Метод проекции градиента

Расширение метода Розена для задачи поиска экстремума выпуклой функции на области, заданной нелинейными ограничениями

Приведенный в статье подход актуален при решении практических задач интеллектуального анализа данных, в частности, для поиска параметров, максимизирующих функцию правдоподобия в алгоритмах кластеризации, классификации и подобных.

Известный метод проекции градиента служит для поиска экстремума функции на области, заданной системой аффинных равенств или неравенств. В статье дается расширение алгоритма для решения задачи поиска минимума выпуклой, Липшиц-непрерывно дифференцируемой функции на области, заданной нелинейными ограничениями специального вида. Приведен итеративный алгоритм поиска минимума, доказана сходимость алгоритма в подпоследовательности к точке глобального минимума на допустимой области.

Утверждение 1

Пусть $\varphi:\mathbb{R}^n\to\mathbb{R}$ дифференцируемая по Липшицу с константой L функция; $x,d\in\mathbb{R}^n$ и $\langle d,\Delta\varphi(x)\rangle>0$. Тогда

1. $\exists t_0>0\ \ \forall t\in(0,t_0]:\varphi(x)-\varphi(x-td)\ge\frac{1}{2}t\langle\Delta\varphi(x),d\rangle$
2. Если для t>0 справедливо $\varphi(x)-\varphi(x-2td)<t\langle\Delta\varphi(x),d\rangle$, то

   имеет место оценка $t\ge\dfrac{\langle\Delta\varphi(x),d\rangle}{2L\|d\|^2}$

Доказательство

1. Из дифференцируемости φ следует: $\varphi(x-td)=\varphi(x)-t\left\langle\Delta\varphi(x),d\right\rangle-o(td)\Rightarrow$

$$\Rightarrow\varphi(x)-\varphi(x-td)=t\left\langle\Delta\varphi(x),d\right\rangle+o(td)=t\|d\|\left(\frac{\left\langle\Delta\varphi(x),d\right\rangle}{\|d\|}+\frac{o(td)}{t\|d\|}\right)$$

Так как $\dfrac{\left\langle\Delta\varphi(x),d\right\rangle}{\|d\|}>0$ и $\lim\limits_{t\to0}\dfrac{o(td)}{t\|d\|}=0$, то существует $t_0>0$, такой что $\forall t\in(0,t_0]$:

$\left|\dfrac{o(td)}{t\|d\|}\right|<\dfrac{1}{2}\dfrac{\left\langle\Delta\varphi(x),d\right\rangle}{\|d\|}$, а следовательно $\forall t\in(0,t_0]$:

$$\varphi(x)-\varphi(x-td)=t\|d\|\left(\frac{\left\langle\Delta\varphi(x),d\right\rangle}{\|d\|}+\frac{o(td)}{t\|d\|}\right)\ge t\|d\|\left(\frac{\left\langle\Delta\varphi(x),d\right\rangle}{\|d\|}-\frac{1}{2}\frac{\left\langle\Delta\varphi(x),d\right\rangle}{\|d\|}\right)=$$

$$=t\|d\|\frac{1}{2}\frac{\left\langle\Delta\varphi(x),d\right\rangle}{\|d\|}=\frac{1}{2}t\left\langle\Delta\varphi(x),d\right\rangle$$

2. Мы имеем $\varphi(x)-\varphi(x-2td)<t\langle\Delta\varphi(x),d\rangle$. Учитывая Липшиц-дифференцируемость φ, получаем :

$$\varphi(x)-\varphi(x-2td)<t\langle\Delta\varphi(x),d\rangle\Rightarrow-t\langle\Delta\varphi(x),d\rangle<\varphi(x-2td)-\varphi(x)=-2t\int_0^1\left\langle\Delta\varphi(x-2tud),d\right\rangle du=$$

$$=-2t\langle\Delta\varphi(x),d\rangle+2t\int_0^1\left\langle\Delta\varphi(x)-\Delta\varphi(x-2tud),d\right\rangle du\Rightarrow$$

$$\Rightarrow t\langle\Delta\varphi(x),d\rangle<2t\int_0^1\left\langle\Delta\varphi(x-2tud)-\Delta\varphi(x),d\right\rangle du\le2t\int_0^1\|\Delta\varphi(x-2tud)-\Delta\varphi(x)\|\|d\|du\le$$

$$\le2tL\int_0^1\|2tud\|\|d\|du=2t2tL\|d\|^2\int_0^1udu=2t^2L\|d\|^2\Rightarrow t>\frac{\langle\Delta\varphi(x),d\rangle}{2L\|d\|^2}$$

$\square$

Утверждение 2

Пусть $\varphi:\mathbb{R}^n \to \mathbb{R}$ дифференцируемая по Липшицу с константой L функция; $x, d \in \mathbb{R}^n$; $\varphi(x) \le 0$ и $\langle d, \Delta\varphi(x)\rangle > 0$. Тогда

1. $\exists t_0 > 0 \ \ \forall t \in (0, t_0] : \varphi(x - td) < 0$
2. Если для t>0 выполняется $\varphi(x - 2td) \ge 0$, то справедливо $t \ge \dfrac{\langle \Delta\varphi(x), d\rangle}{L\|d\|^2}$

Доказательство

1. Из утверждения 1 следует, что для достаточно малых t справедливо $\varphi(x) - \varphi(x - td) \ge \dfrac{1}{2} t \langle \Delta\varphi(x), d\rangle$, а следовательно $0 \ge \varphi(x) \ge \varphi(x - td) + \underbrace{\dfrac{1}{2} t \langle \Delta\varphi(x), d\rangle}_{>0} > \varphi(x - td)$, т.е.

   (*) $\varphi(x - td) < 0$ для достаточно малых t.
2. Пусть $\varphi(x - 2td) \ge 0$. Учитывая Липшиц-дифференцируемость φ, получаем :

$$0 \le \varphi(x - 2td) \underbrace{\Rightarrow}_{\varphi(x)\le 0} 0 \le \varphi(x - 2td) - \varphi(x) = -2t\int_0^1 \langle \Delta\varphi(x - 2tud), d\rangle du =$$

$$= -2t\langle \Delta\varphi(x), d\rangle + 2t\int_0^1 \langle \Delta\varphi(x) - \Delta\varphi(x - 2tud), d\rangle du \Rightarrow$$

$$\Rightarrow 2t\langle \Delta\varphi(x), d\rangle \le 2t\int_0^1 \langle \Delta\varphi(x - 2tud) - \Delta\varphi(x), d\rangle du \le 2t\int_0^1 \|\Delta\varphi(x - 2tud) - \Delta\varphi(x)\| \|d\| du \le$$

$$\le 2tL\int_0^1 \|2tud\|\|d\| du = 2t2tL\|d\|^2 \int_0^1 u du = 2t^2 L\|d\|^2 \Rightarrow t \ge \frac{\langle \Delta\varphi(x), d\rangle}{L\|d\|^2}$$

$\square$

Утверждение 3

Пусть $\varphi:\mathbb{R}^n \to \mathbb{R}$ дифференцируемая по Липшицу с константой L функция; $x, d \in \mathbb{R}^n$; $\varepsilon > 0$, $\varphi(x) < -\varepsilon$. Тогда

1. $\exists t_0 > 0 \ \ \forall t \in (0, t_0] : \varphi(x - td) < 0$
2. Если для t>0 справедливо $\varphi(x - 2td) \ge 0$, то имеет место оценка: $t \ge \dfrac{\langle \Delta\varphi(x), d\rangle + \sqrt{2L\varepsilon\|d\|^2 + \langle \Delta\varphi(x), d\rangle^2}}{2\|d\|^2 L}$

Доказательство

1. Из непрерывности ф-ции φ и отрицательности ее значения в точке x следует существование окрестности точки x, где значения ф-ции φ остаются отрицательными, т.е. $\exists t_0 > 0 \ \ \forall t \in (0, t_0] : \varphi(x - td) < 0$.
2. Пусть $\varphi(x - 2td) \ge 0$. Случай $d = 0$ исключен, так как $\varphi(x) < 0$, таким образом $d \ne 0$.Учитывая Липшиц-дифференцируемость φ, получаем :

$$0 \le \varphi(x - 2td) \underbrace{\Rightarrow}_{\varphi(x)<-\varepsilon} \varepsilon \le \varphi(x - 2td) - \varphi(x) = -2t\int_0^1 \langle \Delta\varphi(x - 2tud), d\rangle du =$$

$$= -2t\langle \Delta\varphi(x), d\rangle + 2t\int_0^1 \langle \Delta\varphi(x) - \Delta\varphi(x - 2tud), d\rangle du \le -2t\langle \Delta\varphi(x), d\rangle + \left|2t\int_0^1 \langle \Delta\varphi(x - 2tud) - \Delta\varphi(x), d\rangle du\right| \le$$

$$\le -2t\langle \Delta\varphi(x), d\rangle + 2t\int_0^1 \|\Delta\varphi(x - 2tud) - \Delta\varphi(x)\| \cdot \|d\| du \le -2t\langle \Delta\varphi(x), d\rangle + 2Lt^2\|d\|^2 \Rightarrow$$

(**) $\qquad \Rightarrow 2Lt^2\|d\|^2 - 2t\langle \Delta\varphi(x), d\rangle \ge \varepsilon$.

Уравнение

(***) $$2Lt^2\|d\|^2 - 2t\langle \Delta\varphi(x), d\rangle - \varepsilon = 0,$$

решаемое относительно t с учетом $d \neq 0$, имеет два действительных

корня: $\dfrac{\langle \Delta\varphi(x), d\rangle \pm \sqrt{2L\varepsilon\|d\|^2 + \langle \Delta\varphi(x), d\rangle^2}}{2\|d\|^2 L}$ среди которых только один положительный:

$\dfrac{\langle \Delta\varphi(x), d\rangle + \sqrt{2L\varepsilon\|d\|^2 + \langle \Delta\varphi(x), d\rangle^2}}{2\|d\|^2 L}$. Учитывая положительность коэффициента перед второй степенью t, наличие двух действительных и только одного положительного корня уравнения (**), приходим к выводу, что положительное решение неравенства (**) должно быть больше корня уравнения (***) , т.е.

$$t \geq \frac{\langle \Delta\varphi(x), d\rangle + \sqrt{2L\varepsilon\|d\|^2 + \langle \Delta\varphi(x), d\rangle^2}}{2\|d\|^2 L}$$

□

Утверждение 4

Пусть $A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \in Mat(m;n)$ - матрица $m \times n$; Ran(A) = m, т.е. строки A линейно независимы, тогда матрица $AA^t \in Mat(m;m)$ инвертируема.

Доказательство

$AA^t = \begin{pmatrix} \langle a_1, a_1\rangle & \cdots & \langle a_1, a_m\rangle \\ \vdots & \ddots & \vdots \\ \langle a_m, a_1\rangle & \cdots & \langle a_m, a_m\rangle \end{pmatrix}$. Предположим, что $\exists d = \begin{pmatrix} d_1 \\ \vdots \\ d_m \end{pmatrix} \in \mathbb{R}^m : AA^t d = 0 \Rightarrow$

$$\Rightarrow 0 = AA^t d = d_1 \begin{pmatrix} \langle a_1, a_1\rangle \\ \vdots \\ \langle a_m, a_1\rangle \end{pmatrix} + \ldots + d_m \begin{pmatrix} \langle a_1, a_m\rangle \\ \vdots \\ \langle a_m, a_m\rangle \end{pmatrix} = \begin{pmatrix} \langle a_1, d_1 a_1\rangle \\ \vdots \\ \langle a_m, d_1 a_1\rangle \end{pmatrix} + \ldots + \begin{pmatrix} \langle a_1, d_m a_m\rangle \\ \vdots \\ \langle a_m, d_m a_m\rangle \end{pmatrix} =$$

$$= \begin{pmatrix} \langle a_1, d_1 a_1 + \ldots + d_m a_m\rangle \\ \vdots \\ \langle a_m, d_1 a_1 + \ldots + d_m a_m\rangle \end{pmatrix} \Rightarrow \begin{cases} \langle a_1, d_1 a_1 + \ldots + d_m a_m\rangle = 0 \\ \vdots \\ \langle a_m, d_1 a_1 + \ldots + d_m a_m\rangle = 0 \end{cases} \Rightarrow$$

$$\Rightarrow 0 = d_1 \underbrace{\langle a_1, d_1 a_1 + \ldots + d_m a_m\rangle}_{=0} + \ldots + d_m \underbrace{\langle a_m, d_1 a_1 + \ldots + d_m a_m\rangle}_{=0} = \underbrace{\langle d_1 a_1 + \ldots + d_m a_m, d_1 a_1 + \ldots + d_m a_m\rangle}_{=0} =$$

$= \|d_1 a_1 + \ldots + d_m a_m\|^2 \Rightarrow d_1 a_1 + \ldots + d_m a_m = 0$. Так как вектора $a_1, \ldots, a_m$ линейно независимы, то получаем, что $d = 0 \Rightarrow Ker(AA^t) = \{0\}$, т.е. AA^t - инъективна, из чего вследствии квадратности следует биективность, то есть инвертируемость.

□

Обозначения

1. Пусть $A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \in Mat(m;n)$ - матрица $m \times n$; $I = \{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$. Обозначим

   $A_I := \begin{pmatrix} a_{i_1} \\ \vdots \\ a_{i_k} \end{pmatrix} \in Mat(k;n)$ - матрица состоящая из соответствующих индексам из I столбцов матрицы A.

2. Если строки A_I линейно независимы, то по утверждению 4 матрица $A_I A_I^t$ инвертируема и можно определить $P_I := E_n - A_I^t (A_I A_I^t)^{-1} A_I \in Mat(n;n)$

Утверждение 5

Пусть $A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix} \in Mat(m;n)$ - матрица $m \times n$; $I = \{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$; $Ran(A_I) = m$. Тогда справедливо:

1. $\forall d \in \mathbb{R}^n : P_I d \in Ker(A_I)$, т.е. $A_I P_I d = 0, \forall d \in \mathbb{R}^n$
2. $P_I P_I = P_I$
3. $P_I = P_I^t$
4. $\forall d \in \mathbb{R}^n : \langle P_I d, d \rangle = \| P_I d \|^2$

Доказательство

1. $A_I P_I = A_I (E_n - A_I^t (A_I A_I^t)^{-1} A_I) = A_I E_n - A_I A_I^t (A_I A_I^t)^{-1} A_I = A_I - A_I = 0 \Rightarrow$
   $A_I P_I d = 0, \forall d \in \mathbb{R}^n$
2. $P_I P_I = (E_n - A_I^t (A_I A_I^t)^{-1} A_I)(E_n - A_I^t (A_I A_I^t)^{-1} A_I) =$
   $= E_n - 2A_I^t (A_I A_I^t)^{-1} A_I + A_I^t (A_I A_I^t)^{-1} A_I A_I^t (A_I A_I^t)^{-1} A_I =$
   $= E_n - 2A_I^t (A_I A_I^t)^{-1} A_I + A_I^t (A_I A_I^t)^{-1} A_I = E_n - A_I^t (A_I A_I^t)^{-1} A_I = P_I$
3. $P_I^t = (E_n - A_I^t (A_I A_I^t)^{-1} A_I)^t = \underbrace{E_n^t}_{E_n} - (A_I^t (A_I A_I^t)^{-1} A_I)^t =$
   $= E_n - A_I^t \left((A_I A_I^t)^{-1} \right)^t \left(A_I^t \right)^t = E_n - A_I^t \left((A_I A_I^t)^t \right)^{-1} A_I =$
   $= E_n - A_I^t \left((A_I^t)^t A_I^t \right)^{-1} A_I = E_n - A_I^t \left(A_I A_I^t \right)^{-1} A_I = P_I$
4. $\langle P_I d, d \rangle \underbrace{=}_{2} \langle P_I P_I d, d \rangle = \langle P_I d, P_I^t d \rangle \underbrace{=}_{3} \langle P_I d, P_I d \rangle = \| P_I d \|^2$

□

Обозначения

$f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ дифференцируемые по Липшицу с константой $L > 0$ функции, где f - функция цели, а $g_1, \ldots, g_m$ - функции ограничений.

Предполагаем, что множество $G := \left\{ x \in \mathbb{R}^n \mid g_1(x), \ldots, g_m(x) \le 0 \right\} \ne \varnothing$ ограниченное, а следовательно, компактное множество.

Для каждой точки $x \in G$ определим множество индексов активных ограничений, т.е. функцию
$I : G \to \wp\left(\{1, ..m\} \right), I(x) := \left\{ i_k \in \{1, \ldots, m\} \mid g_{i_k}(x) \in [-\varepsilon, 0] \right\}$.

Далее, для каждой точки $x \in G$ предполагаем, что градиенты функций активных ограничений линейно независимы, т.е. $Ran\begin{pmatrix} \vartriangle g_{i_1}(x) \\ . \\ . \\ . \\ \vartriangle g_{i_k}(x) \end{pmatrix} = k$, где $\{i_1,...,i_k\} = I(x)$. Обозначив $A(x) := \begin{pmatrix} \vartriangle g_{i_1}(x) \\ . \\ . \\ . \\ \vartriangle g_{i_k}(x) \end{pmatrix} \in Mat(k,n)$, получаем по утверждению 4, что матрица $P(x) := E_n - A(x)^t (A(x)A(x)^t)^{-1} A(x) \in Mat(n;n)$ определена.

Утверждение 6

Пусть $x \in G = \left\{x \in \mathbb{R}^n \mid g_1(x),...,g_m(x) \le 0\right\}$,

$A(x) = \begin{pmatrix} \vartriangle g_{i_1}(x) \\ . \\ . \\ . \\ \vartriangle g_{i_k}(x) \end{pmatrix}$-матрица градиентов ε -активных ограничений в точке x, т.е.

$I(x) = \left\{i_k \in \{1,...,m\} \mid g_{i_k}(x) \in [-\varepsilon, 0]\right\}$,

$d_1(x)$ – проекция градиента функции f в точке x на ядро A(x), т.е.

$d_1(x) := P(x) \vartriangle f(x)^t = \left(E_n - A(x)^t (A(x)A(x)^t)^{-1} A(x)\right) \vartriangle f(x)^t$.

Если $d_1(x) \ne 0$, то существует $t(x) > 0$ и $d(x) \ne 0$, такое что

1. $\|d(x)\| \in \left[\|d_1(x)\|, 2\|d_1(x)\|\right], \langle \Delta f(x), d(x)\rangle \ge \frac{1}{2}\|d_1(x)\|^2$
2. $f(x) - f(x - t(x)d(x)) \ge \frac{1}{4} t(x) \|d_1(x)\|^2$
3. $g_i(x - t(x)d(x)) \le 0, \forall i \in \{1,...,m\}$
4. $t(x) \ge \min\left\{K_1, \dfrac{\langle \Delta g_i(x), d(x)\rangle + \sqrt{K_3 \varepsilon \|d(x)\|^2 + \langle \Delta g_i(x), d(x)\rangle^2}}{K_2}\right\}$, где K_1, K_2, K_3>0, константы, а

   $i \in J(x) = \{1,...,m\} \setminus I(x) = \left\{i \in \{1,...,m\} \middle| g_i(x) < -\varepsilon\right\}$

Доказательство

Итак, определим направление:

$d_2(x) := A(x)^t (A(x)A(x)^t)^{-1} e$, где $e = \begin{pmatrix} 1 \\ . \\ . \\ . \\ 1 \end{pmatrix}$.

$d_2(x) \ne 0$, так как $A(x)d_2(x) = e \ne 0$.

Определим

$$\alpha(x) := \min\left\{1, \frac{\|d_1(x)\|}{\|d_2(x)\|}, \begin{cases} -\dfrac{\|d_1(x)\|^2}{2\langle \Delta f(x), d_2(x)\rangle}, \langle \Delta f(x), d_2(x)\rangle < 0 \\ 1, \langle \Delta f(x), d_2(x)\rangle \ge 0 \end{cases}\right\} > 0, \quad \textbf{(1)}$$

$$d(x) := d_1(x) + \alpha(x) d_2(x) \quad \textbf{(2)}$$

Из утверждения 5 следует, что $\langle \triangle f(x), d_1(x) \rangle = \|d_1(x)\|^2$ **(3)**

и, таким образом $\langle \triangle f(x), d(x) \rangle = \|d_1(x)\|^2 + \alpha(x) \langle \triangle f(x), d_2(x) \rangle$

Из определения (1) следует, что $\alpha(x) \le -\dfrac{\|d_1(x)\|^2}{2\langle \Delta f(x), d_2(x)\rangle}$ если $\langle \Delta f(x), d_2(x)\rangle < 0$ и $\alpha(x) > 0$ в любом случае, поэтому

$$\langle \triangle f(x), d(x) \rangle \ge \begin{cases} \dfrac{\|d_1(x)\|^2}{2}, \langle \Delta f(x), d_2(x)\rangle < 0 \\ \|d_1(x)\|^2, \langle \Delta f(x), d_2(x)\rangle \ge 0 \end{cases} \ge \frac{\|d_1(x)\|^2}{2}, \text{ т.е.}$$

$$\langle \triangle f(x), d(x) \rangle \ge \frac{\|d_1(x)\|^2}{2} \quad \textbf{(4)}$$

Далее, $\langle d_1(x), d_2(x) \rangle = d_1(x)^t d_2(x) = e^t (A(x)A(x)^t)^{-1} A(x) \left(E_n - A(x)^t (A(x)A(x)^t)^{-1} A(x)\right) \triangle f(x)^t =$
$= e^t (A(x)A(x)^t)^{-1} A(x) \triangle f(x)^t - e^t (A(x)A(x)^t)^{-1} A(x)A(x)^t (A(x)A(x)^t)^{-1} A(x) \triangle f(x)^t =$
$= e^t (A(x)A(x)^t)^{-1} A(x) \triangle f(x)^t - e^t (A(x)A(x)^t)^{-1} A(x) \triangle f(x)^t = 0$,

Из чего следует, что $d_1(x)$ и $d_2(x)$ ортогональны, а следовательно, $\|d_1(x) + \alpha d_2(x)\|^2 = \|d_1(x)\|^2 + \alpha^2 \|d_2(x)\|^2, \forall \alpha \in \mathbb{R}$

Т.е. $\|d(x)\|^2 = \|d_1(x) + \alpha(x) d_2(x)\|^2 = \|d_1(x)\|^2 + \alpha(x)^2 \|d_2(x)\|^2 \ge \|d_1(x)\|^2$ **(5)**

С другой, стороны, так как $\alpha(x) \le \dfrac{\|d_1(x)\|}{\|d_2(x)\|}$, то

$$\|d(x)\|^2 = \|d_1(x)\|^2 + \alpha(x)^2 \|d_2(x)\|^2 \le \|d_1(x)\|^2 + \frac{\|d_1(x)\|^2}{\|d_2(x)\|} \|d_2(x)\|^2 \le 2\|d_1(x)\|^2 \quad \textbf{(6)}$$

Из (4), (5) и (6) следует, что пункт 1. утверждения 6 доказан.

Так как $\langle \triangle f(x), d(x) \rangle > 0$, то из утверждения 1 следует, что для достаточно малых $t_0 > 0$, будет выполнено

$$f(x) - f(x - t_0 d(x)) \ge \frac{1}{2} t_0 \langle \Delta f(x), d(x)\rangle \ge \frac{1}{4} t_0 \|d_1(x)\|^2, \quad \textbf{(7)}$$

Из определения $d_1(x)$ и $d_2(x)$ следует, что $\langle \triangle g_i(x), d_1(x) \rangle = 0, \langle \triangle g_i(x), d_2(x) \rangle = 1, \forall i \in I(x)$, так как $\triangle g_i(x)$ - строки матрицы A(x) для i из I(x).

Так как для всех активных индексов $i \in I(x)$ $g_i(x) \le 0$ и

$$\langle \Delta g_i(x), d(x)\rangle = \underbrace{\langle \Delta g_i(x), d_1(x)\rangle}_{=0} + \alpha(x) \underbrace{\langle \Delta g_i(x), d_2(x)\rangle}_{=1} = \alpha(x) > 0,$$

то условия утверждения 2 выполнены и, следовательно, для достаточно малых $t_i > 0, i \in I(x)$ справедливо

$$g_i(x - t_i d(x)) < 0, i \in I(x) \quad \textbf{(8)}$$

Для множества неактивных индексов $J(x) := \{1,...,m\} \setminus I(x) = \{i \in \{1,...,m\} | g_i(x) < -\varepsilon\}$ на основании утверждения 3 найдутся достаточно малые $t_i > 0, i \in J(x)$, такие что

$$g_i(x - t_i d(x)) < 0, i \in J(x) \tag{9}$$

Таким образом, последовательно деля единицу пополам, мы через конечное число шагов придем на основании первых частей утверждений 1, 2 и 3 к такой величине шага $t(x) > 0$, что условия (7), (8) и (9) будут выполнены. Возможны два варианта:

- Эти условия выполняются для первоначального значения $t(x) = 1$.
- Для $t(x) = 1$ эти условия не выполняются, но выполняются для максимального $1 > t(x) > 0$, полученного путем последовательного делением единицы на два. В этом случае для $2t(x)$ хотя бы одно из условий (7), (8) или (9) не выполняются. Из вторых частей утверждений 1, 2 и 3 следует, что для шага $t(x)$ должно выполняться одно из неравенств:

$$t(x) \geq \frac{\langle \Delta f(x), d(x) \rangle}{2L \|d(x)\|^2} \tag{10}$$

$$t(x) \geq \frac{\langle \Delta g_i(x), d(x) \rangle}{L \|d(x)\|^2}, i \in I(x) \tag{11}$$

$$t(x) \geq \frac{\langle \Delta g_i(x), d(x) \rangle + \sqrt{2L\varepsilon \|d(x)\|^2 + \langle \Delta g_i(x), d(x) \rangle^2}}{2\|d(x)\|^2 L}, i \in J(x) \tag{12}$$

Так как $\langle \Delta f(x), d(x) \rangle \geq \frac{1}{2}\|d_1(x)\|$ (4),а также $\|d(x)\|^2 \leq 2\|d_1(x)\|^2$ (6), то из (10) следует

$$t(x) \geq \min\left\{1, \frac{1}{8L}\right\} \tag{13}$$

Так как, G –компактное множество, функции $\vartriangle f$, $\vartriangle g_i$ непрерывны, то существуют положительные константы, такие что

$$\|d_1(x)\| = \left\|\left(E_n - A(x)^t (A(x)A(x)^t)^{-1} A(x)\right) \vartriangle f(x)^t\right\| \leq C_0, \forall x \in G \tag{14}$$

$$\|d_2(x)\| = \left\|A(x)^t (A(x)A(x)^t)^{-1} e\right\| \leq C_1, \forall x \in G \tag{15}$$

$$\left|\langle \Delta f(x), d_2(x) \rangle\right| \leq C_2, \forall x \in G \tag{16}$$

Таким образом, из (1), (15) и (16) следут

$$\alpha(x) \geq \min\left\{1, \frac{\|d_1(x)\|}{C_1}, \begin{cases} \frac{\|d_1(x)\|^2}{2C_2}, \langle \Delta f(x), d_2(x) \rangle < 0 \\ 1, \langle \Delta f(x), d_2(x) \rangle \geq 0 \end{cases}\right\} \geq \min\left\{1, \frac{\|d_1(x)\|}{C_1}, \frac{\|d_1(x)\|^2}{2C_2}\right\} \tag{17}$$

С учетом (16) и (17), из неравенства (11) можно получить следующую оценку:

$$t(x) \geq \frac{\langle \Delta g_i(x), d(x) \rangle}{L\|d(x)\|^2} \geq \frac{\langle \Delta g_i(x), d(x) \rangle}{2L\|d_1(x)\|^2} = \frac{\alpha(x)}{2L\|d_1(x)\|^2} \geq \frac{1}{2L\|d_1(x)\|^2} \min\left\{1, \frac{\|d_1(x)\|}{C_1}, \frac{\|d_1(x)\|^2}{2C_2}\right\} =$$

$$= \frac{1}{2L}\min\left\{\frac{1}{\|d_1(x)\|^2}, \frac{1}{C_1\|d_1(x)\|}, \frac{1}{2C_2}\right\} \geq \frac{1}{2L}\min\left\{\frac{1}{C_0^2}, \frac{1}{C_1 C_0}, \frac{1}{2C_2}\right\}$$

Т.е.

$$t(x) \geq \frac{1}{2L}\min\left\{\frac{1}{C_0^2}, \frac{1}{C_1 C_0}, \frac{1}{2C_2}\right\} \tag{18}$$

С учетом (14) из (12) можно получить оценку

$$t_i \geq \min\left\{1, \frac{\langle \Delta g_i(x), d(x)\rangle + \sqrt{2L\varepsilon\|d(x)\|^2 + \langle \Delta g_i(x), d(x)\rangle^2}}{2C_0{}^2 L}\right\} \qquad \textbf{(19)}$$

Таким образом, для шага t(x) мы получаем: либо t(x)=1 либо справедливо одно из неравенств (13), (18) или (19), а следовательно

$$t(x) \geq \min\left\{1, \frac{1}{8L}, \frac{1}{2LC_0{}^2}, \frac{1}{2LC_1C_0}, \frac{1}{4LC_2}, \frac{\langle \Delta g_i(x), d(x)\rangle + \sqrt{2L\varepsilon\|d(x)\|^2 + \langle \Delta g_i(x), d(x)\rangle^2}}{2C_0{}^2 L}\right\} =$$

$$= \min\left\{K_1, \frac{\langle \Delta g_i(x), d(x)\rangle + \sqrt{K_3\varepsilon\|d(x)\|^2 + \langle \Delta g_i(x), d(x)\rangle^2}}{K_2}\right\}, i \in J(x)$$

где $K_1 := \min\left\{1, \frac{1}{8L}, \frac{1}{2LC_0{}^2}, \frac{1}{2LC_1C_0}, \frac{1}{4LC_2}\right\}$, $K_2 := 2C_0{}^2 L$, $K_3 := 2L$.

Таким образом, для $t(x)$ справедливо:

$$f(x) - f(x - t(x)d(x)) \geq \frac{1}{4}t(x)\|d_1(x)\|^2$$ (пункт 2. утверждения 6)

$$g_i(x - t(x)d(x)) < 0, i \in \{1,\ldots,m\}$$ (пункт 3. утверждения 6)

$$t(x) \geq \min\left\{K_1, \frac{\langle \Delta g_i(x), d(x)\rangle + \sqrt{K_3\varepsilon\|d(x)\|^2 + \langle \Delta g_i(x), d(x)\rangle^2}}{K_2}\right\}, i \in J(x)$$ (т.е. пункт 4. утверждения 6).

□

Утверждение 7

Пусть $x_0 \in G$, итеративно определим следующие величины для k=0,1,2…:

1. $A(x_k)$-матрица ε-активных ограничений в точке x_k,
2. $d_1(x_k)$ – проекция градиента функции f в точке x на ядро $A(x_k)$, т.е.
   $d_1(x_k) := \left(E_n - A(x_k)^t(A(x_k)A(x_k)^t)^{-1}A(x_k)\right)\Delta f(x_k)^t$. Если $d_1(x_k) = 0$, то останавливаемся.
3. Если $d_1(x_k) \neq 0$, то определяем $d_2(x_k) := A(x_k)^t(A(x_k)A(x_k)^t)^{-1}e$, где $e = (1,\ldots,1)^t$,
   $$\alpha(x_k) := \min\left\{1, \frac{\|d_1(x_k)\|}{\|d_2(x_k)\|}, \begin{cases} -\dfrac{\|d_1(x_k)\|^2}{2\langle \Delta f(x_k), d_2(x_k)\rangle}, \langle \Delta f(x_k), d_2(x_k)\rangle < 0 \\ 1, \langle \Delta f(x_k), d_2(x_k)\rangle \geq 0 \end{cases}\right\} > 0,$$
   $d(x_k) := d_1(x_k) + \alpha(x_k)d_2(x_k)$
4. Из утверждения 6 следует, что существует
   $$t(x_k) = \max\{\left(\frac{1}{2}\right)^s \mid s \in \{0,1,2\ldots\}, f(x_k) - f(x_k - \left(\frac{1}{2}\right)^s d(x_k)) \geq \frac{1}{2}\left(\frac{1}{2}\right)^s \langle \Delta f(x_k), d(x_k)\rangle,$$
   ,
   $$g_i(x_k - \left(\frac{1}{2}\right)^s d(x_k)) < 0 \forall i \in \{1,\ldots,m\}\}$$
   при этом справедлива оценка

$$t(x_k) \geq \min\left\{ K_1, \frac{\langle \Delta g_i(x_k), d(x_k)\rangle + \sqrt{K_3\varepsilon \left\| d(x_k)\right\|^2 + \langle \Delta g_i(x_k), d(x_k)\rangle^2}}{K_2} \right\}, i \in J(x_k)$$

5. Определяем $x_{k+1} := x_k - t(x_k)d(x_k)$

6. k:=k+1

Тогда либо алгоритм останавливается для определенного k, для которого $d_1(x_k) = 0$, либо существует сходящаяся к $\overline{x} \in G$ подпоследовательность последовательности x_k, для предела которой справедливо $d_1\left(\overline{x}\right) = 0$.

Доказательство

Пусть для всех k справедливо $d_1(x_k) \neq 0$

Так как по 4. пункту алгоритма по построению $t(x_k)$ следует $g_i(x_{k+1}) = g_i(x_k - t(x_k)d(x_k)) < 0$, то все элементы последовательности x_k находятся в допустимой области G.

Также из пункта 4. и утверждения 6 следует, что
$f(x_k) - f(x_{k+1}) \geq \frac{1}{2} t(x_k)\langle \Delta f(x_k), d(x_k) \geq \frac{1}{4} t(x_k)\left\| d_1(x_k)\right\|^2 > 0$, т.е. последовательность $f(x_k)$ строго монотонно убывающая. Эта последовательность ограничена снизу, так как G компактно, а f непрерывна, следовательно, $f(x_k)$ сходится. В частности из этого следует, что $f(x_k)$ - последовательность Коши (фундаментальная последовательность) и поэтому
$0 = \lim\limits_{k\to\infty}\left(f(x_k) - f(x_{k+1})\right) \geq \frac{1}{4} t(x_k)\left\| d_1(x_k)\right\|^2 > 0$, т.е.

$$\lim_{k\to\infty} t(x_k)\left\| d_1(x_k)\right\|^2 = 0 \qquad \textbf{(1)}$$

Определим $\beta := \inf\limits_k \left\| d_1(x_k)\right\|$

Предположим, что $\beta > 0$, тогда из (1) следует:

$$0 = \lim_{k\to\infty} t(x_k)\left\| d_1(x_k)\right\|^2 \underbrace{\geq}_{\beta \leq \left\| d_1(x_k)\right\|} \beta^2 \limsup_{k\to\infty} t(x_k) \underbrace{\Rightarrow}_{\beta>0} \limsup_{k\to\infty} t(x_k) = 0 \underbrace{\Rightarrow}_{t(x_k)\geq 0} \lim_{k\to\infty} t(x_k) = 0 \qquad \textbf{(2)}$$

С другой стороны из пункта 4. и утверждения 6 следует, что

$$t(x_k) \geq \min\left\{ K_1, \frac{\langle \Delta g_i(x_k), d(x_k)\rangle + \sqrt{K_3\varepsilon \left\| d(x_k)\right\|^2 + \langle \Delta g_i(x_k), d(x_k)\rangle^2}}{K_2} \right\}, i \in J(x_k) \qquad \textbf{(3)}$$

Из (2) и (3) следует, что

$$\lim_{k\to\infty}\left(\min\left\{ \langle \Delta g_i(x_k), d(x_k)\rangle + \sqrt{K_3\varepsilon \left\| d(x_k)\right\|^2 + \langle \Delta g_i(x_k), d(x_k)\rangle^2} \,\middle|\, i \in J(x_k) \right\}\right) = 0 \qquad \textbf{(4)}$$

Так как число индексов конечно, а число элементов последовательности $\min\left\{ \langle \Delta g_i(x_k), d(x_k)\rangle + \sqrt{K_3\varepsilon \left\| d(x_k)\right\|^2 + \langle \Delta g_i(x_k), d(x_k)\rangle^2} \,\middle|\, i \in J(x_k) \right\}$ бесконечно, то существует хотя бы один индекс $i_0 \in \{1,...,m\}$ на котором $\langle \Delta g_i(x_k), d(x_k)\rangle + \sqrt{K_3\varepsilon \left\| d(x_k)\right\|^2 + \langle \Delta g_i(x_k), d(x_k)\rangle^2}$ принимает минимум бесконечное число раз, т.е. существует подпоследовательность индексов $\left(k_l\right)_{l\in\mathbb{N}}$, такая что

$$\langle \Delta g_{i_0}(x_{k_l}), d(x_{k_l})\rangle + \sqrt{K_3\varepsilon \left\| d(x_{k_l})\right\|^2 + \langle \Delta g_{i_0}(x_{k_l}), d(x_{k_l})\rangle^2} =$$

$$= \min\left\{ \langle \Delta g_i(x_{k_l}), d(x_{k_l})\rangle + \sqrt{K_3\varepsilon \left\| d(x_{k_l})\right\|^2 + \langle \Delta g_i(x_{k_l}), d(x_{k_l})\rangle^2} \,\middle|\, i \in J(x_{k_l}) \right\}$$

для всех $l \in \mathbb{N}$

С учетом этого получаем из (4):

$$\lim_{l\to\infty}\left(\langle \Delta g_{i_0}(x_{k_l}), d(x_{k_l})\rangle + \sqrt{K_3\varepsilon\left\|d(x_{k_l})\right\|^2 + \langle \Delta g_{i_0}(x_{k_l}), d(x_{k_l})\rangle^2}\right) = 0 \qquad \textbf{(5)}$$

Из (5) следует, что

$\lim\limits_{l\to\infty}\left\|d(x_{k_l})\right\| = 0 \underbrace{\Rightarrow}_{d_1(x)\le d(x)} \lim\limits_{l\to\infty}\left\|d_1(x_{k_l})\right\| = 0$, т.е. $\beta = \inf\limits_k \left\|d_1(x_k)\right\| \le \lim\limits_{l\to\infty}\left\|d_1(x_{k_l})\right\| = 0$, таким образом, мы пришли к противоречию с предположением, что $\beta > 0$.

Мы получили: $\inf\limits_k \left\|d_1(x_k)\right\| = 0$, а так как $d_1(x_k) \ne 0$, то $\liminf\limits_{k\to\infty}\left\|d_1(x_k)\right\| = 0$, а следовательно, существует подпоследовательность $\left(x_{k_l}\right)_{l\in\mathbb{N}}$, такая что $\lim\limits_{l\to\infty}\left\|d_1(x_{k_l})\right\| = 0$ или $\lim\limits_{l\to\infty} d_1(x_{k_l}) = 0$ **(6)**

Последовательность $\left(x_{k_l}\right)_{l\in\mathbb{N}}$ бесконечна, а набор возможных активных индексов конечен, следовательно, существует набор активных индексов $\overline{I} \subset \{1,...,m\}$, встречающийся на бесконечном количестве элементов последовательности, т.е. $\overline{I} = I(x_{k_l})$ для бесконечного числа $l \in \mathbb{N}$. Следовательно, существует подпоследовательность $\left(x_{k_{l_p}}\right)_{p\in\mathbb{N}}$, такая что множество активных индексов для нее постоянна $\overline{I} = I\left(x_{k_{l_p}}\right), \forall p \in \mathbb{N}$.

Так как G компактное множество, $x_{k_{l_p}} \in G$, то существует сходящаяся к $\overline{x} \in G$ подпоследовательность $\left(x_{k_{l_{p_s}}}\right)_S$.

Учитывая непрерывную дифференцируемость функций $f, g_1, ..., g_m$ и то, что матрицы $A\left(x_{k_{l_p}}\right) = A_{\overline{I}}$ состоят из одних и тех же строк получаем:

$$\begin{aligned}
0 &= \lim_{s\to\infty} d_1(x_{k_{l_{p_s}}}) = \lim_{s\to\infty}\left(E_n - A^t(x_{k_{l_{p_s}}})(A(x_{k_{l_{p_s}}})A^t(x_{k_{l_{p_s}}}))^{-1}A(x_{k_{l_{p_s}}})\right)\triangle f(x_{k_{l_{p_s}}})^t = \\
&\underbrace{=}_{I(x_{k_{l_{p_s}}})=\overline{I}} \lim_{s\to\infty}\left(E_n - A_{\overline{I}}^{\,t}(x_{k_{l_{p_s}}})(A_{\overline{I}}(x_{k_{l_{p_s}}})A_{\overline{I}}^{\,t}(x_{k_{l_{p_s}}}))^{-1}A_{\overline{I}}(x_{k_{l_{p_s}}})\right)\triangle f(x_{k_{l_{p_s}}})^t = \\
&= \left(E_n - A_{\overline{I}}^{\,t}(\lim_{s\to\infty} x_{k_{l_{p_s}}})(A_{\overline{I}}(\lim_{s\to\infty} x_{k_{l_{p_s}}})A_{\overline{I}}^{\,t}(\lim_{s\to\infty} x_{k_{l_{p_s}}}))^{-1}A_{\overline{I}}(\lim_{s\to\infty} x_{k_{l_{p_s}}})\right)\triangle f(\lim_{s\to\infty} x_{k_{l_{p_s}}})^t = \\
&= \left(E_n - A_{\overline{I}}^{\,t}(\overline{x})(A_{\overline{I}}(\overline{x})A_{\overline{I}}^{\,t}(\overline{x}))^{-1}A_{\overline{I}}(\overline{x})\right)\triangle f(\overline{x})^t
\end{aligned} \qquad \textbf{(7)}$$

Если $\overline{I} = I(\overline{x})$, т.е. множество $I(\overline{x})$ активных индексов в точке $\overline{x}$ совпадает с множеством активных индексов $\overline{I}$ в точках $x_{k_{l_{p_s}}}$ из окрестности $\overline{x}$, то $A_{\overline{I}}(\overline{x}) = A(\overline{x})$, и, очевидно,

$$d_1(\overline{x}) = \left(E_n - A^t(\overline{x})(A(\overline{x})A^t(\overline{x}))^{-1}A(\overline{x})\right)\triangle f(\overline{x})^t = \left(E_n - A_{\overline{I}}^{\,t}(\overline{x})(A_{\overline{I}}(\overline{x})A_{\overline{I}}^{\,t}(\overline{x}))^{-1}A_{\overline{I}}(\overline{x})\right)\triangle f(\overline{x})^t,$$

т.е. с учетом (7) $d_1(\overline{x}) = \lim\limits_{s\to\infty} d_1(x_{k_{l_{p_s}}}) = 0$ **(8)**

Предположим, что $\overline{I} \ne I(\overline{x})$. Но множество $I(\overline{x})$ не может быть меньше чем $\overline{I}$ так как если $i \in \overline{I}$, то $g_i(x_{k_{l_{p_s}}}) \in [-\varepsilon, 0], \forall s \in \mathbb{N}$, значит в пределе $g_i(\overline{x}) \le\in [-\varepsilon, 0]$, а следовательно, $i \in I(\overline{x})$, т.е. $\overline{I} \subset I(\overline{x})$. Таким образом, если $\overline{I} \ne I(\overline{x})$, то $\exists i \in I(\overline{x}) \setminus \overline{I}$. В этом случае матрица $A(\overline{x})$ содержит больше (линейно независимых) строк, чем $A_{\overline{I}}(\overline{x})$, поэтому проектор на ядро $A(\overline{x})$ равен проектору на ядро $A_{I(\overline{x})\setminus\overline{I}}(\overline{x})$ умноженному на проектор на ядро $A_{\overline{I}}(\overline{x})$, т.е.

$$\begin{aligned}
&E_n - A^t(\overline{x})(A(\overline{x})A^t(\overline{x}))^{-1}A(\overline{x}) = \\
&= \left(E_n - A_{I(x)\setminus\overline{I}}^{\,t}(\overline{x})(A_{I(x)\setminus\overline{I}}(\overline{x})A_{I(x)\setminus\overline{I}}^{\,t}(\overline{x}))^{-1}A_{I(x)\setminus\overline{I}}(\overline{x})\right)\left(E_n - A_{\overline{I}}^{\,t}(\overline{x})(A_{\overline{I}}(\overline{x})A_{\overline{I}}^{\,t}(\overline{x}))^{-1}A_{\overline{I}}(\overline{x})\right)
\end{aligned}$$

Следовательно

$$\left\|\left(E_n - A^t(\overline{x})(A(\overline{x})A^t(\overline{x}))^{-1}A(\overline{x})\right)\vartriangle f(\overline{x})\right\| =$$

$$= \left\|\left(E_n - A_{I(x)\setminus\overline{I}}{}^t(\overline{x})(A_{I(x)\setminus\overline{I}}(\overline{x})A_{I(x)\setminus\overline{I}}{}^t(\overline{x}))^{-1}A_{I(x)\setminus\overline{I}}(\overline{x})\right)\left(E_n - A_{\overline{I}}{}^t(\overline{x})(A_{\overline{I}}(\overline{x})A_{\overline{I}}{}^t(\overline{x}))^{-1}A_{\overline{I}}(\overline{x})\right)\vartriangle f(\overline{x})\right\|$$

Но норма проекции вектора не может быть больше нормы самого вектора, поэтому

$$\left\|\left(E_n - A^t(\overline{x})(A(\overline{x})A^t(\overline{x}))^{-1}A(\overline{x})\right)\vartriangle f(\overline{x})\right\| =$$

$$= \left\|\left(E_n - A_{I(x)\setminus\overline{I}}{}^t(\overline{x})(A_{I(x)\setminus\overline{I}}(\overline{x})A_{I(x)\setminus\overline{I}}{}^t(\overline{x}))^{-1}A_{I(x)\setminus\overline{I}}(\overline{x})\right)\left(E_n - A_{\overline{I}}{}^t(\overline{x})(A_{\overline{I}}(\overline{x})A_{\overline{I}}{}^t(\overline{x}))^{-1}A_{\overline{I}}(\overline{x})\right)\vartriangle f(\overline{x})\right\| \le$$

$$\le \left\|\left(E_n - A_{\overline{I}}{}^t(\overline{x})(A_{\overline{I}}(\overline{x})A_{\overline{I}}{}^t(\overline{x}))^{-1}A_{\overline{I}}(\overline{x})\right)\vartriangle f(\overline{x})\right\| \underbrace{=}_{(7)} 0$$

Таким образом

$$d_1(\overline{x}) = \left\|\left(E_n - A^t(\overline{x})(A(\overline{x})A^t(\overline{x}))^{-1}A(\overline{x})\right)\vartriangle f(\overline{x})\right\| = 0 \qquad \textbf{(9)}$$

Из (8) и (9) мы получаем, что в любом случае

$d_1(\overline{x}) = 0$

□

Утверждение 8

Пусть $x \in G$ и проекция градиента на ядро матрицы градиентов активных ограничений равно 0, т.е. $d_1(x) = P_I \vartriangle f(x)^t = \left(E_n - A_I{}^t(A_I A_I{}^t)^{-1}A_I\right)\vartriangle f(x)^t = 0$.

Обозначим $u_I := -\vartriangle f(x)A_I{}^t(A_I A_I{}^t)^{-1}A_I = \left(u_I{}^{(1)},...,u_I{}^{(k)}\right) \in Mat(1;k)$, где $k = |I| = \#\{i \in I\}$.

Если существует $i_0 \in I$ такое, что $u_{i_0} < 0$, то для набора индексов $I' := I \setminus \{i_0\}$, полученного из I путем удаления этого i_0-ого индекса справедливо:

1. $d_1'(x) := P_{I'} \vartriangle f(x)^t = \left(E_n - A_{I'}{}^t(A_{I'}A_{I'}{}^t)^{-1}A_{I'}\right)\vartriangle f(x)^t \ne 0$ - проекция градиента на ядро матрицы градиентов активных ограничений с удалением i_0-ого индекса отлична от нуля.
2. $\left\langle d_1'(x), a_{i_0}{}^t\right\rangle = \left\langle P_{I'} \vartriangle f(x)^t, a_{i_0}{}^t\right\rangle = -u_{i_0}\left\|P_{I'}a_{i_0}{}^t\right\|^2 > 0$, где a_{i_0} - i_0-ая строка матрицы градиентов ограничений в точке x.

Доказательство

1. Из $d_1(x) = 0$ имеем:

$$0 = \vartriangle f(x) - \vartriangle f(x)A_I{}^t(A_I A_I{}^t)^{-1}A_I = \vartriangle f(x) + \sum_{i\in I'} u_i \vartriangle g_i(x) + u_{i_0}\vartriangle g_{i_0}(x) \qquad \textbf{(1)}$$

Предположим, что $d_1'(x) = 0$, тогда для $u_I' := -\vartriangle f(x)A_{I'}{}^t(A_{I'}A_{I'}{}^t)^{-1}A_{I'}$ получаем:

$$0 = \vartriangle f(x) - \vartriangle f(x)A_{I'}{}^t(A_{I'}A_{I'}{}^t)^{-1}A_{I'} = \vartriangle f(x) + \sum_{i\in I'} u_i' \vartriangle g_i(x) \qquad \textbf{(2)}$$

Вычитая (2) из (1) получаем:

$$0 = \sum_{i\in I'}\left(u_i - u_i'\right)\vartriangle g_i(x) + u_{i_0}\vartriangle g_{i_0}(x) \qquad \textbf{(3)}$$

Из линейной независимости градиентов активных ограничений, т.е. $\left\{\vartriangle g_i(x) \middle| i \in I\right\}$ из (3) следует, что $u_{i_0} = 0$, что противоречит условию. Т.е. мы доказали, что $d_1'(x) \ne 0$.

2. Сначала докажем, что вектор $P_{I'}a_{i_0}{}^t \ne 0$. Действительно, предположим, что $P_{I'}a_{i_0}{}^t = 0$, следовательно $\left(E_n - A_{I'}{}^t(A_{I'}A_{I'}{}^t)^{-1}A_{I'}\right)a_{i_0}{}^t = 0 \Rightarrow a_{i_0} = a_{i_0}A_{I'}{}^t(A_{I'}A_{I'}{}^t)^{-1}A_{I'}$. Т.е. a_{i_0} представляет собой линейную комбинацию строк из $A_{I'}$, а следовательно строки $\{a_{i_0}\}\bigcup\{a_i \,|\, i \in I'\} = \{a_i \,|\, i \in I\}$ линейно зависимы, что

противоречит условию линейной независимости градиентов активных ограничений. Таким образом, мы доказали, что $P_{I'}a_{i_0}{}^t \neq 0$. **(4)**

Далее из утверждения 5 следует:

$$a_{i_0}P_{I'}a_{i_0}{}^t = \left\langle a_{i_0}{}^t, P_{I'}a_{i_0}{}^t \right\rangle = \left\langle a_{i_0}{}^t, P_{I'}P_{I'}a_{i_0}{}^t \right\rangle = \left\langle P_{I'}{}^t a_{i_0}{}^t, P_{I'}a_{i_0}{}^t \right\rangle =$$
$$= \left\langle P_{I'}a_{i_0}{}^t, P_{I'}a_{i_0}{}^t \right\rangle = \left\| P_{I'}a_{i_0}{}^t \right\|^2 > 0 \quad \textbf{(5)}$$

Из $d_1(x) = 0$ имеем:

$0 = \vartriangle f(x) + \sum_{i \in I'} u_i a_i + u_{i_0}a_{i_0}$, умножаем на $P_{I'}a_{i_0}{}^t$:

$$0 = \vartriangle f(x)P_{I'}a_{i_0}{}^t + \sum_{i \in I'} u_i a_i P_{I'}a_{i_0}{}^t + u_{i_0}a_{i_0}P_{I'}a_{i_0}{}^t \quad \textbf{(6)}$$

Так как $P_{I'}$ - проектор на ядро $A_{I'}$, то $a_i P_{I'} = 0, \forall i \in I'$, следовательно из (6) с учетом (5) получаем:

$$0 = \vartriangle f(x)P_{I'}a_{i_0}{}^t + u_{i_0}\left\| P_{I'}a_{i_0}{}^t \right\|^2 \Rightarrow \left\langle P_{I'} \vartriangle f(x)^t, a_{i_0}{}^t \right\rangle = \vartriangle f(x)P_{I'}a_{i_0}{}^t = -\underbrace{u_{i_0}}_{<0}\underbrace{\left\| P_{I'}a_{i_0}{}^t \right\|^2}_{>0} > 0, \quad \textbf{(7)}$$

что доказывает последнее утверждение.

□

Утверждение 9

Пусть $x \in G = \left\{ x \in \mathbb{R}^n \mid g_1(x), \dots, g_m(x) \le 0 \right\}$,

$A(x)$ -матрица градиентов ε -активных ограничений в точке x, т.е. $I(x) = \left\{ i_k \in \{1, \dots, m\} \mid g_{i_k}(x) \in [-\varepsilon, 0] \right\}$,

$d_1(x)$ – равная нулю проекция градиента функции f в точке x на ядро A(x), т.е.

$d_1(x) = P_I \vartriangle f(x)^t = \left(E_n - A_I{}^t (A_I A_I{}^t)^{-1} A_I \right) \vartriangle f(x)^t = 0$.

Обозначим $u_I := -\vartriangle f(x) A_I{}^t (A_I A_I{}^t)^{-1} A_I = \left(u_I{}^{(1)}, \dots, u_I{}^{(k)} \right) \in Mat(1;k)$, где $k = |I| = \#\{ i \in I \}$.

Пусть существует $i_0 \in I$ такое, что $u_{i_0} < 0$. Определим набор индексов $I' := I \setminus \{ i_0 \}$, полученный из I путем удаления этого i_0 -ого индекса.

Определим

$d_1' := P_{I'} \vartriangle f(x)^t = \left(E_n - A_{I'}{}^t (A_{I'} A_{I'}{}^t)^{-1} A_{I'} \right) \vartriangle f(x)^t$

$d_2' := A_{I'}{}^t (A_{I'} A_{I'}{}^t)^{-1} e$, где $e = (1, \dots, 1)^t$

$$\alpha' := \min \left\{ 1, \frac{\|d_1'\|}{\|d_2'\|}, \begin{cases} -\dfrac{\|d_1'\|^2}{2\langle \Delta f(x), d_2' \rangle}, \langle \Delta f(x), d_2' \rangle < 0 \\ 1, \langle \Delta f(x), d_2' \rangle \ge 0 \end{cases}, \begin{cases} -\dfrac{\langle \Delta g_{i_0}(x), d_1' \rangle}{2\langle \Delta g_{i_0}(x), d_2' \rangle}, \langle \Delta g_{i_0}(x), d_2' \rangle < 0 \\ 1, \langle \Delta g_{i_0}(x), d_2' \rangle \ge 0 \end{cases} \right\} > 0$$

Отметим, что так как по утверждению 8 $\langle \Delta g_{i_0}(x), d_1' \rangle > 0$, то $\alpha' > 0$

$d' := d_1' + \alpha' d_2'$

Тогда найдется такое положительное число t, что будет справедливы следующие неравенства:

$$f(x) - f(x - td') \ge \frac{1}{2} t \langle \Delta f(x), d' \rangle \ge \frac{1}{4} t \|d_1'\|^2 > 0$$

$$g_i(x - td') < 0, i \in \{1, \dots, m\}$$

$$t \ge \min \left\{ K_0, K_1 |u_{i_0}| \left\| P_{I'} \Delta g_{i_0}(x)^t \right\|^2, \frac{\langle \Delta g_i(x), d' \rangle + \sqrt{K_3 \varepsilon \|d_1'\|^2 + \langle \Delta g_i(x), d' \rangle^2}}{K_2} \right\},$$

где K_0, K_1, K_2, K_3>0, константы, а $i \in J(x) = \{1,...,m\} \setminus I(x) = \{i \in \{1,...,m\} | g_i(x) < -\varepsilon\}$

Доказательство

Из утверждения 8 следует, что $d_1' \neq 0$.

Из ортогональности d_1' и d_2' следует, что $\|d'\| \geq \|d_1'\|$.

Из определения $\alpha' \leq \frac{\|d_1'\|}{\|d_2'\|}$ следует, что $\|d'\|^2 \leq 2\|d_1'\|^2$

Далее, так как $\langle \vartriangle f(x), d' \rangle = \langle \vartriangle f(x), d_1' \rangle + \alpha' \langle \vartriangle f(x), d_2' \rangle = \|d_1'\|^2 + \alpha' \langle \vartriangle f(x), d_2' \rangle \geq \frac{1}{2}\|d_1'\|^2 > 0$, то из утверждения 1 следует, что для достаточно малых $t_0 > 0$, будет выполнено

$$f(x) - f(x - t_0 d') \geq \frac{1}{2} t_0 \langle \Delta f(x), d' \rangle \geq \frac{1}{4} t_0 \|d_1'\|^2 \quad \textbf{(1)}$$

Из определения d_1' и d_2' следует, что $\langle \vartriangle g_i(x), d_1' \rangle = 0, \langle \vartriangle g_i(x), d_2' \rangle = 1, \forall i \in I'(x)$.

Так как для всех индексов $i \in I'$ $g_i(x) \leq 0$ и $\langle \Delta g_i(x), d' \rangle = \underbrace{\langle \Delta g_i(x), d_1' \rangle}_{=0} + \alpha' \underbrace{\langle \Delta g_i(x), d_2' \rangle}_{=1} = \alpha' > 0$,

то условия утверждения 2 выполнены и, следовательно, для достаточно малых $t_i > 0, i \in I'$ справедливо

$$g_i(x - t_i d') < 0, i \in I' \quad \textbf{(2)}$$

Для индекса $i_0 \in I \setminus I'$ справедливо вследствие утверждения 8: $\langle d_1', \vartriangle g_{i_0}(x) \rangle > 0$ **(3)**

Из определения α' следует $\alpha' \leq -\frac{\langle \Delta g_{i_0}(x_k), d_1' \rangle}{2\langle \Delta g_{i_0}(x_k), d_2' \rangle}$ если $\langle \Delta g_{i_0}(x_k), d_2' \rangle < 0$ и следовательно

$$\langle d', \vartriangle g_{i_0}(x) \rangle = \langle d_1', \vartriangle g_{i_0}(x) \rangle + \alpha' \langle d_2', \vartriangle g_{i_0}(x) \rangle \geq \begin{cases} \frac{1}{2} \langle d_1', \vartriangle g_{i_0}(x) \rangle, \langle \Delta g_{i_0}(x_k), d_2' \rangle < 0 \\ \langle d_1', \vartriangle g_{i_0}(x) \rangle, \langle \Delta g_{i_0}(x_k), d_2' \rangle \geq 0 \end{cases}$$

Т.е. $\langle d', \vartriangle g_{i_0}(x) \rangle \geq \frac{1}{2} \langle d_1', \vartriangle g_{i_0}(x) \rangle > 0$ **(4)**

Итак, мы имеем: $g_{i_0}(x) \leq 0$, $\langle d', \vartriangle g_{i_0}(x) \rangle > 0$, т.е. условия утверждения 2 для g_{i_0} тоже выполнены и, следовательно, для достаточно малых $t_{i_0} > 0$ справедливо

$$g_i(x - t_{i_0} d') < 0 \quad \textbf{(5)}$$

Для множества неактивных индексов $J(x) := \{1,...,m\} \setminus I(x) = \{i \in \{1,...,m\} | g_i(x) < -\varepsilon\}$ на основании утверждения 3 найдутся достаточно малые $t_i > 0, i \in J(x)$, такие что

$$g_i(x - t_i d(x)) < 0, i \in J(x) \quad \textbf{(6)}$$

Таким образом, последовательно деля единицу пополам, мы через конечное число шагов придем на основании первых частей утверждений 1, 2 и 3 к такой величине шага $t > 0$, что условия (1), (2), (5) и (6) будут выполнены.

Если для $t=1$ эти условия не выполняются, но выполняются для максимального $1>t>0$, полученного путем последовательного делением единицы на два, то в этом случае для $2t$ хотя бы одно из условий (1), (2), (5) или (6) не выполняются. Из вторых частей утверждений 1, 2 и 3 следует, что для шага t должно выполняться одно из неравенств:

$$t \geq \frac{\langle \Delta f(x), d' \rangle}{2L\|d'\|^2} \geq \frac{\|d_1'\|^2}{4L\|d'\|^2} \geq \frac{\|d_1'\|^2}{8L\|d_1'\|^2} = \frac{1}{8L} \tag{7}$$

$$t \geq \frac{\langle \Delta g_i(x), d' \rangle}{L\|d'\|^2} \geq \frac{\alpha'}{L\|d'\|^2}, i \in I'(x) \tag{8}$$

$$t \geq \frac{\langle \Delta g_{i_0}(x), d' \rangle}{L\|d'\|^2} \geq \frac{\langle \Delta g_{i_0}(x), d_1' \rangle}{2L\|d'(x)\|^2} = \frac{-u_{i_0}\left\|P_{I'}\Delta g_{i_0}(x)^t\right\|^2}{2L\|d'\|^2} \tag{9}$$

$$t \geq \frac{\langle \Delta g_i(x), d' \rangle + \sqrt{2L\varepsilon\|d'\|^2 + \langle \Delta g_i(x), d' \rangle^2}}{2\|d'\|^2 L}, i \in J(x) \tag{10}$$

Так как, G –компактное множество, функции $\vartriangle f$, $\vartriangle g_i$ непрерывны, то существуют положительные константы, такие что

$$\|d_1'\| \leq \max\left\{\left\|\left(E_n - A_I(x)^t(A_I(x)A_I(x)^t)^{-1}A_I(x)\right)\vartriangle f(x)^t\right\| \middle| I \subset \{1,...,m\}\right\} \leq C_0, \forall x \in G \tag{11}$$

$$\|d_2'\| \leq \max\left\{\left\|A_I(x)^t(A_I(x)A_I(x)^t)^{-1}e\right\| \middle| I \subset \{1,...,m\}\right\} \leq C_1, \forall x \in G \tag{12}$$

$$\left|\langle \Delta f(x), d_2' \rangle\right| \leq C_2, \forall x \in G \tag{13}$$

$$\left|\langle \Delta g_{i_0}(x), d_2' \rangle\right| \leq C_3, \forall x \in G \tag{14}$$

Поэтому

$$\alpha' \geq \min\left\{1, \frac{\|d_1'\|}{C_1}, \frac{\|d_1'\|^2}{2C_2}, \frac{\langle \Delta g_{i_0}(x), d_1' \rangle}{2C_3}\right\} = \min\left\{1, \frac{\|d_1'\|}{C_1}, \frac{\|d_1'\|^2}{2C_2}, \frac{|u_{i_0}|\left\|P_{I'}\Delta g_{i_0}(x)^t\right\|^2}{2C_3}\right\} \tag{15}$$

Из (11) и (12) следует:

$$\|d'\| = \|d_1' + \alpha' d_2'\| \leq \underbrace{\|d_1'\|}_{C_0} + \underbrace{|\alpha'|}_{\leq 1}\underbrace{\|d_2\|}_{\leq C_1} \leq C_0 + C_1 \tag{16}$$

Если справедливо (8), то из (15) и (16) следует:

$$\begin{aligned} t &\geq \frac{\alpha'}{L\|d'(x)\|^2} \geq \min\left\{\frac{1}{L\|d'(x)\|^2}, \frac{1}{C_1 L\|d'(x)\|}, \frac{1}{2C_2 L}, \frac{|u_{i_0}|\left\|P_{I'}\Delta g_{i_0}(x)^t\right\|^2}{2C_3 L\|d'(x)\|^2}\right\} \geq \\ &\geq \min\left\{\frac{1}{L(C_0+C_1)^2}, \frac{1}{C_1 L(C_0+C_1)}, \frac{1}{2C_2 L}, \frac{|u_{i_0}|\left\|P_{I'}\Delta g_{i_0}(x)^t\right\|^2}{2C_3 L(C_0+C_1)^2}\right\}, i \in I'(x) \end{aligned} \tag{17}$$

Если справедливо (9), то из (16) следует:

$$t \geq \frac{|u_{i_0}|\left\|P_{I'}\Delta g_{i_0}(x)^t\right\|^2}{2L(C_0+C_1)^2} \tag{18}$$

Если справедливо (10), то из (16) следует:

$$t \geq \frac{\langle \Delta g_i(x), d' \rangle + \sqrt{2L\varepsilon\|d_1'\|^2 + \langle \Delta g_i(x), d' \rangle^2}}{2(C_0+C_1)^2 L}, i \in J(x) \tag{19}$$

Из (7), (17), (18), (19) следует:

$$t \geq \min\left\{K_0, K_1 \left|u_{i_0}\right| \left\|P_{I'} \Delta g_{i_0}(x)^t\right\|^2, \frac{\langle \Delta g_i(x), d' \rangle + \sqrt{K_3 \varepsilon \left\|d_1'\right\|^2 + \langle \Delta g_i(x), d' \rangle^2}}{K_2}\right\}, \quad \textbf{(20)}$$

где $K_0, K_1, K_2, K_3 > 0$, константы, а $i \in J(x) = \{1,\dots,m\} \setminus I(x) = \{i \in \{1,\dots,m\} \mid g_i(x) < -\varepsilon\}$

□

Утверждение 10

Пусть $l := 0$, $x_0 \in G$.

По алгоритму, описанному в утверждении 7 итеративно определим следующие величины для k=0,1,2…:

1. $A(x_k)$-матрица ε-активных ограничений в точке x_k,
2. $d_1(x_k)$ – проекция градиента функции f в точке x на ядро $A(x_k)$, т.е. $d_1(x_k) := \left(E_n - A(x_k)^t (A(x_k)A(x_k)^t)^{-1} A(x_k)\right) \vartriangle f(x_k)^t$. Если $d_1(x_k) = 0$, то переходим к пункту 7..
3. Если $d_1(x_k) \neq 0$, то определяем $d_2(x_k) := A(x_k)^t (A(x_k)A(x_k)^t)^{-1} e$, где $e = (1,\dots,1)^t$,

$$\alpha(x_k) := \min\left\{1, \frac{\|d_1(x_k)\|}{\|d_2(x_k)\|}, \begin{cases} -\dfrac{\|d_1(x_k)\|^2}{2\langle \Delta f(x_k), d_2(x_k)\rangle}, \langle \Delta f(x_k), d_2(x_k)\rangle < 0 \\ 1, \langle \Delta f(x_k), d_2(x_k)\rangle \geq 0 \end{cases}\right\} > 0,$$

$$d(x_k) := d_1(x_k) + \alpha(x_k) d_2(x_k)$$

4. Из утверждения 6 следует, что существует

$$t(x_k) = \max\{\left(\frac{1}{2}\right)^s \mid s \in \{0,1,2\dots\}, f(x_k) - f(x_k - \left(\frac{1}{2}\right)^s d(x_k)) \geq \frac{1}{2}\left(\frac{1}{2}\right)^s \langle \Delta f(x_k), d(x_k)\rangle, ,$$
$$g_i(x_k - \left(\frac{1}{2}\right)^s d(x_k)) < 0 \forall i \in \{1,\dots,m\}\}$$

при этом справедлива оценка

$$t(x_k) \geq \min\left\{K_1, \frac{\langle \Delta g_i(x_k), d(x_k)\rangle + \sqrt{K_3 \varepsilon \|d(x_k)\|^2 + \langle \Delta g_i(x_k), d(x_k)\rangle^2}}{K_2}\right\}, i \in J(x_k)$$

5. Определяем $x_{k+1} := x_k - t(x_k) d(x_k)$
6. k:=k+1, переходим к пункту 1.

Тогда по утверждению 7 либо мы выходим из цикла последовательности действий 1-6 алгоритма с определенным k, для которого $d_1(x_k) = 0$ либо существует сходящаяся к $\overline{x} \in G$ подпоследовательность последовательности x_k, для предела которой справедливо $d_1(\overline{x}) = 0$.

Обозначим полученную в результате последовательности действий 1-6 алгоритма (возможно предельную) точку $\overline{x}_l$ в которой $d_1(\overline{x}_l) = 0$.

7. Обозначим $u_I(\overline{x}_l) := -\vartriangle f(\overline{x}_l) A_I^t (A_I A_I^t)^{-1} A_I = \left(u_I^{(1)}(\overline{x}_l), \dots, u_I^{(k)}(\overline{x}_l)\right) \in Mat(1;k)$, где $k = |I| = \#\{i \in I\}, I = I(x)$ - матрица градиентов ε-активных ограничений в точке $\overline{x}_l$.
8. Если $u_I(\overline{x}_l) \geq 0$, то алгоритм завершается.
9. Если существует $i_0 \in I(x)$ такое, что $u_I^{i_0}(\overline{x}_l) < 0$, то для набора индексов $I' := I \setminus \{i_0\}$, полученный из I путем удаления этого i_0-ого индекса со свойством $u_I^{i_0}(\overline{x}_l) = \min\left\{u_I^i(\overline{x}_l) < 0 \middle| i \in I\right\}$ определим:

$$d_1{}'(\overline{x}_l) := P_{I'} \vartriangle f(\overline{x}_l)^t = \left(E_n - A_{I'}{}^t (A_{I'} A_{I'}{}^t)^{-1} A_{I'}\right) \vartriangle f(\overline{x}_l)^t$$

$d_2{}'(\overline{x}_l) := A_{I'}{}^t (A_{I'} A_{I'}{}^t)^{-1} e$, где $e = (1,....,1)^t$

$$\alpha'(\overline{x}_l) := \min\{1, \frac{\|d_1'(\overline{x}_l)\|}{\|d_2'(\overline{x}_l)\|}, \begin{cases} -\frac{\|d_1'(\overline{x}_l)\|^2}{2\langle \Delta f(\overline{x}_l), d_2'(\overline{x}_l)\rangle}, \langle \Delta f(\overline{x}_l), d_2'(\overline{x}_l)\rangle < 0 \\ 1, \langle \Delta f(\overline{x}_l), d_2'(\overline{x}_l)\rangle \ge 0 \end{cases},$$

$$\begin{cases} -\frac{\langle \Delta g_{i_0}(\overline{x}_l), d_1'(\overline{x}_l)\rangle}{2\langle \Delta g_{i_0}(\overline{x}_l), d_2'(\overline{x}_l)\rangle}, \langle \Delta g_{i_0}(\overline{x}_l), d_2'(\overline{x}_l)\rangle < 0 \\ 1, \langle \Delta g_{i_0}(\overline{x}_l), d_2'(\overline{x}_l)\rangle \ge 0 \end{cases}\}$$

$$d'(\overline{x}_l) := d_1'(\overline{x}_l) + \alpha'(\overline{x}_l) d_2'(\overline{x}_l)$$

По утверждению 9 найдется такое положительное число $t'(\overline{x}_l)$, что будет справедливы следующие неравенства:

$$f(\overline{x}_l) - f(\overline{x}_l - t'(\overline{x}_l) d'(\overline{x}_l)) \ge \frac{1}{2} t'(\overline{x}_l) \langle \Delta f(\overline{x}_l), d'(\overline{x}_l)\rangle \ge \frac{1}{4} t'(\overline{x}_l) \|d_1'(\overline{x}_l)\|^2 > 0 \quad \textbf{(1)}$$

$$g_i(\overline{x}_l - t'(\overline{x}_l) d'(\overline{x}_l)) < 0, i \in \{1,...,m\} \quad \textbf{(2)}$$

$$t(\overline{x}_l) \ge \min\{K_0, K_1 \left|u_I{}^{i_0}(\overline{x}_l)\right| \left\|P_{I'} \Delta g_{i_0}(\overline{x}_l)^t\right\|^2,$$
$$\frac{\langle \Delta g_i(\overline{x}_l), d'(\overline{x}_l)\rangle + \sqrt{K_3 \varepsilon \|d_1'(\overline{x}_l)\|^2 + \langle \Delta g_i(\overline{x}_l), d'(\overline{x}_l)\rangle^2}}{K_2}\}, i \in J(\overline{x}_l) \quad \textbf{(3)}$$

Присваиваем

$l := l + 1$

$x_{k+1} := \overline{x}_l - t'(\overline{x}_l) d'(\overline{x}_l)$, из (2) следует, что $x_{k+1} \in G$.

$k := k + 1$, переходим снова к пункту 1. алгоритма.

По построению алгоритма, если полученная таким образом последовательность $(\overline{x}_l)_l$ конечна, то существует $l \in \mathbb{N}$:

$$\vartriangle f(\overline{x}_l) + \sum_{i \in I(\overline{x}_l)} u_{\overline{I}}{}^i(\overline{x}_l) \vartriangle g_i(\overline{x}_l) = 0$$

$$u_{I(\overline{x}_l)}{}^i(\overline{x}_l) \ge 0, \forall i \in I(\overline{x}_l)$$

$\left|g_i(\overline{x}_l)\right| \le \varepsilon, \forall i \in I(\overline{x}_l)$, где $I(\overline{x}_l)$ - множество активных ограничений в точке $\overline{x}_l$.

Если она бесконечна, то последовательность, сгенерированная в ходе выполнения алгоритма, имеет сходящуюся к точке $\overline{x} \in G$ подпоследовательность, для которой для некого набора индексов $\overline{I} \subset I(\overline{x}) \subset \{1,...,m\}$ выполняются следующие условия:

$$\vartriangle f(\overline{x}) + \sum_{i \in \overline{I}} u_{\overline{I}}{}^i(\overline{x}) \vartriangle g_i(\overline{x}) = 0$$

$$u_{\overline{I}}{}^i(\overline{x}) \ge 0, \forall i \in \overline{I}$$

$$\left|g_i(\overline{x})\right| \le \varepsilon, \forall i \in \overline{I}$$

Доказательство

Рассмотрим случай, когда последовательность $(\overline{x}_l)_l$ для которой $d_1(\overline{x}_l) = 0$ и $\exists i(l) \in I(\overline{x}_l) : u_I{}^{i(l)}(\overline{x}_l) < 0$, бесконечна. Из компактности G и $\overline{x}_{l_s} \in G$ следует, что существует $\overline{x} \in G$ и сходящаяся к ней подпоследовательность $(\overline{x}_{l_s})_s$ т.е.

$$\lim_{s \to \infty} \overline{x}_{l_s} = \overline{x} \quad \textbf{(4)}$$

Последовательность $\left(\overline{x}_{l_s}\right)_{s\in\mathbb{N}}$ бесконечна, а набор возможных активных индексов конечен, следовательно, существует набор активных индексов $\overline{I}\subset\{1,...,m\}$, встречающийся на бесконечном количестве элементов последовательности, т.е. $\overline{I}=I(x_{l_s})$ для бесконечного числа $s\in\mathbb{N}$. Следовательно, существует подпоследовательность $\left(\overline{x}_{l_{s_p}}\right)_{p\in\mathbb{N}}$, такая, что множество активных индексов для нее постоянна $\overline{I}=I\left(\overline{x}_{l_{s_p}}\right),\forall p\in\mathbb{N}$.

Учитывая непрерывную дифференцируемость функций $f,g_1,...,g_m$ и то, что матрицы $A\left(\overline{x}_{l_{s_p}}\right)=A_{\overline{I}}$ состоят из одних и тех же строк, получаем:

$$\begin{aligned}
0&=\lim_{p\to\infty}d_1(\overline{x}_{l_{s_p}})=\lim_{p\to\infty}\left(E_n-A^t(\overline{x}_{l_{s_p}})(A(\overline{x}_{l_{s_p}})A^t(\overline{x}_{l_{s_p}}))^{-1}A(\overline{x}_{l_{s_p}})\right)\vartriangle f(\overline{x}_{l_{s_p}})^t=\\
&\underbrace{=}_{I(\overline{x}_{l_{s_p}})=\overline{I}}\lim_{p\to\infty}\left(E_n-A_{\overline{I}}{}^t(\overline{x}_{l_{s_p}})(A_{\overline{I}}(\overline{x}_{l_{s_p}}A_{\overline{I}}{}^t(\overline{x}_{l_{s_p}}))^{-1}A_{\overline{I}}(\overline{x}_{l_{s_p}})\right)\vartriangle f(\overline{x}_{l_{s_p}})^t=\\
&=\left(E_n-A_{\overline{I}}{}^t(\lim_{p\to\infty}\overline{x}_{l_{s_p}})(A_{\overline{I}}(\lim_{p\to\infty}\overline{x}_{l_{s_p}})A_{\overline{I}}{}^t(\lim_{p\to\infty}\overline{x}_{l_{s_p}}))^{-1}A_{\overline{I}}(\lim_{p\to\infty}\overline{x}_{l_{s_p}})\right)\vartriangle f(\lim_{p\to\infty}\overline{x}_{l_{s_p}})^t=\\
&=\left(E_n-A_{\overline{I}}{}^t(\overline{x})(A_{\overline{I}}(\overline{x})A_{\overline{I}}{}^t(\overline{x}))^{-1}A_{\overline{I}}(\overline{x})\right)\vartriangle f(\overline{x})^t
\end{aligned}\tag{5}$$

$\overline{I}\subset I(\overline{x})$, так как если $i\in\overline{I}$, то $g_i(\overline{x}_{l_{s_p}})\in[-\varepsilon,0],\forall p\in\mathbb{N}$, значит в пределе $g_i(\overline{x})\le\in[-\varepsilon,0]$, а следовательно, $i\in I(\overline{x})$. Множество $I(\overline{x})$, таким образом, отличается от $\overline{I}$ на множество индексов

$$I(\overline{x})\setminus\overline{I}=\left\{i\in\{1,...,m\}\,\middle|\,g_i(\overline{x})=-\varepsilon,\,g_i(\overline{x}_{l_{s_p}})<-\varepsilon,\forall p\in\mathbb{N}\right\}\tag{6}$$

Из (1) следует, что последовательность $f(\overline{x}_l)$ строго монотонно убывает, поэтому сходится, а следовательно,

$$\lim_{l\to\infty}f(\overline{x}_l)=\lim_{p\to\infty}f(\overline{x}_{l_{s_p}})=f(\lim_{p\to\infty}\overline{x}_{l_{s_p}})=f(\overline{x})\tag{7}$$

Предположим, что существует $i_0\in\overline{I}:u_{\overline{I}}^{i_0}(\overline{x})=\left(-\vartriangle f(\overline{x})A_{\overline{I}}(\overline{x})^t(A_{\overline{I}}(\overline{x})A_{\overline{I}}(\overline{x})^t)^{-1}A_{\overline{I}}(\overline{x})\right)_{i_0}<0$. Выберем i_0 такой, что $u_{\overline{I}}^{i_0}(\overline{x})$ принимает самое малое значение среди всех $u_{\overline{I}}^{i}(\overline{x}),i\in\overline{I}$. Тогда из непрерывной дифференцируемости $f,g_1,...,g_m$ и того, что матрицы $A(\overline{x}_{l_{s_p}})=A_{\overline{I}}(\overline{x}_{l_{s_p}})$ состоят из одних и тех же строк, получаем, что

$$u_{\overline{I}}^{i_0}(\overline{x}_{l_{s_p}})=\left(-\vartriangle f(\overline{x}_{l_{s_p}})A_{\overline{I}}(\overline{x}_{l_{s_p}})^t(A_{\overline{I}}(\overline{x}_{l_{s_p}})A_{\overline{I}}(\overline{x}_{l_{s_p}})^t)^{-1}A_{\overline{I}}(\overline{x}_{l_{s_p}})\right)_{i_0}<\frac{1}{2}u_{\overline{I}}^{i_0}(\overline{x})<0,\forall p\ge p_0\tag{8}$$

и $u_{\overline{I}}^{i_0}(\overline{x}_{l_{s_p}})$ будут принимать самые малые значения из всех $u_{\overline{I}}^{i}(\overline{x}_{l_{s_p}}),i\in\overline{I},\forall p\ge p_0$. Тогда на каждом шаге p из множества активных индексов $\overline{I}$ будет удаляться именно индекс i_0, т.е. множество $\overline{I}'(\overline{x}_{l_{s_p}})=\overline{I}\setminus\{i_0\}$ будет, начиная с $p\ge p_0$, постоянно.

Учитывая $d_1(\overline{x}_{l_{s_p}})=0,\forall p\in\mathbb{N}$ получаем из (1), (2), (3), что:

$$f(\overline{x}_{l_{s_p}})-f(\overline{x}_{l_{s_p}+1})\ge\frac{1}{4}t(\overline{x}_{l_{s_p}})\left\|d_1'(\overline{x}_{l_{s_p}})\right\|^2>0,\tag{9}$$

$$\begin{gathered}
t(\overline{x}_{l_{s_p}})\ge\min\{K_0,K_1\left|u_{\overline{I}}^{i_0}(\overline{x}_{l_{s_p}})\right|\left\|P_{\overline{I}}\Delta g_{i_0}(\overline{x}_{l_{s_p}})^t\right\|^2,\\
\frac{\langle\Delta g_i(\overline{x}_{l_{s_p}}),d'(\overline{x}_{l_{s_p}})\rangle+\sqrt{K_3\varepsilon\left\|d_1'(\overline{x}_l)\right\|^2+\langle\Delta g_i(\overline{x}_{l_{s_p}}),d'(\overline{x}_{l_{s_p}})\rangle^2}}{K_2}\}
\end{gathered}\tag{10}$$

Из (9) мы получаем, что последовательность $\left(f(\overline{x}_{l_{s_p}})\right)_{p\in\mathbb{N}}$ строго монотонно убывающая, снизу на компактном множестве ограничена, следовательно, сходящаяся, следовательно, последовательность Коши. Получаем из (9):

$$\lim_{p\to\infty} t(\overline{x}_{l_{s_p}})\left\|d_1'(\overline{x}_{l_{s_p}})\right\|^2 = 0 \qquad \textbf{(11)}$$

Определим $\beta := \inf_k \left\|d_1'(\overline{x}_{l_{s_p}})\right\|$

Предположим, что $\beta > 0$, тогда из (11) следует:

$$0 = \lim_{p\to\infty} t(\overline{x}_{l_{s_p}})\left\|d_1'(\overline{x}_{l_{s_p}})\right\|^2 \underbrace{\geq}_{\beta\leq\left\|d_1'(\overline{x}_{l_{s_p}})\right\|} \beta^2 \limsup_{k\to\infty} t(x_k) \underbrace{\Rightarrow}_{\beta>0} \limsup_{k\to\infty} t(\overline{x}_{l_{s_p}}) = 0 \underbrace{\Rightarrow}_{t(\overline{x}_{l_{s_p}})\geq 0} \lim_{k\to\infty} t(\overline{x}_{l_{s_p}}) = 0 \qquad \textbf{(12)}$$

Но с учетом (10) выражение (12) возможно только если

$$\lim_{p\to\infty} \underbrace{\left|u_{\overline{I}}^{i_0}(\overline{x}_{l_{s_p}})\right|}_{\to u_{\overline{I}}^{i_0}(\overline{x})>0} \left\|P_{\overline{I}'}\Delta g_{i_0}(\overline{x}_{l_{s_p}})^t\right\|^2 = 0 \qquad \textbf{(13)}$$

или

$$\lim_{p\to\infty} \frac{\langle \Delta g_i(\overline{x}_{l_{s_p}}), d'(\overline{x}_{l_{s_p}})\rangle + \sqrt{K_3\varepsilon\left\|d_1'(\overline{x}_{l_{s_p}})\right\|^2 + \langle \Delta g_i(\overline{x}_{l_{s_p}}), d'(\overline{x}_{l_{s_p}})\rangle^2}}{K_2} = 0 \qquad \textbf{(14)}$$

Из (14) следует $\lim_{p\to\infty}\left\|d_1'(\overline{x}_{l_{s_p}})\right\| = 0$, что противоречит предположению $\inf_k \left\|d_1'(\overline{x}_{l_{s_p}})\right\| = \beta > 0$, таким образом, случай (14) исключен.

Если выполняется условие (13), то с учетом (8) получаем , что

$$\lim_{p\to\infty}\left\|P_{\overline{I}'}\Delta g_{i_0}(\overline{x}_{l_{s_p}})^t\right\|^2 = 0 \qquad \textbf{(15)}$$

Учитывая непрерывную дифференцируемость $f, g_1, ..., g_m$ получаем из (4) и (15):

$P_{\overline{I}'}\Delta g_{i_0}(\overline{x})^t = 0 \Rightarrow \left(E_n - A_{\overline{I}'}{}^t(A_{\overline{I}'}A_{\overline{I}'}{}^t)^{-1}A_{\overline{I}'}\right)\Delta g_{i_0}(\overline{x})^t = 0 \Rightarrow \Delta g_{i_0}(\overline{x}) = \Delta g_{i_0}(\overline{x})A_{\overline{I}'}{}^t(A_{\overline{I}'}A_{\overline{I}'}{}^t)^{-1}A_{\overline{I}'}$,

следовательно, строки $\left\{\Delta g_{i_0}(\overline{x}), \Delta g_i(\overline{x}) \middle| i \in \overline{I}'\right\} = \left\{\Delta g_i(\overline{x}) \middle| i \in \overline{I}\right\}$ линейно зависимы, что противоречит условию о линейной независимости градиентов активных ограничений $\left\{\Delta g_i(\overline{x}) \middle| i \in I(\overline{x})\right\}$. Таким образом, случай (13) тоже исключен, поэтому $\beta = \inf_k \left\|d_1'(\overline{x}_{l_{s_p}})\right\| = 0$, переходя к подпоследовательности , получаем:

$\lim_{p\to\infty} d_1'(\overline{x}_{l_{s_p}}) = 0$. Учитывая непрерывную дифференцируемость $f, g_1, ..., g_m$ и того, что

$d_1'(\overline{x}_{l_{s_p}}) = \left(E_n - A_{\overline{I}'}{}^t(\overline{x}_{l_{s_p}})(A_{\overline{I}'}(\overline{x}_{l_{s_p}})A_{\overline{I}'}{}^t(\overline{x}_{l_{s_p}}))^{-1}A_{\overline{I}'}(\overline{x}_{l_{s_p}})\right)\vartriangle f(\overline{x}_{l_{s_p}})^t$, где матрицы $A_{\overline{I}'}$ состоят из одних и тех же строк, переходя к пределу, получаем:

$$\left(E_n - A_{\overline{I}'}{}^t(\overline{x})(A_{\overline{I}'}(\overline{x})A_{\overline{I}'}{}^t(\overline{x}))^{-1}A_{\overline{I}'}(\overline{x})\right)\vartriangle f(\overline{x})^t = 0 \qquad \textbf{(16)}$$

Из (5) и (16) получаем:

$$0 = \vartriangle f(\overline{x}) - \vartriangle f(\overline{x})A_{\overline{I}}{}^t(A_{\overline{I}}A_{\overline{I}}{}^t)^{-1}A_{\overline{I}} = \vartriangle f(\overline{x}) + \sum_{i\in\overline{I}'} u_{\overline{I}}{}^i(\overline{x})\vartriangle g_i(\overline{x}) + u_{\overline{I}}{}^{i_0}(\overline{x})\vartriangle g_{i_0}(\overline{x})$$

$0 = \vartriangle f(\overline{x}) - \vartriangle f(\overline{x})A_{\overline{I}'}{}^t(A_{\overline{I}'}A_{\overline{I}'}{}^t)^{-1}A_{\overline{I}'} = \vartriangle f(\overline{x}) + \sum_{i\in\overline{I}'} u_{\overline{I}'}{}^i(\overline{x})\vartriangle g_i(\overline{x})$, вычитая

$0 = \sum_{i\in I'}\left(u_{\overline{I}}{}^i(\overline{x}) - u_{\overline{I}'}{}^i(\overline{x})\right)\vartriangle g_i(\overline{x}) + \underbrace{u_{\overline{I}}{}^{i_0}(\overline{x})}_{<0}\vartriangle g_{i_0}(\overline{x})$, что опять противоречит линейной независимости

$\left\{\Delta g_i(\overline{x}) \middle| i \in \overline{I}\right\}$. Следовательно, предположение, что существует

$i_0 \in \overline{I} : u_{\overline{I}}{}^{i_0}(\overline{x}) = \left(-\vartriangle f(\overline{x})A_{\overline{I}}(\overline{x})^t(A_{\overline{I}}(\overline{x})A_{\overline{I}}(\overline{x})^t)^{-1}A_{\overline{I}}(\overline{x})\right)_{i_0} < 0$ было неверным, а следовательно

$$-\vartriangle f(\overline{x})A_{\overline{I}}(\overline{x})^t(A_{\overline{I}}(\overline{x})A_{\overline{I}}(\overline{x})^t)^{-1}A_{\overline{I}}(\overline{x}) \geq 0 \qquad \textbf{(17)}$$

Резюмируя (16) и (17), получаем:

последовательность, сгенерированная в ходе выполнения алгоритма, имеет сходящуюся к точке $\overline{x} \in G$ подпоследовательность, для которой для некого набора индексов $\overline{I} \subset \{1,...,m\}$ выполняются следующие условия:

$$\triangle f(\overline{x}) + \sum_{i \in \overline{I}} u_{\overline{I}}{}^{i}(\overline{x}) \triangle g_i(\overline{x}) = 0 \quad \textbf{(18)}$$

$$u_{\overline{I}}{}^{i}(\overline{x}) \geq 0, \forall i \in \overline{I} \quad \textbf{(19)}$$

$$\left|g_i(\overline{x})\right| \leq \varepsilon, \forall i \in \overline{I} \quad \textbf{(20)}$$

□

Утверждение 11

Пусть для сходящейся к нулю последовательности $(\varepsilon_k)_{k\in\mathbb{N}}$, $\lim_{k\to\infty} \varepsilon_k = 0$ по алгоритму, описанному в утверждении 10, сконструирована последовательность $(x_k)_{k\in\mathbb{N}}, x_k \in G$ такая что

$$\triangle f(x_k) + \sum_{i \in I(x_k)} u_{I(x_k)}{}^{i}(x_k) \triangle g_i(x_k) = 0 \quad \textbf{(1)}$$

$$u_{I(x_k)}{}^{i}(x_k) \geq 0, \forall i \in I(x_k) \quad \textbf{(2)}$$

$$\left|g_i(x_k)\right| \leq \varepsilon_k, \forall i \in I(x_k) \quad \textbf{(3)}$$

Тогда существует сходящаяся к $\overline{x} \in G$ подпоследовательность $(x_{k_l})_{l\in\mathbb{N}}$, набор индексов $\overline{I} \subset \{1,...,m\}$ и числа $u_1,...,u_m$ для которых выполняются условия:

$$\lim_{l\to\infty} x_{k_l} = \overline{x} \in G$$

$$\lim_{l\to\infty} u_{I(x_{k_l})}{}^{i}(x_{k_l}) = u_i, \forall i \in \overline{I}$$

$$\triangle f(\overline{x}) + \sum_{i=1}^{m} u^i \triangle g_i(\overline{x}) = 0$$

$$u^i \geq 0, \forall i \in \{1,...,m\}$$

$$u_i g_i(\overline{x}) = 0, \forall i \in \{1,...,m\}$$

Точка $\overline{x}$ является глобальным минимумом функции f на G.

Доказательство

Все элементы последовательности $(x_k)_{k\in\mathbb{N}}$ находятся в компактном множестве G, следовательно, $(x_k)_{k\in\mathbb{N}}$ содержит сходящуюся к $\overline{x} \in G$ подпоследовательность $(x_{k_l})_{l\in\mathbb{N}}$.

Число наборов индексов $I(x_{k_l})$ конечно, а число элементов последовательности $(x_{k_l})_{l\in\mathbb{N}}$ бесконечно, следовательно, существует набор индексов $\overline{I} \subset \{1,...,m\}$, повторяющийся в последовательно бесконечное число раз.

Следовательно, можно выделить подпоследовательность $\left(x_{k_{l_p}}\right)_{p\in\mathbb{N}}$ для которой будет справедливо:

$$\overline{I} = I(x_{k_{l_p}}), \forall p \in \mathbb{N}.$$

Учитывая непрерывную дифференцируемость функций $f, g_1,..., g_m$, а также тот факт, что все $A(x_{k_{l_p}}) = A_{\overline{I}}(x_{k_{l_p}})$ состоят из одинаковых строк, то, переходя к пределу с учетом (1), (2), (3), получим:

$$u_i := \lim_{p\to\infty} u_{I(x_{k_{l_p}})}{}^{i}(x_{k_{l_p}}) = -\triangle f(\overline{x}) A_{\overline{I}}(\overline{x})^t (A_{\overline{I}}(\overline{x}) A_{\overline{I}}(\overline{x})^t)^{-1} A_{\overline{I}}(\overline{x}) \geq 0, \forall i \in \overline{I} \quad \textbf{(4)}$$

$$\vartriangle f(\overline{x})+\sum_{i\in\overline{I}}u^i\vartriangle g_i(\overline{x})=\left(E_n-A_{\overline{I}}(\overline{x})^t(A_{\overline{I}}(\overline{x})A_{\overline{I}}(\overline{x})^t)^{-1}A_{\overline{I}}(\overline{x})\right)\vartriangle f(\overline{x})^t=$$

$$=\lim_{p\to\infty}\left(E_n-A_{\overline{I}}(x_{k_{l_p}})^t(A_{\overline{I}}(x_{k_{l_p}})A_{\overline{I}}(x_{k_{l_p}})^t)^{-1}A_{\overline{I}}(x_{k_{l_p}})\right)\vartriangle f(x_{k_{l_p}})^t= \quad \textbf{(5)}$$

$$=\lim_{p\to\infty}\left(\vartriangle f(x_{k_{l_p}})+\sum_{i\in\overline{I}}u_{I(x_{k_{l_p}})}{}^i(x_{k_{l_p}})\vartriangle g_i(x_{k_{l_p}})\right)=0$$

$$\left|g_i(x_{k_{l_p}})\right|\le\varepsilon_{k_{l_p}},\forall i\in\overline{I}\Rightarrow g_i(\overline{x})=0,\forall i\in\overline{I}\Rightarrow u_i g_i(\overline{x})=0,\forall i\in\overline{I} \quad \textbf{(6)}$$

Определим $u_i:=0,\forall i\notin\overline{I}$, тогда из (5) мы получим:

$$\vartriangle f(\overline{x})+\sum_{i=1}^{m}u^i\vartriangle g_i(\overline{x})=0 \quad \textbf{(7)}$$

Из (4) $u_i\ge 0,\forall i\in\{1,...,m\}$ **(8)**

Из (6) $u_i g_i(\overline{x})=0,\forall i\in\{1,...,m\}$ **(9)**

Определим функцию $L:\mathbb{R}^n\times\mathbb{R}_+^m: L(x,u):=f(x)+\sum_{i=1}^{m}u^i g_i(x)$

Для каждого фиксированного $u\ge 0$ функция L является выпуклой по x, так как $f,g_1,...,g_m$ -выпуклые функции, а $u\ge 0$. L также дифференцируемая функция, поэтому
$L(x,u)\ge L(\overline{x},u)+\left\langle \Delta_x L(\overline{x},u),x-\overline{x}\right\rangle,\forall x\in\mathbb{R}^n$.

Если $x\in G$, то $g_i(x)\le 0,\forall i\in\{1,...,m\}$, следовательно

$$f(x)\ge f(x)+\sum_{i=1}^{m}\underbrace{u^i g_i(x)}_{\le 0}=L(x,u)\ge L(\overline{x},u)+\left\langle \Delta_x L(\overline{x},u),x-\overline{x}\right\rangle=$$

$$=f(\overline{x})+\sum_{i=1}^{m}\underbrace{u^i g_i(\overline{x})}_{=0}+\left\langle \underbrace{\vartriangle f(\overline{x})+\sum_{i=1}^{m}u^i\vartriangle g_i(\overline{x})}_{=0},x-\overline{x}\right\rangle=f(\overline{x}) \quad \textbf{(10)}$$

Т.е. $\overline{x}$ - минимум f на G .

□