} \left\langle F^{'}(x_k) \left(L_k\tau_k I_n + F^{'}(x_k)^\top F^{'}(x_k) \right)^{-1} F^{'}(x_k)^\top F(x_k), F(x_k) \right\rangle \leq \\
&\leq \{\text{лемма 2}\} \leq \frac{\tau_k}{2} + \frac{f_2(x_k)}{2\tau_k} - \frac{\eta_k(2-\eta_k)}{2\tau_k} \left(\frac{\|F(x_k)\|^2\mu}{L_k\tau_k + \mu} \right) = \\
&= \frac{\tau_k}{2} + \frac{f_2(x_k)}{2\tau_k} \left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k\tau_k + \mu} \right).
\end{aligned}
$$

Первая часть (7) выведена, рассмотрим $\tau_k < \frac{\mu}{L_k}$:

$$
\begin{aligned}
&f_1(x_{k+1}) \leq \frac{\tau_k}{2} + \frac{f_2(x_k)}{2\tau_k} \left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k\tau_k + \mu} \right) = \frac{\tau_k}{2} + \\
&+ \frac{f_2(x_k)}{2\tau_k} \left(1 - \eta_k(2-\eta_k) \left(1 + \frac{L_k\tau_k}{\mu} \right)^{-1} \right) = \\
&= \left\{ (1+y)^{-1} = 1 - y + \frac{y^2}{(1+\xi)^3}, |y| < 1, \text{ для некоторого } \xi \in (-1, 1] \right\} = \\
&= \frac{\tau_k}{2} + \frac{f_2(x_k)(1-\eta_k)^2}{2\tau_k} + \\
&+ \frac{\eta_k(2-\eta_k)L_k f_2(x_k)}{2\mu} - \frac{\eta_k(2-\eta_k)L_k^2 f_2(x_k)\tau_k}{2\mu^2(1+\xi)^3},
\end{aligned}
$$

вторая часть (7) по формуле Тейлора с остаточным членом в форме Лагранжа выведена. Докажем (8), согласно доказанному неравенству (7):

$$f_1(x_{k+1}) \leq \frac{\tau_k}{2} + \left[\begin{array}{l} \dfrac{f_2(x_k)}{2\tau_k} \underbrace{\left(1 - \dfrac{\eta_k(2-\eta_k)\mu}{L_k\tau_k + \mu}\right)}_{\in(0,1)} \leq \dfrac{f_2(x_k)}{2\tau_k} \leq \dfrac{f_2(x_k)L_k}{2\mu} \leq \\ \qquad \leq \{L_k \leq 2L_F\} \leq \dfrac{f_2(x_k)L_F}{\mu}, \tau_k \geq \dfrac{\mu}{L_k}; \\ \dfrac{f_2(x_k)(1-\eta_k)^2}{2\tau_k} + \dfrac{\eta_k(2-\eta_k)L_k f_2(x_k)}{2\mu} - \\ - \underbrace{\dfrac{\eta_k(2-\eta_k)L_k^2 f_2(x_k)\tau_k}{2\mu^2(1+\xi)^3}}_{\geq 0} \leq \\ \leq \{\eta_k = 1, L_k \leq 2L_F\} \leq \dfrac{f_2(x_k)L_F}{\mu}, \text{ иначе.} \end{array}\right.$$

□

Теорема 2 использует явную формулу проксимального отображения для вывода условий сходимости к стационарной точке.

Теорема 2. Пусть выполнено предположение 1. Рассмотрим последовательность $\{x_k\}_{k\in\mathbb{Z}_+}$, вычисляемую по схеме 1 с правилом (4), в котором $\tau_k > 0$ и $\eta_k > 0$, $\eta_k(2-\eta_k) \geq c > 0$, $k \in \mathbb{Z}_+$. Дополнительно предположим ограниченность нормы матрицы Якоби: существует $M_F > 0$, для которого выполнено $\left\|F^{'}(x)\right\| \leq M_F$ при всех $x \in \mathcal{F}$. Тогда при $k \in \mathbb{N}$:

$$\min_{i\in\overline{0,k-1}} \left\{\|\nabla f_2(x_i)\|^2\right\} \leq \frac{8\left(2L_F \max\limits_{i\in\overline{0,k-1}}\{\tau_i\} + M_F^2\right)}{\eta(2-\eta)k} \sum_{i=0}^{k-1} \left(\frac{1}{2}\left(\tau_i - f_1(x_i)\right)^2 + \right.$$
$$\left. + \tau_i\left(f_1(x_i) - f_1(x_{i+1})\right)\right), \quad \eta \in \operatorname*{Argmin}_{k\in\mathbb{Z}_+}\{\eta_k(2-\eta_k)\}.$$

Доказательство. Используя рассуждения из теоремы 1 получаем следующее:

$$\begin{aligned} &f_1(x_k) - f_1(x_{k+1}) \geq f_1(x_k) - f_1(y_{k+1}) \geq f_1(x_k) - \frac{\tau_k}{2} - \frac{f_2(x_k)}{2\tau_k} + \\ &+ \frac{\eta_k(2-\eta_k)}{2\tau_k}\left\langle \left(L_k\tau_k I_n + F^{'}(x_k)^\top F^{'}(x_k)\right)^{-1} F^{'}(x_k)^\top F(x_k), F^{'}(x_k)^\top F(x_k)\right\rangle \geq \\ &\geq \left\{\text{ограничение спектра матрицы снизу}, \nabla f_2(x_k) = 2F^{'}(x_k)^\top F(x_x)\right\} \geq \\ &\geq f_1(x_k) - \frac{\tau_k}{2} - \frac{f_2(x_k)}{2\tau_k} + \frac{\eta_k(2-\eta_k)}{2\tau_k}\left(\frac{\|\nabla f_2(x_k)\|^2}{4(L_k\tau_k + M_F^2)}\right). \end{aligned} \tag{12}$$

Из (12) следует:

$$\begin{aligned} f_1(x_{k+1}) &\leq \frac{\tau_k}{2} + \frac{f_2(x_k)}{2\tau_k} - \frac{\eta_k(2-\eta_k)\|\nabla f_2(x_k)\|^2}{8\tau_k(L_k\tau_k + M_F^2)} \leq \\ &\leq \frac{\tau_k}{2} + \frac{f_2(x_k)}{2\tau_k} - \frac{\eta(2-\eta)\|\nabla f_2(x_k)\|^2}{8\tau_k\left(2L_F\tau_k + M_F^2\right)}. \end{aligned}$$

Тогда, оценивая сверху минимальное значение квадрата нормы градиента $\nabla f_2(x_k)$ для $k \in \mathbb{N}$, выведем соотношение:

$$\frac{\eta(2-\eta)\left\|\nabla f_2(x_k)\right\|^2}{8\tau_k\left(2L_F \max\limits_{i\in\overline{0,k-1}}\{\tau_i\}+M_F^2\right)} \leq \frac{\eta(2-\eta)\left\|\nabla f_2(x_k)\right\|^2}{8\tau_k\left(2L_F\tau_k+M_F^2\right)} \leq \frac{\tau_k}{2}+\frac{f_2(x_k)}{2\tau_k}-$$

$$-f_1(x_{k+1}) \Rightarrow \frac{\eta(2-\eta)\left\|\nabla f_2(x_k)\right\|^2}{8\left(2L_F \max\limits_{i\in\overline{0,k-1}}\{\tau_i\}+M_F^2\right)} \leq \frac{\tau_k^2}{2}+\frac{f_2(x_k)}{2}-\tau_k f_1(x_{k+1}) \Rightarrow$$

$$\Rightarrow \frac{\eta(2-\eta)k}{8\left(2L_F \max\limits_{i\in\overline{0,k-1}}\{\tau_i\}+M_F^2\right)} \min_{i\in\overline{0,k-1}}\left\{\left\|\nabla f_2(x_i)\right\|^2\right\} \leq$$

$$\leq \frac{\eta(2-\eta)}{8\left(2L_F \max\limits_{i\in\overline{0,k-1}}\{\tau_i\}+M_F^2\right)} \sum_{i=0}^{k-1}\left\|\nabla f_2(x_i)\right\|^2 \leq$$

$$\leq \sum_{i=0}^{k-1}\left(\frac{\tau_i^2}{2}+\frac{f_2(x_i)}{2}-\tau_i f_1(x_{i+1})\right)=$$

$$=\sum_{i=0}^{k-1}\left(\frac{1}{2}\left(\tau_i-f_1(x_i)\right)^2+\tau_i\left(f_1(x_i)-f_1(x_{i+1})\right)\right) \Rightarrow \min_{i\in\overline{0,k-1}}\left\{\left\|\nabla f_2(x_i)\right\|^2\right\} \leq$$

$$\leq \frac{8\left(2L_F \max\limits_{i\in\overline{0,k-1}}\{\tau_i\}+M_F^2\right)}{\eta(2-\eta)k} \sum_{i=0}^{k-1}\left(\frac{1}{2}\left(\tau_i-f_1(x_i)\right)^2+\tau_i\left(f_1(x_i)-f_1(x_{i+1})\right)\right).$$

□

В теореме 3 используется стратегия выбора $\tau_k = \mathrm{O}\left(f_1(x_k)\right)$ для вывода линейной сходимости к решению системы (1) при выполнении условия Поляка–Лоясиевича.

Теорема 3. Пусть условия теоремы 1 верны, определим на каждой итерации метода трех квадратов максимальный коэффициент линейной сходимости α_k:

$$\alpha_k \in \left[\frac{1}{2}+\frac{L_k f_1(x_k)+(1-\eta_k)^2\mu}{2\left(L_k f_1(x_k)+\mu\right)}, 1\right), \quad k \in \mathbb{Z}_+.$$

Дополнительно предположим, что на каждой итерации $\tau_k = c_k f_1(x_k)$, $k \in \mathbb{Z}_+$:

$$c_k \in \left[\frac{L_k f_1(x_k)+(1-\eta_k)^2\mu}{(2\alpha_k-1)(L_k f_1(x_k)+\mu)}, \quad \alpha_k-\frac{1}{2}-\frac{\mu}{2L_k f_1(x_k)}+\right.$$
$$\left.+\sqrt{\alpha_k^2+\alpha_k\left(\frac{\mu}{L_k f_1(x_k)}-1\right)+\frac{1}{4}\left(\frac{\mu}{L_k f_1(x_k)}+1\right)^2-\frac{(1-\eta_k)^2\mu}{L_k f_1(x_k)}}\right]. \tag{13}$$

Тогда метод трех квадратов с вычислением x_{k+1} по правилу (4) глобально сходится не хуже, чем линейно к решению задачи (2) $\lim\limits_{k\to+\infty} x_k = x^* : F(x^*) = \mathbf{0}_m$ со следующей оценкой:

$$f_1(x_k) \leq f_1(x_0)\prod_{i=0}^{k-1}\alpha_i, \quad k \in \mathbb{Z}_+, \prod_{i=0}^{-1}\alpha_i \overset{def}{=} 1.$$

Доказательство. Для удобства рассмотрим следующую декомпозицию задачи поиска области определения $c_k, k \in \mathbb{Z}_+$:

$$\left[\begin{aligned} &1. \tau_k \in \left[\frac{f_1(x_k)}{c_k^{-1}}, f_1(x_k)\right], c_k^{-1} \geq 1; \\ &2. \tau_k \in \left[f_1(x_k), c_k f_1(x_k)\right], c_k \geq 1. \end{aligned}\right.$$

Выведем границы значений c_k для первого случая. Воспользуемся результатом теоремы 1:

$$\begin{aligned} f_1(x_{k+1}) \leq \frac{\tau_k}{2} + \frac{f_2(x_k)}{2\tau_k}\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k\tau_k + \mu}\right) \leq \frac{f_1(x_k)}{2} + \\ + \frac{c_k^{-1} f_1(x_k)}{2}\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k f_1(x_k) + \mu}\right) = \\ = f_1(x_k)\left(\frac{c_k^{-1}+1}{2} - \frac{\eta_k(2-\eta_k)c_k^{-1}\mu}{2\left(L_k f_1(x_k) + \mu\right)}\right) \end{aligned}$$

По определению линейной сходимости:

$$\left(\frac{c_k^{-1}+1}{2} - \frac{\eta_k(2-\eta_k)c_k^{-1}\mu}{2\left(L_k f_1(x_k) + \mu\right)}\right) \in [0, \alpha_k], \alpha_k \in (0, 1).$$

Тогда:

$$\begin{aligned} &0 \leq \frac{c_k^{-1}+1}{2} - \frac{\eta_k(2-\eta_k)c_k^{-1}\mu}{2\left(L_k f_1(x_k) + \mu\right)} \leq \alpha_k < 1 \Rightarrow \\ &\Rightarrow -1 \leq c_k^{-1} - \frac{\eta_k(2-\eta_k)c_k^{-1}\mu}{L_k f_1(x_k) + \mu} \leq 2\alpha_k - 1 \Rightarrow \\ &\Rightarrow -1 \leq c_k^{-1}\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k f_1(x_k) + \mu}\right) \leq 2\alpha_k - 1 \Rightarrow \\ &\Rightarrow -\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k f_1(x_k) + \mu}\right)^{-1} \leq c_k^{-1} \leq (2\alpha_k - 1)\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k f_1(x_k) + \mu}\right)^{-1} \Rightarrow \\ &\Rightarrow 0 \leq c_k^{-1} \leq (2\alpha_k - 1)\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k f_1(x_k) + \mu}\right)^{-1} = \\ &= (2\alpha_k - 1)\left(\frac{L_k f_1(x_k) + \mu}{L_k f_1(x_k) + (1 - \eta_k(2-\eta_k))\mu}\right). \end{aligned}$$

Применим ограничения на c_k^{-1} из условия:

$$1 \leq c_k^{-1} \leq (2\alpha_k - 1)\left(\frac{L_k f_1(x_k) + \mu}{L_k f_1(x_k) + (1-\eta_k)^2\mu}\right).$$

Следовательно,

$$c_k \in \left[\frac{L_k f_1(x_k) + (1-\eta_k)^2\mu}{(2\alpha_k - 1)(L_k f_1(x_k) + \mu)}, 1\right].$$

Определим для таких c_k допустимые значения α_k:

$$\begin{aligned} &c_k^{-1} = (2\alpha_k - 1)\left(\frac{L_k f_1(x_k) + \mu}{L_k f_1(x_k) + (1-\eta_k)^2\mu}\right) \geq 1 \Rightarrow \\ &\Rightarrow \alpha_k \in \left[\frac{1}{2} + \frac{L_k f_1(x_k) + (1-\eta_k)^2\mu}{2\left(L_k f_1(x_k) + \mu\right)}, 1\right). \end{aligned} \tag{14}$$

Теперь рассмотрим второй случай:

$$\tau_k \in [f_1(x_k), c_k f_1(x_k)], c_k \geq 1.$$

По теореме 1:

$$\begin{aligned} f_1(x_{k+1}) \leq \frac{\tau_k}{2} + \frac{f_2(x_k)}{2\tau_k}\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k\tau_k + \mu}\right) \leq \frac{c_k f_1(x_k)}{2} + \\ + \frac{f_1(x_k)}{2}\left(1 - \frac{\eta_k(2-\eta_k)\mu}{L_k c_k f_1(x_k) + \mu}\right) = \\ = f_1(x_k)\left(\frac{c_k+1}{2} - \frac{\eta_k(2-\eta_k)\mu}{2\left(L_k f_1(x_k) c_k + \mu\right)}\right). \end{aligned}$$

По определению линейной сходимости:

$$\left(\frac{c_k+1}{2} - \frac{\eta_k(2-\eta_k)\mu}{2\left(L_k f_1(x_k) c_k + \mu\right)}\right) \in [0, \alpha_k], \alpha_k \in (0,1).$$

Распишем подробнее данные неравенства:

$$\begin{aligned} & 0 \leq \frac{c_k+1}{2} - \frac{\eta_k(2-\eta_k)\mu}{2\left(L_k f_1(x_k) c_k + \mu\right)} \leq \alpha_k < 1 \Rightarrow \\ & \Rightarrow -1 \leq c_k - \frac{\eta_k(2-\eta_k)\mu}{L_k f_1(x_k) c_k + \mu} \leq 2\alpha_k - 1 \Rightarrow \\ & \Rightarrow -\mu - L_k f_1(x_k) c_k \leq (c_k)^2 L_k f_1(x_k) + c_k\mu - \eta_k(2-\eta_k)\mu \leq \\ & \leq (2\alpha_k - 1)\left(L_k f_1(x_k) c_k + \mu\right) \Rightarrow \\ & \Rightarrow 0 \leq (c_k)^2 L_k f_1(x_k) + c_k\left(\mu + L_k f_1(x_k)\right) + (1-\eta_k)^2\mu \leq 2\alpha_k\left(L_k f_1(x_k) c_k + \mu\right). \end{aligned}$$

Рассмотрим левое неравенство:

$$0 \leq (c_k)^2 L_k f_1(x_k) + c_k\left(\mu + L_k f_1(x_k)\right) + (1-\eta_k)^2\mu.$$

Введем замену переменных $b_k = \frac{L_k f_1(x_k)}{\mu}$, подставим в неравенства:

$$0 \leq b_k(c_k)^2 + c_k(1+b_k) + (1-\eta_k)^2 \leq 2\alpha_k(b_k c_k + 1).$$

Рассмотрим дискриминант полинома второй степени, записанного в центральном неравенстве выше:

$$(1+b_k)^2 - 4b_k(1-\eta_k)^2 = b_k^2 + b_k(2 - 4(1-\eta_k)^2) + 1.$$

При тех η_k, при которых дискриминант отрицателен, новые ограничения на c_k в левом неравенстве не возникают. При остальных η_k получаются следующие корни:

$$c_k = -\frac{1}{2} - \frac{1}{2b_k} \pm \sqrt{\frac{1}{4} + \frac{1}{2b_k} - \frac{(1-\eta_k)^2}{b_k} + \frac{1}{4b_k^2}}.$$

Левый корень не накладывает дополнительные ограничения, так как он отрицателен. Правый корень не накладывает ограничений, потому что он левее $c_k = 1$:

$$\begin{aligned} & -\frac{1}{2} - \frac{1}{2b_k} + \sqrt{\frac{1}{4} + \frac{1}{2b_k} - \frac{(1-\eta_k)^2}{b_k} + \frac{1}{4b_k^2}} \vee 1 \Rightarrow \\ & \Rightarrow \sqrt{\frac{1}{4} + \frac{1}{2b_k} - \frac{(1-\eta_k)^2}{b_k} + \frac{1}{4b_k^2}} \vee \frac{3}{2} + \frac{1}{2b_k} \Rightarrow \\ & \Rightarrow \frac{1}{4} + \frac{1}{2b_k} - \frac{(1-\eta_k)^2}{b_k} + \frac{1}{4b_k^2} \vee \frac{9}{4} + \frac{3}{2b_k} + \frac{1}{4b_k} \Rightarrow -\frac{(1-\eta_k)^2}{b_k} \leq 2 + \frac{1}{b_k}; \end{aligned}$$

то есть дополнительные ограничения не накладываются:

$$c_k \geq \max\left\{1, -\frac{1}{2} - \frac{1}{2b_k} + \sqrt{\frac{1}{4} + \frac{1}{2b_k} - \frac{(1-\eta_k)^2}{b_k} + \frac{1}{4b_k^2}}\right\} = 1.$$

Перейдем к правому неравенству сразу с подстановкой $b_k = \frac{L_k f_1(x_k)}{\mu}$:

$$\begin{aligned} &b_k(c_k)^2 + c_k(1+b_k) + (1-\eta_k)^2 \leq 2\alpha_k(b_k c_k + 1) \Rightarrow \\ &\Rightarrow b_k(c_k)^2 + c_k(1 + b_k - 2\alpha_k b_k) + (1-\eta_k)^2 - 2\alpha_k \leq 0. \end{aligned}$$

Ограничения задаются полиномом второй степени, дискриминант которого равен:

$$\begin{aligned} &(1 + b_k - 2\alpha_k b_k)^2 - 4b_k((1-\eta_k)^2 - 2\alpha_k) = \\ &= \alpha_k^2(4b_k^2) + \alpha_k(4b_k(1-b_k)) + ((b_k+1)^2 - 4b_k(1-\eta_k)^2). \end{aligned}$$

Дискриминант полинома является параболой по α_k с положительным коэффициентом при старшем члене. Рассмотрим значение данной параболы в точке вершины $\alpha_k = \frac{b_k-1}{2b_k}$:

$$\begin{aligned} &4b_k^2\left(\frac{b_k-1}{2b_k}\right)^2 + 4b_k\left(\frac{b_k-1}{2b_k}\right)(1-b_k) + (b_k+1)^2 - 4b_k(1-\eta_k)^2 = \\ &= 4b_k\eta_k(2-\eta_k) \geq 0, \eta_k \in (0,2). \end{aligned}$$

То есть множество допустимых значений c_k непусто. Тогда граничные точки c_k равны:

$$c_k = \alpha_k - \frac{b_k+1}{2b_k} \pm \sqrt{\alpha_k^2 + \frac{\alpha_k(1-b_k)}{b_k} + \left(\frac{1+b_k}{2b_k}\right)^2 - \frac{(1-\eta_k)^2}{b_k}}.$$

У параболы свободный член неположителен и равен $(1-\eta_k)^2 - 2\alpha_k$, так как по (14) $\alpha_k \geq \frac{1}{2}$ и $\eta_k \in (0,2)$. Поэтому по теореме Виета некратные ненулевые корни многочлена второй степени обладают разными знаками, сама парабола имеет положительный коэффициент при старшем члене, что по ограничению $c_k \geq 1$ для множества допустимых значений составляет отрезок:

$$\begin{aligned} c_k \in &\left[1, \alpha_k - \frac{b_k+1}{2b_k} + \sqrt{\alpha_k^2 + \frac{\alpha_k(1-b_k)}{b_k} + \left(\frac{1+b_k}{2b_k}\right)^2 - \frac{(1-\eta_k)^2}{b_k}}\right] = \\ = &\left[1, \alpha_k - \frac{1}{2} - \frac{\mu}{2L_k f_1(x_k)} + \right. \\ &\left. + \sqrt{\alpha_k^2 + \alpha_k\left(\frac{\mu}{L_k f_1(x_k)} - 1\right) + \frac{1}{4}\left(\frac{\mu}{L_k f_1(x_k)} + 1\right)^2 - \frac{(1-\eta_k)^2\mu}{L_k f_1(x_k)}}\right]. \end{aligned}$$

Определим для таких c_k допустимые значения α_k:

$$
\begin{aligned}
&c_k = \alpha_k - \frac{b_k+1}{2b_k} + \sqrt{\alpha_k^2 + \frac{\alpha_k(1-b_k)}{b_k} + \left(\frac{1+b_k}{2b_k}\right)^2 - \frac{(1-\eta_k)^2}{b_k}} \geq 1 \Rightarrow \\
&\Rightarrow \alpha_k^2 + \frac{\alpha_k(1-b_k)}{b_k} + \left(\frac{1+b_k}{2b_k}\right)^2 - \frac{(1-\eta_k)^2}{b_k} \geq \left(1 + \frac{1+b_k}{2b_k} - \alpha_k\right)^2 \Rightarrow \\
&\Rightarrow \alpha_k(1-b_k) - (1-\eta_k)^2 \geq 1 + 2b_k - \alpha_k(1+b_k) - 2\alpha_k b_k \Rightarrow 2\alpha_k(b_k+1) \geq \\
&\geq 2b_k + 1 + (1-\eta_k)^2 \Rightarrow \alpha_k \geq \frac{2b_k + 1 + (1-\eta_k)^2}{2(b_k+1)} = \frac{1}{2} + \frac{b_k + (1-\eta_k)^2}{2(b_k+1)} \Rightarrow \\
&\Rightarrow \alpha_k \in \left[\frac{1}{2} + \frac{L_k f_1(x_k) + (1-\eta_k)^2\mu}{2\left(L_k f_1(x_k) + \mu\right)}, 1\right).
\end{aligned}
$$

Объединяя рассмотренные случаи, получаем искомое (13):

$$
\begin{cases}
c_k \in \left[\dfrac{L_k f_1(x_k) + (1-\eta_k)^2\mu}{(2\alpha_k - 1)(L_k f_1(x_k) + \mu)}, \quad \alpha_k - \dfrac{1}{2} - \dfrac{\mu}{2L_k f_1(x_k)} + \right. \\
\qquad \left. + \sqrt{\alpha_k^2 + \alpha_k\left(\dfrac{\mu}{L_k f_1(x_k)} - 1\right) + \dfrac{1}{4}\left(\dfrac{\mu}{L_k f_1(x_k)} + 1\right)^2 - \dfrac{(1-\eta_k)^2\mu}{L_k f_1(x_k)}}\right]. \\
\alpha_k \in \left[\frac{1}{2} + \frac{L_k f_1(x_k) + (1-\eta_k)^2\mu}{2(L_k f_1(x_k) + \mu)}, 1\right).
\end{cases}
$$

Установлена линейная сходимость метода трех квадратов, что по свойствам минимизируемой нормы приводит итерационный процесс к решению $\lim\limits_{k\to+\infty} x_k = x^* : F(x^*) = \mathbf{0}_m$. Данная сходимость носит глобальный характер, так как для $x_0 \in \mathcal{F} : \mathcal{L}(f_1(x_0)) \subseteq \mathcal{F}$, верно вложение

$$
x_k \in \mathcal{L}(f_1(x_k)) \subseteq \mathcal{L}(f_1(x_0)) \subseteq \mathcal{F}, \quad k \in \mathbb{Z}_+.
$$

Из доказательства непосредственно выводится оценка сходимости:

$$
f_1(x_k) \leq f_1(x_0) \prod_{i=0}^{k-1} \alpha_i, \quad k \in \mathbb{N}.
$$

□

Графики и таблицы с результатами экспериментов

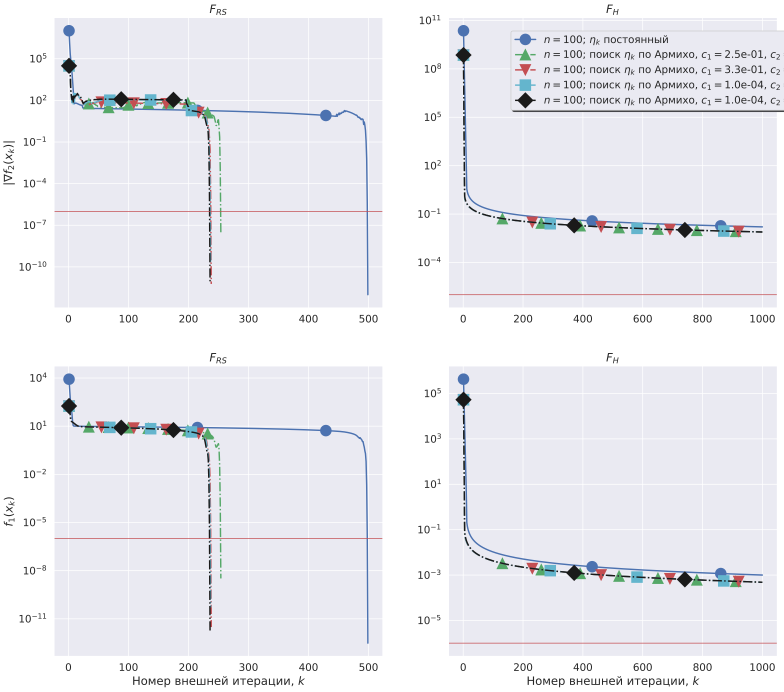

Рис. 1. Сходимость алгоритма 1 с константной стратегией поиска t_k, n — размерность задачи

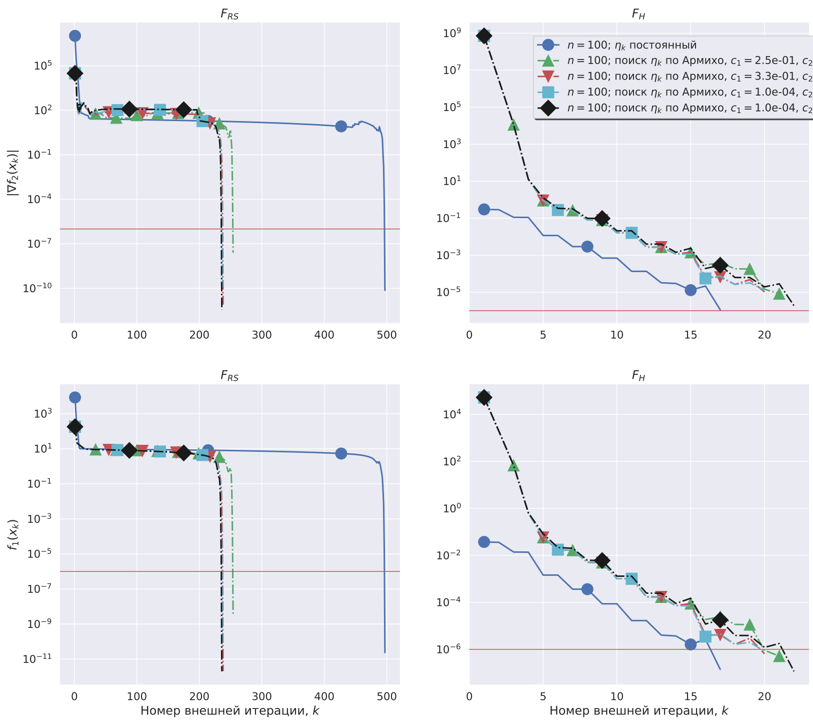

Рис. 2. Сходимость алгоритма 1 со стратегией экстраполяции при поиске t_k, n — размерность задачи

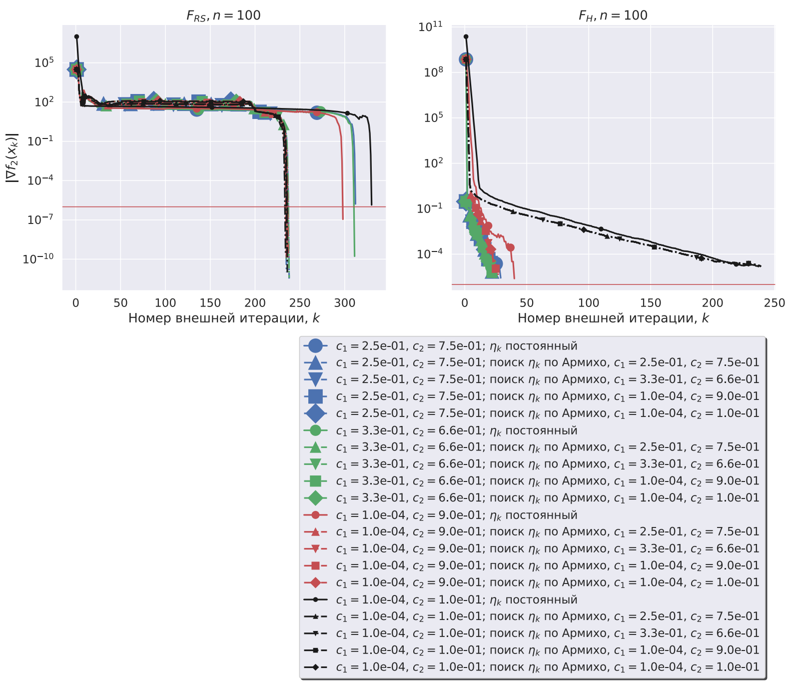

Рис. 3. Сходимость алгоритма 1 по $\|\nabla f_2(x_k)\|$ с поиском t_k по правилу Армихо, n — размерность задачи, c_1 и c_2 — факторы наклона лучей в условии Армихо

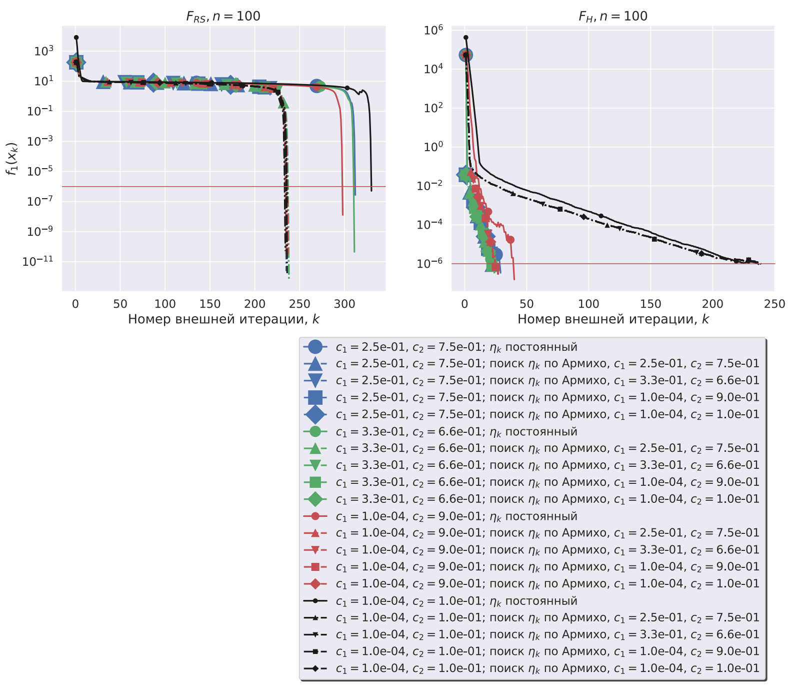

Рис. 4. Сходимость алгоритма 1 по значению функции $f_1(x_k)$ с поиском t_k по правилу Армихо, n — размерность задачи, c_1 и c_2 — факторы наклона лучей в условии Армихо

Список литературы (References)

Поляк Б. Т. Градиентные методы минимизации функционалов //Журнал вычислительной математики и математической физики. — 1963. — Т. 3. — №. 4. — С. 643–653.

Polyak B. T. Gradientnye metody minimizacii funkcionalov [Gradient methods for minimizing functionals] //Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki. — 1963. — Vol. 3. — No. 4. — P. 643–653 (in Russian).

Nesterov Yu. Flexible Modification of Gauss–Newton Method //CORE Discussion Papers. — 2021.