Методы безусловной минимизации функций многих переменных

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Определение

Методы получения точек, соответствующих условию (2), называются методами спуска.

Определение

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Методы минимизации функций многих переменных

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Методы безусловной минимизации функций многих переменных

Определение

Поверхностью уровня функции f(x) называется множество точек, в которых функция принимает постоянное значение,
то есть f(x)=const.

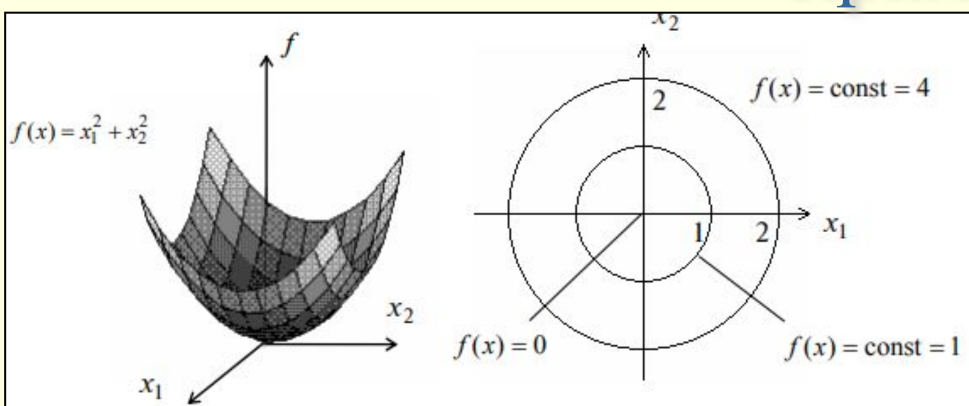
Рассмотрим задачу

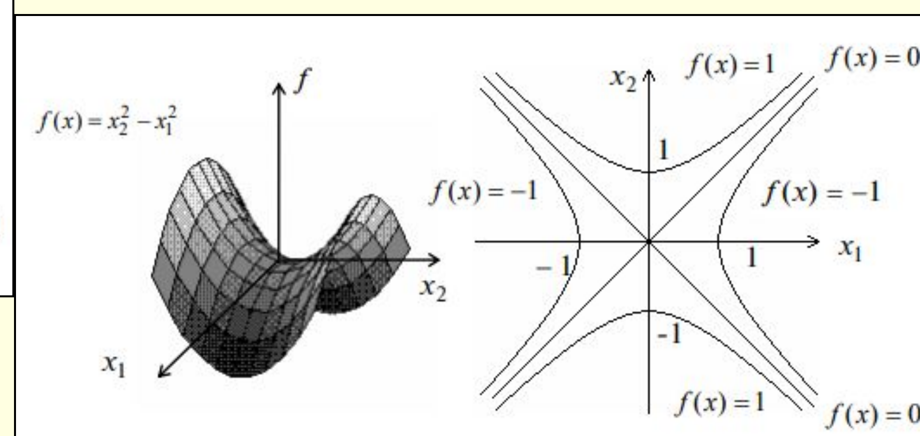
$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) \geq f(x_1) \geq \dots f(x_k) \geq f(x_{k+1}) \geq \dots \qquad (2)$$

Примеры





Необходимое условие оптимальности для дифференцируемых функций

Пусть задано пространство E^n и $X \subset E^n$ некоторое множество.

Определение

Функция f(x) называется x_0 **дифференцируемой в точке** $\mathbf{x}_0 \in X$, если существует такой вектор $\nabla f(x_0)$, что для любого x в окрестности справедливо асимптотическое равенство:

$$f(x) = f(x_0) + \left(\nabla f(x_0), x - x_0\right) + o(\|x - x_0\|)$$

$\left(\nabla f(x_0), x - x_0\right)$- скалярное произведение векторов по правилу $(x, y) = \sum_i x_i y_i$;

$\|x\| = \sqrt{(x, x)}$ - евклидова норма вектора;

$o(\alpha)$ - величина бесконечно малая по сравнению с α

Определение

Вектор $\nabla f(x_0)$ называется градиентом функции f(x) в точке x_0 и для $x \in R^n$

$$\nabla f(x_0) = \left(\frac{\partial f(x_0)}{\partial x_1}, \frac{\partial f(x_0)}{\partial x_2}, \dots, \frac{\partial f(x_0)}{\partial x_n} \right)$$

Утверждение

Оптимальное решение x^*, минимизирующее дифференцируемую функцию f(x) на множестве X, находится либо на границе $\overline{X}$ множества X, либо на множестве решений уравнения $\nabla \mathbf{f}(\mathbf{x}) = 0$

Если f(x) дифференцируема, то направлением наибыстрейшего убывания функции f(x) в точке x^k является направление антиградиента $-\nabla f(x^k)$

Определение

Матрицей Гессе H(x) дважды непрерывно дифференцируемой в точке x функции f(x) называется матрица частных производных второго порядка, вычисленных в данной точке

$$H(x)=\begin{pmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{pmatrix};$$

Пример

Для функции $f(x)=x_1^2+x_2^4$ вычислить градиент и найти матрицу Гессе в точках $x^0=(0,0)^T$, $x^1=(1,1)^T$.

$$\nabla f(x)=\left(2x_1,4x_2^3\right)^T, H(x)=\begin{pmatrix} 2 & 0 \\ 0 & 12x_2^2 \end{pmatrix}; \quad \nabla f(x^0)=(0,0)^T, H(x^0)=\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix};$$

$$\nabla f(x^1)=(2,4)^T, H(x^1)=\begin{pmatrix} 2 & 0 \\ 0 & 12 \end{pmatrix}.$$

Определение

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \ldots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей

последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Методы безусловной минимизации функций многих переменных

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

В различных вариантах методов спуска используются разные способы выбора скаляра λ (метод с постоянным шагом, с дроблением шага, метод наискорейшего спуска)

Метод покоординатного спуска (конфигураций или Хука-Дживса)

Метод конфигураций включает два основных этапа:

)Поиск вокруг базисной точки (исследующий поиск);

)Поиск в направлении, выбранном для оптимизации (поиск по образцу)

Метод покоординатного спуска

Основная идея метода покоординатного спуска

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

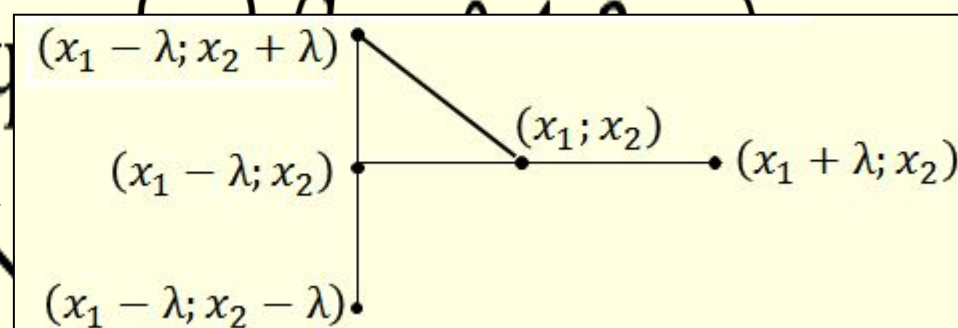
Метод покоординатного спуска

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей

последовательности точек (x_k) $(k = 0, 1, 2, \ldots)$ такой, что



$$f(x_0) > f(x_1) > \ldots > f(x_k) > f(x_{k+1}) > \ldots \quad (2)$$

Метод покоординатного спуска

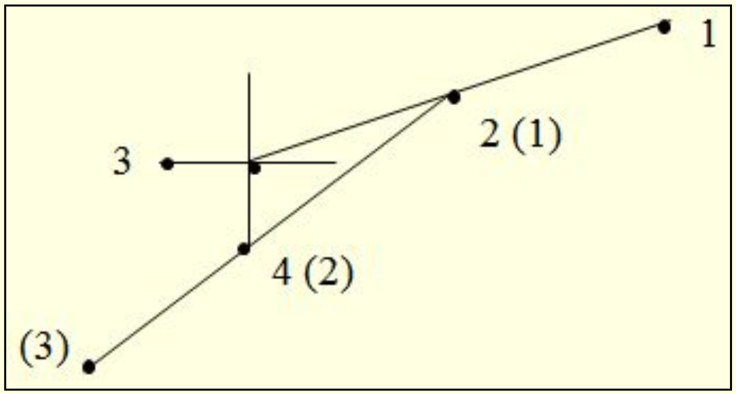
Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Иллюстрация для двух координат



Алгоритм метода покоординатного спуска

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Шаги 1 и 2 алгоритма метода покоординатного спуска осуществляют пробный поиск, а шаг 3 является ускоряющим шагом по направлению $x_{k+1} - x_k$. На шаге 4 длина шага сокращается.

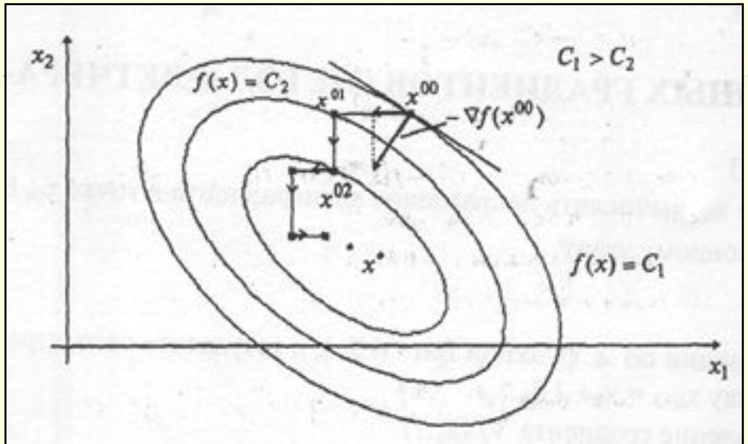
Метод покоординатного спуска обладает следующими преимуществами перед градиентными методами:

1) он не требует задания целевой функции в явном виде, что является существенным достоинством при решении сложных экономических и технических задач;

2) метод позволяет легко учитывать ограничения, накладываемые на переменные, а также на допустимую область поиска.

Недостатком метода покоординатного спуска является то, что он может останавливаться вблизи локального минимума, не в состоянии обеспечить дальнейшее улучшение.

Геометрическая интерпретация



ПРИМЕР

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \dots f(x_k) > f(x_{k+1}) > \dots \quad (2)$$

1. Зададим $\varepsilon = 0{,}1$, $x^1 = (2;-4)^T$, $\lambda = 1$, $\alpha = 1$, $n = 2$. Положим $e = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $k = j = 1$.

2^{11}. Вычислим

$$Y_1 = x^1 = (2;-4)^T,\ f(Y_1) = 2,$$

$$Y_1 + \lambda e_1 = (3;-4)^T,\ f(Y_1 + \lambda e_1) = 1.$$

3^{11}. Так как $f(Y_1 + \lambda e_1) < f(Y_1)$, то $Y_2 = Y_1 + \lambda e_1 = (3;-4)^T$.

4^{11}. Так как $j = 1 < n = 2$, то $j = 2$. Перейти к шагу 2.

2^{12}. Вычислим

$$Y_2 + \lambda e_2 = (3;-3)^T,\ f(Y_2 + \lambda e_2) = 4.$$

3^{12}. Так как $f(Y_2 + \lambda e_2) > f(Y_2)$, то $Y_2 - \lambda e_2 = (3;-5)^T$, $f(Y_2 - \lambda e_2) = 0$. Так как $f(Y_2 - \lambda e_2) < f(Y_2)$, то $Y_3 = Y_2 - \lambda e_2 = (3;-5)^T$.

4^{12}. Так как $j = n = 2$ и $f(Y_3) < f(x^1)$, то переход к шагу 5.

5^{12}. Положим $x^3 = Y_3 = (3;-5)^T$, $Y_1 = x^2 + \alpha(x^2 - x^1) = (4;-6)^T$. Примем $k = 2$, $j = 1$. Переход к шагу 2.

2^{21}. Вычислим

$$Y_1 = (4;-6)^T,\ f(Y_1) = 2,$$

$$Y_1 + \lambda e_1 = (5;-6)^T,\ f(Y_1 + \lambda e_1) = 5.$$

3^{21}. Так как $f(Y_1 + \lambda e_1) > f(Y_1)$, то $Y_1 - \lambda e_1 = (3;-6)^T$, $f(Y_1 - \lambda e_1) = 1$. Так как $f(Y_1 - \lambda e_1) < f(Y_1)$, то $Y_2 = Y_1 - \lambda e_1 = (3;-6)^T$.

4^{21}. Так как $j = 1 < n = 2$, то $j = 2$. Перейти к шагу 2.

2^{22}. Вычислим

$$Y_2 + \lambda e_2 = (3;-5)^T,\ f(Y_2 + \lambda e_2) = 0.$$

3^{22}. Так как $f(Y_2 + \lambda e_2) < f(Y_2)$, то $Y_3 = Y_2 + \lambda e_2 = (3;-5)^T$.

4^{22}. Так как $j = n = 2$ и $f(Y_3) = f(x^2)$, то переход к шагу 6.

6^{22}. Поскольку $\lambda = 1 > \varepsilon = 0{,}1$, то положим $\lambda = \frac{\lambda}{2} = \frac{1}{2}$, $Y_1 = x^2 = (3;-5)^T$, $x^3 = x^2 = (3;-5)^T$. Примем $k = 3$, $j = 1$. Переход к шагу 2.

2^{31}. Вычислим

$$Y_1 = (3;-5)^T,\ f(Y_1) = 0,$$

$$Y_1 + \lambda e_1 = (3{,}5;-5)^T,\ f(Y_1 + \lambda e_1) = \frac{1}{4}.$$

3^{31}. Так как $f(Y_1 + \lambda e_1) > f(Y_1)$, то $Y_1 - \lambda e_1 = (2{,}5;-5)^T$, $f(Y_1 - \lambda e_1) = \frac{1}{4}$. Так как $f(Y_1 - \lambda e_1) > f(Y_1)$, то $Y_2 = Y_1 = (3;-5)^T$.

4^{31}. Так как $j = 1 < n = 2$, то $j = 2$. Перейти к шагу 2.

2^{32}. Вычислим

$$Y_2 + \lambda e_2 = (3;-4{,}5)^T,\ f(Y_2 + \lambda e_2) = \frac{1}{4}.$$

3^{32}. Так как $f(Y_2 + \lambda e_2) > f(Y_2)$, то $Y_2 - \lambda e_2 = (3;-5{,}5)^T$, $f(Y_2 - \lambda e_2) = \frac{1}{4}$. Так как $f(Y_2 - \lambda e_2) > f(Y_2)$, то $Y_3 = Y_2 = (3;-5)^T$.

4^{32}. Так как $j = n = 2$ и $f(Y_3) = f(x^3)$, то переход к шагу 6.

6^{32}. Поскольку $\lambda = \frac{1}{2} > \varepsilon = 0{,}1$, то положим $\lambda = \frac{\lambda}{2} = \frac{1}{4}$, $Y_1 = x^3 = (3;-5)^T$, $x^4 = x^3 = (3;-5)^T$. Примем $k = 4$, $j = 1$. Переход к шагу 2.

2^{41}. Вычислим

$$Y_1 = (3;-5)^T,\ f(Y_1) = 0,$$

$$Y_1 + \lambda e_1 = (3{,}25;-5)^T,\ f(Y_1 + \lambda e_1) = \frac{1}{16}.$$

3^{41}. Так как $f(Y_1 + \lambda e_1) > f(Y_1)$, то $Y_1 - \lambda e_1 = (2{,}75;-5)^T$, $f(Y_1 - \lambda e_1) = \frac{1}{16}$. Так как $f(Y_1 - \lambda e_1) > f(Y_1)$, то $Y_2 = Y_1 = (3;-5)^T$.

4^{41}. Так как $j = 1 < n = 2$, то $j = 2$. Перейти к шагу 2.

2^{42}. Вычислим

$$Y_2 + \lambda e_2 = (3;-4{,}75)^T,\ f(Y_2 + \lambda e_2) = \frac{1}{16}.$$

3^{42}. Так как $f(Y_2 + \lambda e_2) > f(Y_2)$, то $Y_2 - \lambda e_2 = (3;-5{,}25)^T$, $f(Y_2 - \lambda e_2) = \frac{1}{16}$. Так как $f(Y_2 - \lambda e_2) > f(Y_2)$, то $Y_3 = Y_2 = (3;-5)^T$.

4^{42}. Так как $j = n = 2$ и $f(Y_3) = f(x^4)$, то переход к шагу 6.

6^{42}. Поскольку $\lambda = 0{,}25 > \varepsilon = 0{,}1$, то положим $\lambda = \frac{\lambda}{2} = \frac{1}{8}$, $Y_1 = x^4 = (3;-5)^T$, $x^5 = x^4 = (3;-5)^T$. Примем $k = 5$, $j = 1$. Переход к шагу 2.

2^{51}. Вычислим

$$Y_1 = (3;-5)^T,\ f(Y_1) = 0,$$

$$Y_1 + \lambda e_1 = (3{,}125;-5)^T,\ f(Y_1 + \lambda e_1) = \frac{1}{64}.$$

3^{51}. Так как $f(Y_1+\lambda e_1)>f(Y_1)$, то $Y_1-\lambda e_1=(2{,}875;-5)^T$, $f(Y_1-\lambda e_1)=\frac{1}{64}$. Так как $f(Y_1-\lambda e_1)>f(Y_1)$, то $Y_2=Y_1=(3;-5)^T$.

4^{51}. Так как $j=1<n=2$, то $j=2$. Перейти к шагу 2.

2^{52}. Вычислим

$$Y_2+\lambda e_2=(3;-4{,}875)^T,\ f(Y_2+\lambda e_2)=\frac{1}{64}.$$

3^{52}. Так как $f(Y_2+\lambda e_2)>f(Y_2)$, то $Y_2-\lambda e_2=(3;-5{,}125)^T$, $f(Y_2-\lambda e_2)=\frac{1}{64}$. Так как $f(Y_2-\lambda e_2)>f(Y_2)$, то $Y_3=Y_2=(3;-5)^T$.

4^{52}. Так как $j=n=2$ и $f(Y_3)=f(x^5)$, то переход к шагу 6.

6^{52}. Поскольку $\lambda=0{,}125>\varepsilon=0{,}1$, то положим $\lambda=\frac{\lambda}{2}=\frac{1}{16}$, $Y_1=x^5=(3;-5)^T$, $x^6=x^5=(3;-5)^T$. Примем $k=6$, $j=1$. Переход к шагу 2.

2^{61}. Вычислим

$$Y_1=(3;-5)^T,\ f(Y_1)=0,$$

$$Y_1+\lambda e_1=(3{,}0625;-5)^T,\ f(Y_1+\lambda e_1)=\frac{1}{256}.$$

3^{61}. Так как $f(Y_1+\lambda e_1)>f(Y_1)$, то $Y_1-\lambda e_1=(2{,}9375;-5)^T$, $f(Y_1-\lambda e_1)=\frac{1}{256}$. Так как $f(Y_1-\lambda e_1)>f(Y_1)$, то $Y_2=Y_1=(3;-5)^T$.

4^{61}. Так как $j=1<n=2$, то $j=2$. Перейти к шагу 2.

2^{62}. Вычислим

$$Y_2+\lambda e_2=(3;-4{,}9375)^T,\ f(Y_2+\lambda e_2)=\frac{1}{256}.$$

3^{62}. Так как $f(Y_2+\lambda e_2)>f(Y_2)$, то $Y_2-\lambda e_2=(3;-5{,}0625)^T$, $f(Y_2-\lambda e_2)=\frac{1}{256}$. Так как $f(Y_2-\lambda e_2)>f(Y_2)$, то $Y_3=Y_2=(3;-5)^T$.

4^{62}. Так как $j=n=2$ и $f(Y_3)=f(x^6)$, то переход к шагу 6.

6^{62}. Поскольку $\lambda=0{,}0625<\varepsilon=0{,}1$, то $x^*=x^5=(3;\ -5)^T$, $f(x^*)=0$.

Метод наискорейшего спуска

Идея метода наискорейшего спуска

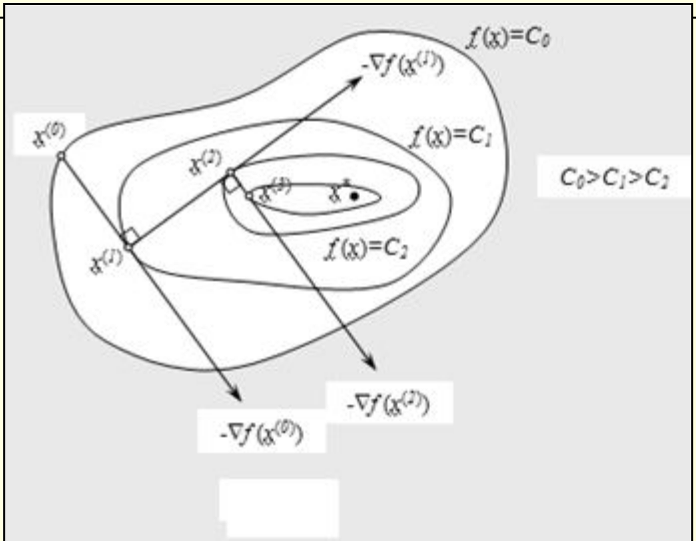
Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

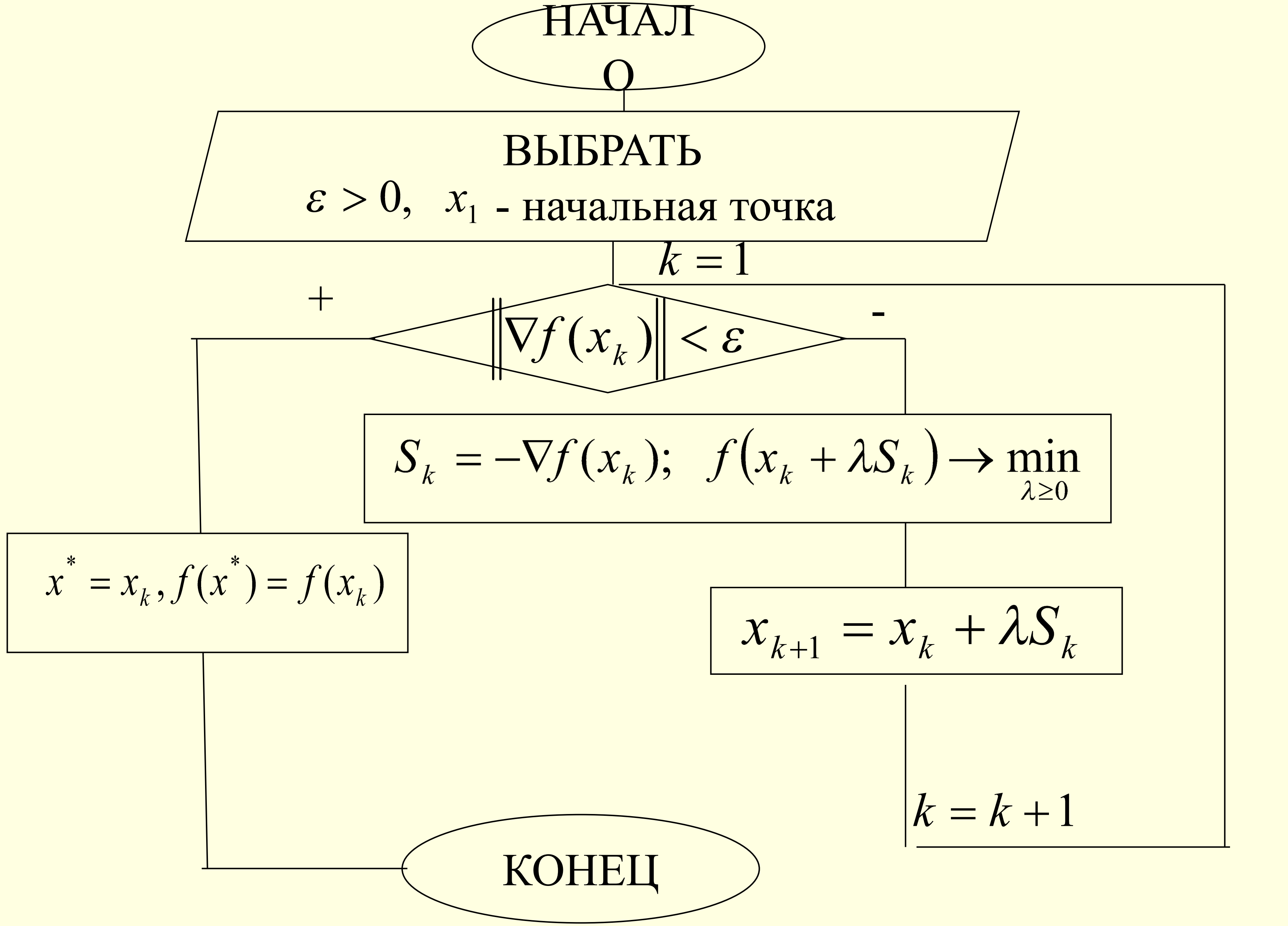
и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Геометрическая интерпретация



Метод наискорейшего спуска



Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Метод наискорейшего спуска

Замечания:

Метод наискорейшего спуска сходится достаточно быстро, если для минимизирующей функции поверхности уровня близки к сферам. Однако, если поверхности уровня минимизирующей функции сильно вытянуты в некотором направлении, то метод сходится, но может застревать в локальном минимуме.

В двумерном случае рельеф поверхности напоминает рельеф местности с оврагом. Поэтому такие функции называются «овражными». Вдоль направления, характеризующего дно оврага «овражная» функция меняется незначительно, а в других направлениях, характеризующих склон оврага происходит резкое изменение функции. Если начальная точка попадает на склон оврага, то направление градиентного спуска оказывается перпендикулярным к дну оврага и очередное приближение попадает на противоположный склон оврага. В результате, вместо того, чтобы двигаться вдоль дна оврага в направлении к точке минимума, траектория спуска совершает зигзагообразные скачки поперек оврага, почти не приближаясь к точке минимума.

ПРИМЕР

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

1. Зададим $\varepsilon = 0{,}1$, $x^0 = (-3;5)^T$. Найдем градиент функции в произвольной точке $\nabla f(x) = (2(x_1 - 3);\ 2(x_2 + 5))^T$.

2. Вычислим значение функции в начальной точке $f(x^0) = 136$.

3. В качестве направления поиска выберем вектор градиент в текущей точке $\nabla f(x^0) = (-12;\ 20)^T$.

4. Проверим критерий выхода $\|\nabla f(x^0)\| \le \varepsilon$: $\|\nabla f(x^0)\| = 23{,}324 > \varepsilon = 0{,}1$.

5. Сделаем шаг вдоль направления антиградиента:

$$x^1 = x^0 - \lambda_1 \cdot \nabla f(x^0) = \begin{pmatrix} -3 \\ 5 \end{pmatrix} - \lambda_1 \cdot \begin{pmatrix} -12 \\ 20 \end{pmatrix} = \begin{pmatrix} 12\lambda_1 - 3 \\ 5 - 20\lambda_1 \end{pmatrix}.$$

6. Вычислим значение функции в новой точке: $f(x^1) = 136 \cdot (2\lambda_1 - 1)^2$.

7. Найдем такой шаг, чтобы целевая функция достигала минимума вдоль этого направления. Из необходимых условий существования экстремума функции $f'(x) = 0$ следует, что $\lambda_1 = \dfrac{1}{2}$. Тогда имеем: $x^1 = \begin{pmatrix} 3 \\ -5 \end{pmatrix}$.

8. Вектор градиент в текущей точке $\nabla f(x^1) = (0;\ 0)^T$.

9. Проверим критерий выхода $\|\nabla f(x^1)\| \le \varepsilon$: $\|\nabla f(x^1)\| = 0 < \varepsilon = 0{,}1$.

Следовательно, $x^* = (3;\ -5)^T$, $f(x^*) = 0$.

Метод сопряженных градиентов

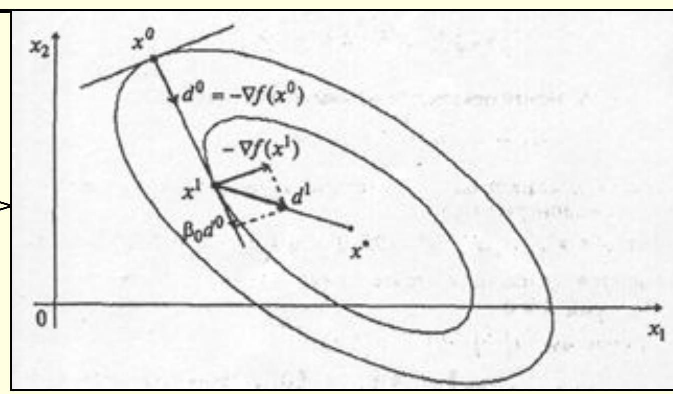
Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Геометрическая интерпретация



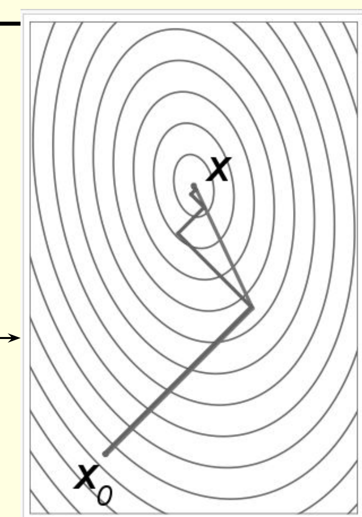


Иллюстрация последовательных приближений метода наискорейшего спуска и метода сопряжённых градиентов к точке экстремума.

Обоснование метода сопряженных градиентов

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

Определение

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \ldots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

Нулевая итерация

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \ldots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей

последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Метод сопряженных градиентов

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Замечание: Если требуется найти глобальный минимум функции f(x), То для строго выпуклой f(x) Решение этой задачи аналогично поиску локального минимума функции. В случае, когда f(x) имеет несколько локальных минимумов, поиск глобального минимума осуществляется в результате перебора всех локальных минимумов.

ПРИМЕР

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

1. Зададим $\varepsilon = 0{,}1$, $x^0 = (-3;5)^T$. Найдем градиент функции в произвольной точке $\nabla f(x) = \left(2(x_1 - 3);\ 2(x_2 + 5)\right)^T$.

2. Вектор градиент в текущей точке $\nabla f(x^0) = (-12;\ 20)^T$.

3. Проверим критерий выхода $\left\|\nabla f(x^0)\right\| \le \varepsilon$: $\left\|\nabla f(x^0)\right\| = 23{,}324 > \varepsilon = 0{,}1$.

4. Вычислим новую точку:

$$x^1 = x^0 - \lambda_1 \cdot \nabla f(x^0) = \begin{pmatrix} -3 \\ 5 \end{pmatrix} - \lambda_1 \cdot \begin{pmatrix} -12 \\ 20 \end{pmatrix} = \begin{pmatrix} 12\lambda_1 - 3 \\ 5 - 20\lambda_1 \end{pmatrix}.$$

5. Вычислим значение функции в новой точке: $f(x^1) = 136 \cdot (2\lambda_1 - 1)^2$.

6. Найдем такой шаг, чтобы целевая функция достигала минимума вдоль этого направления. Из необходимых условий существования экстремума функции $f'(x) = 0$ следует, что $\lambda_1 = \dfrac{1}{2}$. Тогда имеем: $x^1 = \begin{pmatrix} 3 \\ -5 \end{pmatrix}$.

7. Вектор градиент в текущей точке $\nabla f(x^1) = (0;\ 0)^T$.

8. Проверим критерий выхода $\left\|\nabla f(x^1)\right\| \le \varepsilon$: $\left\|\nabla f(x^1)\right\| = 0 < \varepsilon = 0{,}1$.

Следовательно, $x^* = (3;\ -5)^T$, $f(x^*) = 0$.

Метод Ньютона

Метод Ньютона относится к градиентным методам второго порядка, в котором направление минимизации выбирается умножением вектора антиградиента на матрицу, обратную матрице Гессе.

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \ldots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Метод Ньютона

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Алгоритм метода Ньютона

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Алгоритм метода Ньютона

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

ПРИМЕР

Рассмотрим задачу
$$f(x) \to min, x \in X \subseteq R^n$$
и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \ldots)$ таких, что

1. Зададим $\varepsilon_1 = 0{,}1$, $\varepsilon_2 = 0{,}1$, $x^0 = (-3;5)^T$, $M = 10$. Найдем градиент функции в произвольной точке $\nabla f(x) = \left(2(x_1 - 3);\ 2(x_2 + 5)\right)^T$ и матрицу Гессе

$$H(x) = \begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

2. Положим $k = 0$.

3^0. Вычислим $\nabla f(x^0)$: $\nabla f(x^0) = (-12;\ 20)^T$.

4^0. Проверим выполнение условия выхода:
$$\left\|\nabla f(x^0)\right\| \le \varepsilon_1 :\ \left\|\nabla f(x^0)\right\| = 23{,}324 > \varepsilon = 0{,}1.$$
Переходим к шагу 5.

5^0. Проверим условие $k \ge M$: $k = 0 < 10$. Переходим к шагу 6.

6^0. Вычислим $H(x^0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$, $H^{-1}(x^0) = \begin{pmatrix} 0{,}5 & 0 \\ 0 & 0{,}5 \end{pmatrix}$.

7^0. Проверим выполнение условия $H^{-1}(x^0) > 0$. Так как $\Delta_1 = 0{,}5 > 0$, $\Delta_2 = 0{,}25 > 0$, то согласно критерию Сильвестра $H^{-1}(x^0) > 0$.

8^0. Определим $d_0 = -H^{-1}(x^0) \cdot \nabla f(x^0) = \begin{pmatrix} -0{,}5 & 0 \\ 0 & -0{,}5 \end{pmatrix} \cdot \begin{pmatrix} -2 \\ 20 \end{pmatrix} = \begin{pmatrix} 6 \\ -10 \end{pmatrix}$.

9^0. Вычислим $x^1 = x^0 + t_0 d_0$ $(t_0 = 1)$: $x^1 = \begin{pmatrix} -3 \\ 5 \end{pmatrix} + \begin{pmatrix} 6 \\ -10 \end{pmatrix} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$.

10^0. Проверим выполнение условия:
$$\left\|x^1 - x^0\right\| < \varepsilon_2 :\ \left\|x^1 - x^0\right\| = 11{,}6619 > \varepsilon_2 = 0{,}1.$$
Полагаем $k = 1$, переходим к шагу 3.

3^1. Вычислим $\nabla f(x^1)$: $\nabla f(x^1) = (0;\ 0)^T$.

4^1. Проверим выполнение условия выхода:
$$\left\|\nabla f(x^1)\right\| \le \varepsilon_1 :\ \left\|\nabla f(x^1)\right\| = 0 < \varepsilon = 0{,}1.$$

Следовательно, $x^* = (3;\ -5)^T$, $f(x^*) = 0$.

Работа: «Методы многомерного поиска»

ЦЕЛЬ РАБОТЫ Ознакомиться с методами могомерного поиска. Сравнить различные алгоритмы по эффективности на тестовых примерах.

ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

Рассмотрим задачу

$$f(x) \to min, x \in X \subseteq R^n$$

и процедуру построения минимизирующей последовательности точек $\{x_k\}$ $(k = 0, 1, 2, \dots)$ таких, что

$$f(x_0) > f(x_1) > \cdots f(x_k) > f(x_{k+1}) > \cdots \qquad (2)$$

ВАРИАНТЫ ТЕСТОВЫХ ЗАДАНИЙ

1) $f(x) = x_1^2 + 2x_2^2 - 4x_1 + 2x_2, \quad x^0 = (1,0)$
2) $f(x) = 2x_1^2 + x_2^2 - 12x_1, \quad x^0 = (5,3)$
3) $f(x) = x_1^2 + x_2^2 - 2x_1 - 4x_2, \quad x^0 = (5,10)$
4) $f(x) = x_1^2 + x_2^2 - 4x_1 - 2x_2, \quad x^0 = (4,5)$
5) $f(x) = x_1^2 + 4x_2^2 - 6x_1 - 32x_2, \quad x^0 = (0,0)$
6) $f(x) = 2x_1^2 + x_2^2 - x_1x_2, \quad x^0 = (1,2)$
7) $f(x) = x_1^2 + 3x_2^2 - 2x_1x_2, \quad x^0 = (2,0)$

СОДЕРЖАНИЕ ОТЧЕТА

- титульный лист; цель работы; задание;
- таблицы с результатами исследований по каждому методу, сравнение эффективности работы разных алгоритмов;
- безмашинный вариант реализации двух итераций каждым методом;
- выводы по всем пунктам задания.