Метод градиентного спуска

Идея МГС

Даны множество $\mathcal{K}$ и функция

$$f : \mathcal{K} \to \mathbb{R},$$

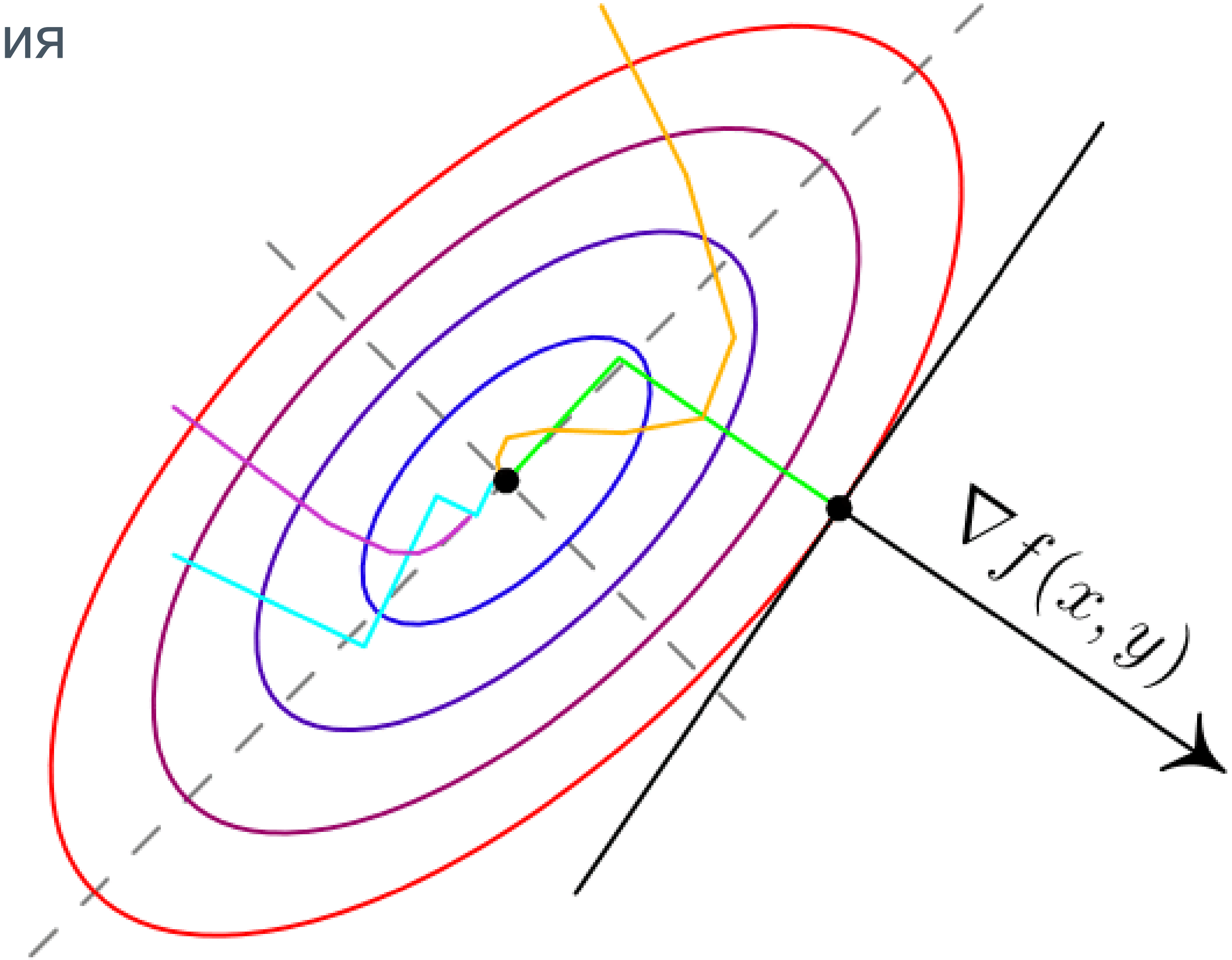
требуется найти точку

$$x^* \in \mathcal{K},$$

такую, что

$f(x) \geq f(x^*)$ для всех $x \in \mathcal{K}$

$$f(x) \to \min_{x \in \mathcal{K}}.$$

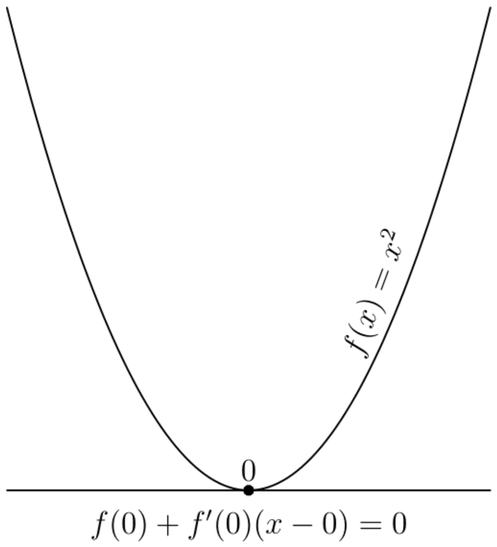


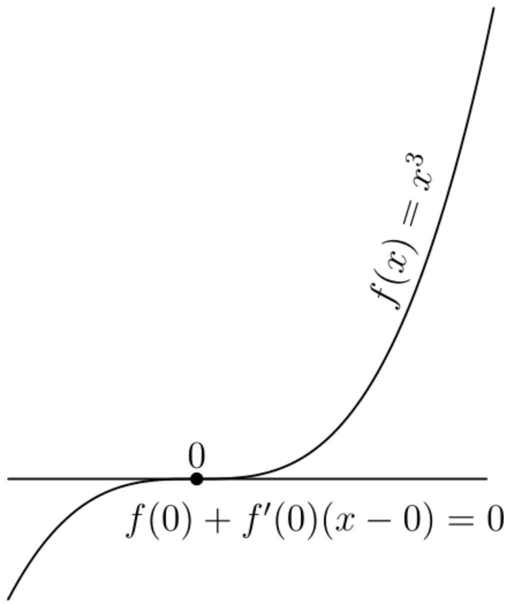
Математические основы

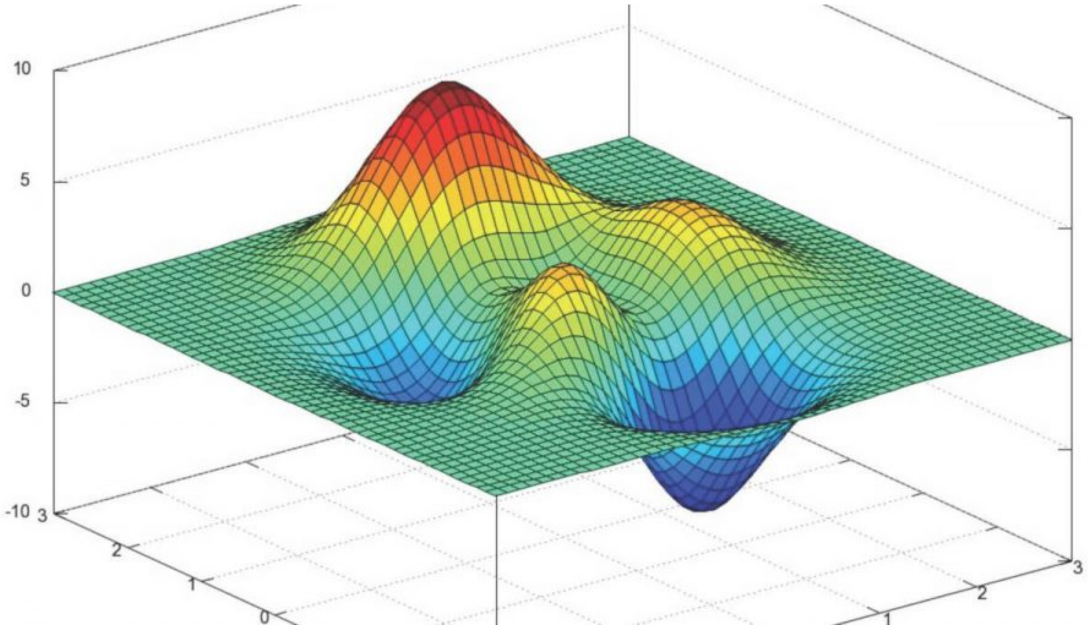
$$f(x) \to \min_{x \in \mathbb{R}}.$$

$$f'(x^*) = 0 \qquad f(x) \approx f(x^*) + f'(x^*)(x - x^*).$$

$$\nabla f(x^*) = 0 \qquad f(x) \approx f(x^*) + \nabla f(x^*)^T (x - x^*)$$







Квадратичные функции

$$f(x) = f(x_1, x_2, \ldots, x_n) = \frac{1}{2}\sum_{i,j=1}^{n} a_{ij}x_ix_j - \sum_{i=1}^{n} b_ix_i + c$$

$$f(x) = \frac{1}{2}x^TAx - b^Tx + c,$$

$$\frac{\partial}{\partial x_i}f(x_1, x_2, \ldots, x_n) = a_{ii}x_i + \sum_{j\neq i}\frac{1}{2}(a_{ij} + a_{ji})x_j - b_i,$$

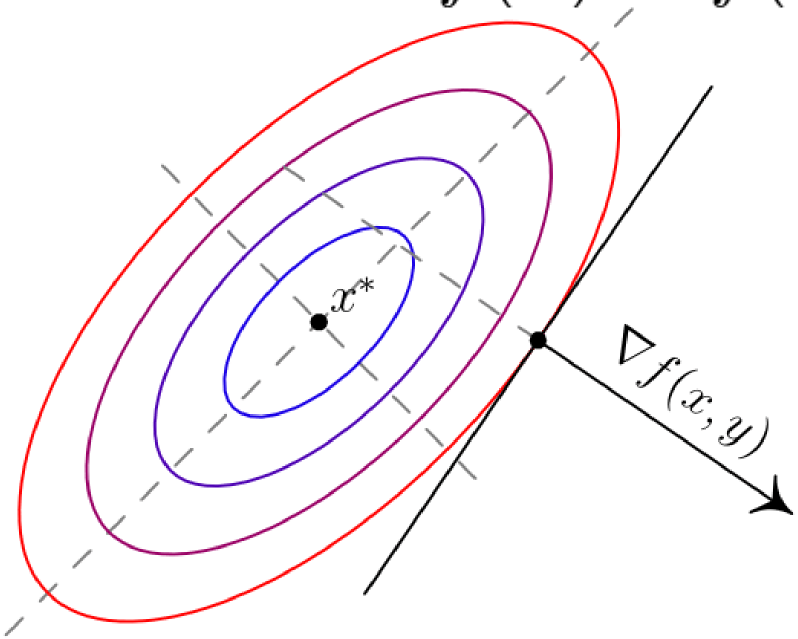
$$\nabla f(x) = Ax - b,$$

Применение свойств градиента

дана точка x, *найти точку* $\bar{x}$ *такую, что* $f(\bar{x}) < f(x)$

$$\bar{x} = x - \alpha \nabla f(x),\ \alpha > 0$$

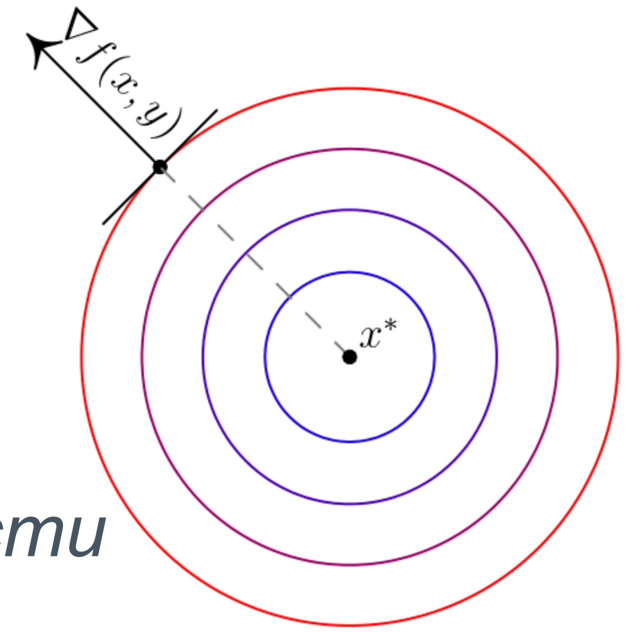
$$f(\bar{x}) \approx f(x) - \alpha \|\nabla f(x)\|^2 < f(x)$$



$f(x,y) = \frac{3}{2}(x-y)^2 + \frac{1}{3}(x+y)^2$

$$f(x) = c$$

$\nabla f(x)$ *задает нормаль к поверхности*



$f(x,y) = x^2 + y^2$

Градиентный спуск

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k).$$

Величина α_k называется *размером шага*

(в машинном обучении — *скорость обучения*)

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$$

$\alpha_k < \dfrac{2}{L}$ гарантирует убывание $f(x_k)$.

Модификации градиентного спуска

- Инерционные или ускоренные градиентные методы
- Метод Чебышева
- Метод сопряженных градиентов
- Метод Нестерова
- Стохастический градиентный спуск
- Субградиентный спуск
- Proximal методы

Инерционные или ускоренные градиентные методы

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k) + \beta_k (x_k - x_{k-1}).$$

Такие методы обладают двумя важными свойствами:

1. Они практически не усложняют обычный градиентный спуск в вычислительном плане.
2. При аккуратном подборе α_k, β_k такие методы на порядок быстрее, чем обычный градиентный спуск даже с оптимально подобранным шагом.