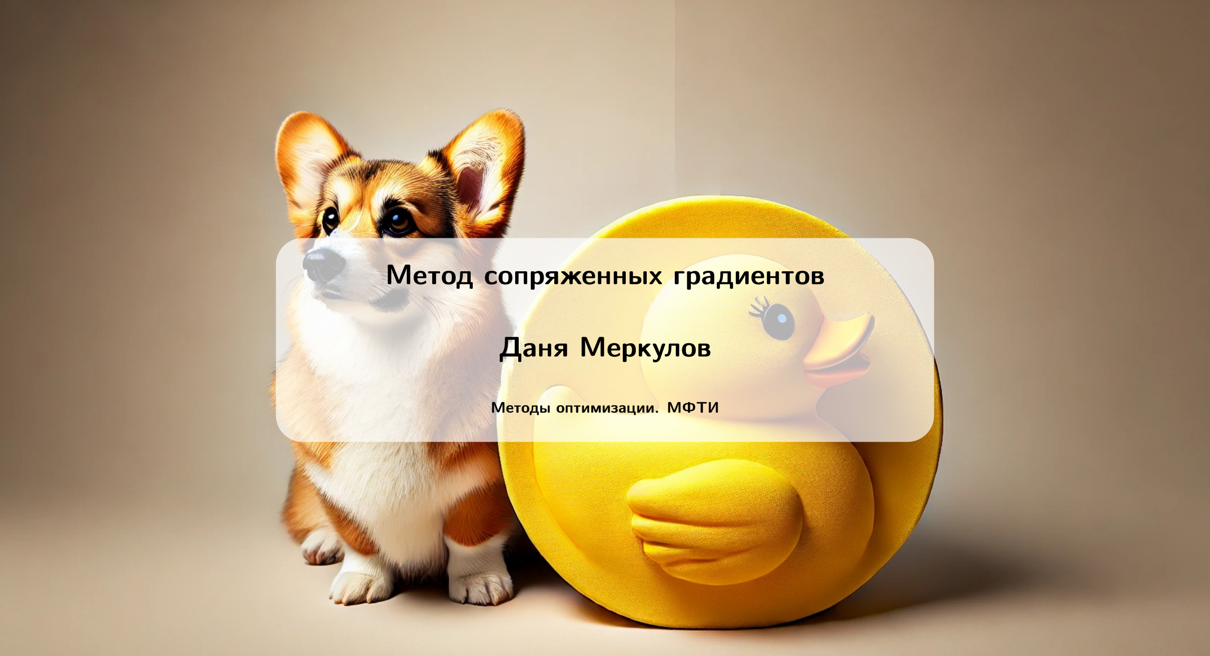


Метод сопряженных градиентов

Даня Меркулов

Методы оптимизации. МФТИ

Квадратичная задача оптимизации

$$\min_{x \in \mathbb{R}^n} f(x) = \min \frac{1}{2} x^T A x - b^T x + c$$

$A \in \mathbb{S}^n_{++}$

- Ускоренная сх-ть
- Гарантированно сходится за не более, чем n итераций в точной арифм.
- Если у матрицы A k различных СЗ, то метод сойдётся за k итераций
- Стоимость итерации такая же как у HB, NAG

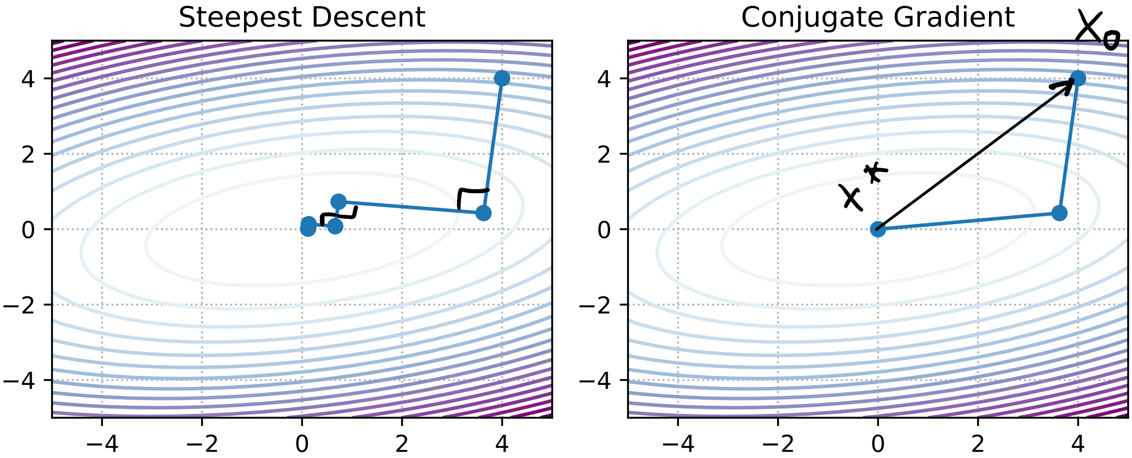
Сильно выпуклая квадратичная функция

Рассмотрим следующую квадратичную задачу оптимизации:

$$
\min_{x \in \mathbb{R}^n} f(x) = \min_{x \in \mathbb{R}^n} \frac{1}{2} x^\top A x - b^\top x + c, \text{ где } A \in \mathbb{S}^n_{++}. \tag{1}
$$

Условия оптимальности

$$
Ax^* = b
$$



Наискорейший спуск aka точный линейный поиск

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

$$
\boxed{\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_{k+1}) = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))}
$$

Более теоретический, чем практический подход к выбору шага. Он также позволяет анализировать сходимость, но точный линейный поиск может быть численно сложным, если вычисление функции занимает слишком много времени или требует слишком много ресурсов.

Интересное теоретическое свойство этого метода заключается в том, что каждая следующая итерация метода ортогональна предыдущей:

$$
\frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T \nabla^2 f(x_k) \nabla f(x_k)}
$$

$$
\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))
$$

$f(x) = \frac{1}{2} x^T A x - b^T x \qquad \nabla f = Ax - b, \quad \nabla^2 f = A$

$\nabla f(x_{k+1})^T \nabla f(x_k) = 0$

$(Ax_{k+1} - b)^T (Ax_k - b) = 0$

$\boxed{\alpha_k = \frac{g_k^T g_k}{g_k^T A g_k}}$

$(A(x_k - \alpha_k g_k) - b)^T \cdot g_k = 0 \to (g_k - \alpha_k A g_k)^T g_k = 0$

Наискорейший спуск aka точный линейный поиск

$$\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_{k+1}) = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))$$

Более теоретический, чем практический подход к выбору шага. Он также позволяет анализировать сходимость, но точный линейный поиск может быть численно сложным, если вычисление функции занимает слишком много времени или требует слишком много ресурсов.

Интересное теоретическое свойство этого метода заключается в том, что каждая следующая итерация метода ортогональна предыдущей:

$$\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))$$

Условия оптимальности:

Наискорейший спуск aka точный линейный поиск

$$\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_{k+1}) = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))$$

Более теоретический, чем практический подход к выбору шага. Он также позволяет анализировать сходимость, но точный линейный поиск может быть численно сложным, если вычисление функции занимает слишком много времени или требует слишком много ресурсов.

Интересное теоретическое свойство этого метода заключается в том, что каждая следующая итерация метода ортогональна предыдущей:

$$\alpha_k = \text{arg}\min_{\alpha \in \mathbb{R}^+} f(x_k - \alpha \nabla f(x_k))$$

Условия оптимальности:

$$\nabla f(x_k)^T \nabla f(x_{k+1}) = 0$$

Оптимальное значение для квадратичных функций

$$\nabla f(x_k)^\top A(x_k - \alpha \nabla f(x_k)) - \nabla f(x_k)^\top b = 0 \qquad \alpha_k = \frac{\nabla f(x_k)^T \nabla f(x_k)}{\nabla f(x_k)^T A \nabla f(x_k)}$$

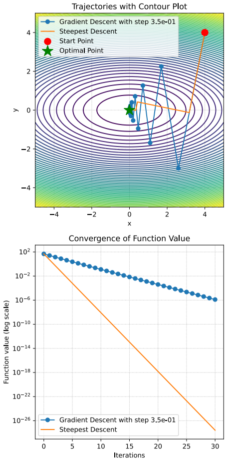
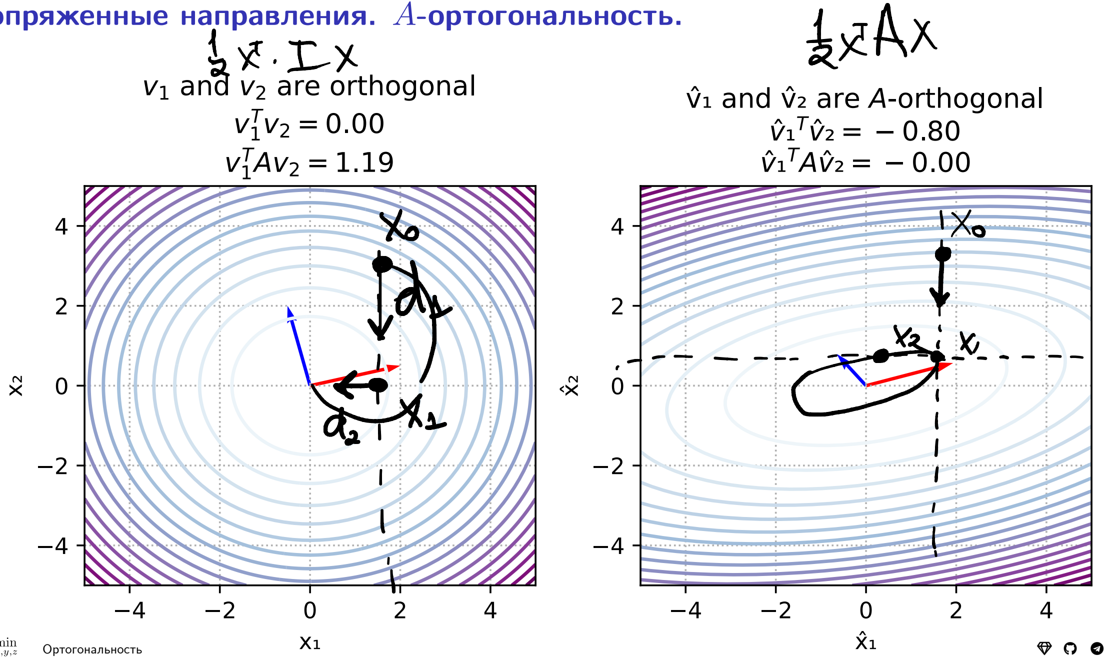


Рис. 1: Наискорейший спуск

Открыть в Colab ♣

Ортогональность

Сопряженные направления. A-ортогональность.



Сопряженные направления. A-ортогональность.

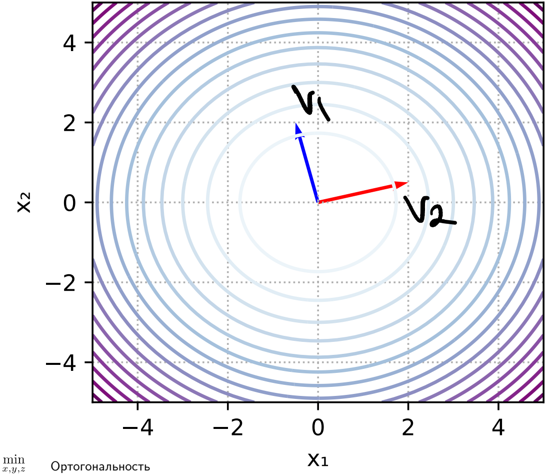
$A = I$ идеальный мир

реальный мир $A \neq I$

v_1 and v_2 are orthogonal

$v_1^T v_2 = 0.00$

$v_1^T A v_2 = 1.19$

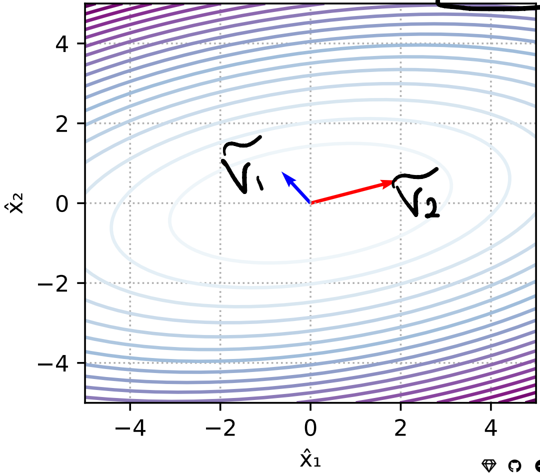


$\hat{v}_1$ and $\hat{v}_2$ are A-orthogonal

$\hat{v}_1^T \hat{v}_2 = -0.80$

$\hat{v}_1^T A \hat{v}_2 = -0.00$

$\boxed{\tilde{v}_1^T A \tilde{v}_2 = 0}$



Сопряженные направления. A-ортогональность.

Предположим, у нас есть две системы координат и квадратичная функция $f(x) = \frac{1}{2}x^T I x$ выглядит так, как на левой части изображения 2, в то время как в других координатах она выглядит как $\boxed{f(\hat{x}) = \frac{1}{2}\hat{x}^T A \hat{x}}$, где $A \in \mathbb{S}^n_{++}$.

$$\frac{1}{2}x^T I x \qquad\qquad \frac{1}{2}\hat{x}^T A \hat{x}$$

Поскольку $\boxed{A = Q\Lambda Q^T}$:

$$\frac{1}{2}\hat{x}^T A \hat{x} = \frac{1}{2}\hat{x}^T Q \Lambda Q^T \hat{x} = \frac{1}{2}\underbrace{\hat{x}^T Q \Lambda^{\frac{1}{2}}}_{x^T}\underbrace{\Lambda^{\frac{1}{2}} Q^T \hat{x}}_{x} = \frac{1}{2}x^T I x$$

Сопряженные направления. A-ортогональность.

Предположим, у нас есть две системы координат и квадратичная функция $f(x) = \frac{1}{2}x^T I x$ выглядит так, как на левой части изображения 2, в то время как в других координатах она выглядит как $f(\hat{x}) = \frac{1}{2}\hat{x}^T A \hat{x}$, где $A \in \mathbb{S}^n_{++}$.

$$
\frac{1}{2}x^T I x \qquad\qquad \frac{1}{2}\hat{x}^T A \hat{x}
$$

Поскольку $A = Q\Lambda Q^T$:

$$
\frac{1}{2}\hat{x}^T A \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda Q^T \hat{x}
$$

Сопряженные направления. A-ортогональность.

Предположим, у нас есть две системы координат и квадратичная функция $f(x) = \frac{1}{2}x^T I x$ выглядит так, как на левой части изображения 2, в то время как в других координатах она выглядит как $f(\hat{x}) = \frac{1}{2}\hat{x}^T A \hat{x}$, где $A \in \mathbb{S}^n_{++}$.

$$\frac{1}{2}x^T I x \qquad\qquad \frac{1}{2}\hat{x}^T A \hat{x}$$

Поскольку $A = Q\Lambda Q^T$:

$$\frac{1}{2}\hat{x}^T A \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda Q^T \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}} Q^T \hat{x}$$

Сопряженные направления. A-ортогональность.

Предположим, у нас есть две системы координат и квадратичная функция $f(x) = \frac{1}{2}x^T I x$ выглядит так, как на левой части изображения 2, в то время как в других координатах она выглядит как $f(\hat{x}) = \frac{1}{2}\hat{x}^T A \hat{x}$, где $A \in \mathbb{S}^n_{++}$.

$$\frac{1}{2}x^T I x \qquad\qquad \frac{1}{2}\hat{x}^T A \hat{x}$$

Поскольку $A = Q\Lambda Q^T$:

$$\frac{1}{2}\hat{x}^T A \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda Q^T \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Q^T \hat{x} = \frac{1}{2}x^T I x$$

Сопряженные направления. A-ортогональность.

Предположим, у нас есть две системы координат и квадратичная функция $f(x) = \frac{1}{2}x^T I x$ выглядит так, как на левой части изображения 2, в то время как в других координатах она выглядит как $f(\hat{x}) = \frac{1}{2}\hat{x}^T A \hat{x}$, где $A \in \mathbb{S}^n_{++}$.

$$\frac{1}{2}x^T I x \qquad\qquad \frac{1}{2}\hat{x}^T A \hat{x}$$

Поскольку $A = Q\Lambda Q^T$:

$$\frac{1}{2}\hat{x}^T A \hat{x} = \frac{1}{2}\hat{x}^T Q \overset{x}{\boxed{\Lambda Q^T \hat{x}}} = \frac{1}{2}\hat{x}^T Q \Lambda^{\frac{1}{2}} \Lambda^{\frac{1}{2}} Q^T \hat{x} = \frac{1}{2}x^T I x \text{ и } \boxed{\hat{x} = Q\Lambda^{-\frac{1}{2}} x}$$

Сопряженные направления. A-ортогональность.

Предположим, у нас есть две системы координат и квадратичная функция $f(x) = \frac{1}{2}x^T I x$ выглядит так, как на левой части изображения 2, в то время как в других координатах она выглядит как $f(\hat{x}) = \frac{1}{2}\hat{x}^T A \hat{x}$, где $A \in \mathbb{S}^n_{++}$.

$$\frac{1}{2}x^T I x \qquad\qquad \frac{1}{2}\hat{x}^T A \hat{x}$$

Поскольку $A = Q\Lambda Q^T$:

$$\frac{1}{2}\hat{x}^T A \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda Q^T \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}} Q^T \hat{x} = \frac{1}{2}x^T I x \text{ и } \hat{x} = Q\Lambda^{-\frac{1}{2}} x$$

Сопряженные направления. A-ортогональность.

Предположим, у нас есть две системы координат и квадратичная функция $f(x) = \frac{1}{2}x^T I x$ выглядит так, как на левой части изображения 2, в то время как в других координатах она выглядит как $f(\hat{x}) = \frac{1}{2}\hat{x}^T A \hat{x}$, где $\boxed{A \in \mathbb{S}^n_{++}.}$

$$\frac{1}{2}x^T I x \qquad\qquad \frac{1}{2}\hat{x}^T A \hat{x}$$

Поскольку $A = Q\Lambda Q^T$:

$$\frac{1}{2}\hat{x}^T A \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda Q^T \hat{x} = \frac{1}{2}\hat{x}^T Q\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}Q^T \hat{x} = \frac{1}{2}x^T I x \text{ и } \hat{x} = Q\Lambda^{-\frac{1}{2}}x$$

A-ортогональные векторы

Векторы $x \in \mathbb{R}^n$ и $y \in \mathbb{R}^n$ называются A-ортогональными (или A-сопряженными), если

$$\boxed{x^T A y = 0} \qquad \Leftrightarrow \qquad x \perp_A y$$

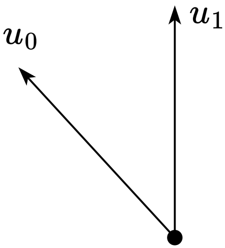
Когда $A = I$, A-ортогональность превращается в ортогональность.

Процесс Грама-Шмидта

n ЛНЗ

Вход: n линейно независимых векторов $\boxed{u_0, \ldots, u_{n-1}}$.

Выход: n линейно независимых попарно ортогональных векторов $\boxed{d_0, \ldots, d_{n-1}}$.



$n=2$

Рис. 3: Иллюстрация процесса Грама-Шмидта

Процесс Грама-Шмидта

Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.

Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

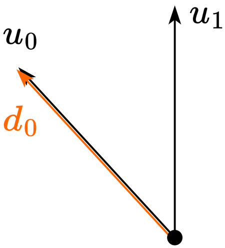


Рис. 4: Иллюстрация процесса Грама-Шмидта

Процесс Грама-Шмидта

Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.

Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

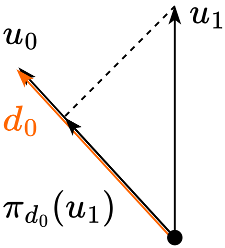


Рис. 5: Иллюстрация процесса Грама-Шмидта

Процесс Грама-Шмидта

Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.

Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

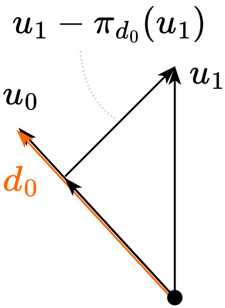


Рис. 6: Иллюстрация процесса Грама-Шмидта

Процесс Грама-Шмидта

Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.

Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

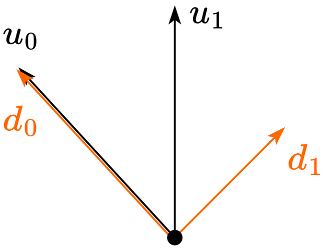
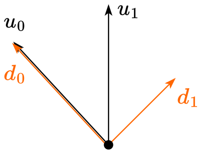
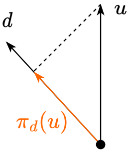


Рис. 7: Иллюстрация процесса Грама-Шмидта

Процесс Грама-Шмидта

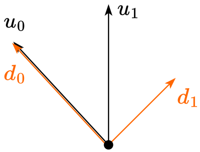
Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.

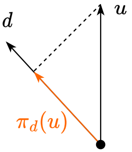




$$\pi_d(u) = \frac{\langle d, u \rangle}{\|d\|_2^2} d$$

Процесс Грама-Шмидта



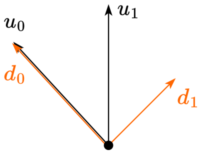


$$\pi_d(u) = \frac{\langle d, u\rangle}{\|d\|_2^2} d$$

Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.
Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

$$d_0 = u_0$$

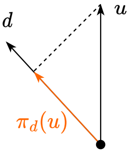
Процесс Грама-Шмидта



Вход: n линейно независимых векторов $u_0, \dots, u_{n-1}$.
Выход: n линейно независимых попарно ортогональных векторов $d_0, \dots, d_{n-1}$.

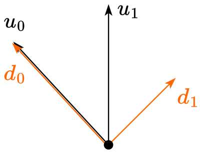
$$d_0 = u_0$$
$$d_1 = u_1 - \pi_{d_0}(u_1)$$

$$d_2 = u_2 - \pi_{d_0}(u_2) - \pi_{d_1}(u_2)$$



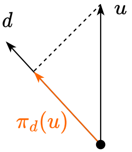
$$\pi_d(u) = \frac{\langle d, u\rangle}{\|d\|_2^2} d$$

Процесс Грама-Шмидта



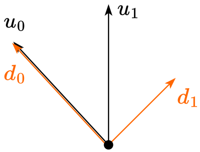
Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.
Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

$$
\begin{aligned}
d_0 &= u_0 \\
d_1 &= u_1 - \pi_{d_0}(u_1) \\
d_2 &= u_2 - \pi_{d_0}(u_2) - \pi_{d_1}(u_2)
\end{aligned}
$$



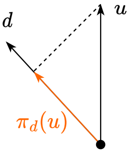
$$
\pi_d(u) = \frac{\langle d, u \rangle}{\|d\|_2^2} d
$$

Процесс Грама-Шмидта



Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.
Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

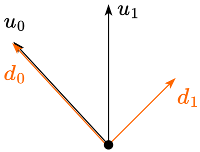
$$
\begin{aligned}
d_0 &= u_0 \\
d_1 &= u_1 - \pi_{d_0}(u_1) \\
d_2 &= u_2 - \pi_{d_0}(u_2) - \pi_{d_1}(u_2) \\
&\vdots
\end{aligned}
$$



$$\pi_d(u) = \frac{\langle d, u\rangle}{\|d\|_2^2} d$$

Процесс Грама-Шмидта

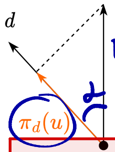
Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.
Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.



$a^T b = \|a\| \cdot \|b\| \cdot \cos\alpha$

$$
\begin{aligned}
d_0 &= u_0 \\
d_1 &= u_1 - \pi_{d_0}(u_1) \\
d_2 &= u_2 - \pi_{d_0}(u_2) - \pi_{d_1}(u_2) \\
&\vdots
\end{aligned}
$$

$$
\boxed{d_k = u_k - \sum_{i=0}^{k-1} \pi_{d_i}(u_k)}
$$

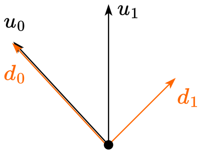


$\Pi = u \cdot \cos\alpha = \cancel{\|u\|} \cdot \frac{u^T d}{\cancel{\|u\|} \cdot \|d\|} \cdot \frac{d}{\|d\|}$

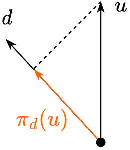
$$
\boxed{\pi_d(u) = \frac{\langle d, u \rangle}{\|d\|_2^2} d}
$$

Процесс Грама-Шмидта

Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.
Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

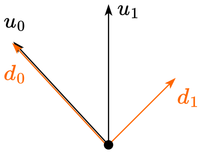


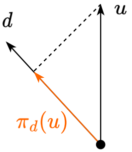
$$
\begin{aligned}
d_0 &= u_0 \\
d_1 &= u_1 - \pi_{d_0}(u_1) \\
d_2 &= u_2 - \pi_{d_0}(u_2) - \pi_{d_1}(u_2) \\
&\vdots \\
d_k &= u_k - \sum_{i=0}^{k-1} \pi_{d_i}(u_k)
\end{aligned}
$$



$$
\pi_d(u) = \frac{\langle d, u \rangle}{\|d\|_2^2} d
$$

Процесс Грама-Шмидта





$$\pi_d(u) = \frac{\langle d, u\rangle}{\|d\|_2^2} d$$

Вход: n линейно независимых векторов $u_0, \ldots, u_{n-1}$.
Выход: n линейно независимых попарно ортогональных векторов $d_0, \ldots, d_{n-1}$.

$$\begin{aligned}
d_0 &= u_0 \\
d_1 &= u_1 - \pi_{d_0}(u_1) \\
d_2 &= u_2 - \pi_{d_0}(u_2) - \pi_{d_1}(u_2) \\
&\vdots \\
d_k &= u_k - \sum_{i=0}^{k-1} \pi_{d_i}(u_k)
\end{aligned}$$

$$\boxed{d_k = u_k + \sum_{i=0}^{k-1} \beta_{ik} d_i \qquad \beta_{ik} = -\frac{\langle d_i, u_k\rangle}{\langle d_i, d_i\rangle}} \tag{2}$$

Метод сопряженных направлений (CD)

Общая идея

выбор длины шага

- В изотропном случае $A = I$ метод наискорейшего спуска, запущенный из произвольной точки в n ортогональных линейно независимых направлениях, сойдется за n шагов в точных арифметических вычислениях. Мы пытаемся построить аналогичную процедуру в случае $A \neq I$ с использованием концепции A-ортогональности.

Общая идея

- В изотропном случае $A = I$ метод наискорейшего спуска, запущенный из произвольной точки в n ортогональных линейно независимых направлениях, сойдется за n шагов в точных арифметических вычислениях. Мы пытаемся построить аналогичную процедуру в случае $A \neq I$ с использованием концепции A-ортогональности.
- Предположим, у нас есть набор из n линейно независимых A-ортогональных направлений $d_0, \ldots, d_{n-1}$ (которые будут вычислены с помощью процесса Грама-Шмидта).

Общая идея

- В изотропном случае $A = I$ метод наискорейшего спуска, запущенный из произвольной точки в n ортогональных линейно независимых направлениях, сойдется за n шагов в точных арифметических вычислениях. Мы пытаемся построить аналогичную процедуру в случае $A \neq I$ с использованием концепции A-ортогональности.
- Предположим, у нас есть набор из n линейно независимых A-ортогональных направлений $d_0, \ldots, d_{n-1}$ (которые будут вычислены с помощью процесса Грама-Шмидта).
- Мы хотим построить метод, который идет из x_0 в x^* для квадратичной задачи с шагами α_i, который, фактически, является разложением $x^* - x_0$ в некотором базисе:

$$
x^* = x_0 + \sum_{i=0}^{n-1} \alpha_i d_i \qquad x^* - x_0 = \sum_{i=0}^{n-1} \alpha_i d_i
$$

Общая идея

- В изотропном случае $A = I$ метод наискорейшего спуска, запущенный из произвольной точки в n ортогональных линейно независимых направлениях, сойдется за n шагов в точных арифметических вычислениях. Мы пытаемся построить аналогичную процедуру в случае $A \neq I$ с использованием концепции A-ортогональности.
- Предположим, у нас есть набор из n линейно независимых A-ортогональных направлений $d_0, \ldots, d_{n-1}$ (которые будут вычислены с помощью процесса Грама-Шмидта).
- Мы хотим построить метод, который идет из x_0 в x^* для квадратичной задачи с шагами α_i, который, фактически, является разложением $x^* - x_0$ в некотором базисе:

$$x^* - x_0 = \sum_{i=0}^{n-1} \alpha_i d_i \qquad \boxed{x^* = x_0 + \sum_{i=0}^{n-1} \alpha_i d_i} \qquad x^* - x_0 = \sum_{i=0}^{n-1} \alpha_i d_i$$

- Мы докажем, что α_i и d_i могут быть построены очень эффективно с вычислительной точки зрения (метод сопряженных градиентов).

Идея метода сопряженных направлений (CD)

Таким образом, мы формулируем алгоритм:

1. Пусть $k = 0$ и $x_k = x_0$, посчитаем $d_k = d_0 = -\nabla f(x_0)$.

Идея метода сопряженных направлений (CD)

Таким образом, мы формулируем алгоритм:

1. Пусть $k = 0$ и $x_k = x_0$, посчитаем $d_k = d_0 = -\nabla f(x_0)$.
2. С помощью процедуры точного линейного поиска находим оптимальную длину шага. Вычисляем α минимизируя $f(x_k + \alpha_k d_k)$ по формуле

$$\boxed{\alpha_k = -\frac{d_k^\top (Ax_k - b)}{d_k^\top A d_k}} \qquad \alpha_k = \frac{g_k^T g_k}{g_k^T A g_k} \tag{3}$$

$$-\frac{-g_k^T g_k}{g_k^T A g_k} = \frac{g_k^T g_k}{g_k^T A g_k} \qquad d_k = -g_k$$

Идея метода сопряженных направлений (CD)

Таким образом, мы формулируем алгоритм:

1. Пусть $k = 0$ и $x_k = x_0$, посчитаем $d_k = d_0 = -\nabla f(x_0)$.
2. С помощью процедуры точного линейного поиска находим оптимальную длину шага. Вычисляем α минимизируя $f(x_k + \alpha_k d_k)$ по формуле

$$\alpha_k = -\frac{d_k^\top (Ax_k - b)}{d_k^\top A d_k} \tag{3}$$

3. Выполняем шаг алгоритма:

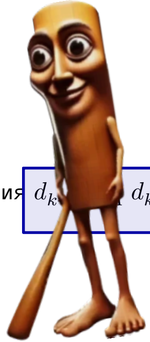
$$x_{k+1} = x_k + \alpha_k d_k$$

Идея метода сопряженных направлений (CD)

Таким образом, мы формулируем алгоритм:

1. Пусть $k = 0$ и $x_k = x_0$, посчитаем $d_k = d_0 = -\nabla f(x_0)$.
2. С помощью процедуры точного линейного поиска находим оптимальную длину шага. Вычисляем α минимизируя $f(x_k + \alpha_k d_k)$ по формуле

$$\alpha_k = -\frac{d_k^\top (Ax_k - b)}{d_k^\top A d_k} \tag{3}$$



BAD IDEA

3. Выполняем шаг алгоритма:

$$x_{k+1} = x_k + \alpha_k d_k$$

4. Обновляем направление: $\boxed{d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k}$ в целях сохранения d_k … d_k, где β_k вычисляется по формуле:

$$\beta_k = \frac{\nabla f(x_{k+1})^\top A d_k}{d_k^\top A d_k}.$$

$(-g_{k+1} + \beta_k d_k)^\top d_k = 0$

Идея метода сопряженных направлений (CD)

Таким образом, мы формулируем алгоритм:

1. Пусть $k = 0$ и $x_k = x_0$, посчитаем $d_k = d_0 = -\nabla f(x_0)$.
2. С помощью процедуры точного линейного поиска находим оптимальную длину шага. Вычисляем α минимизируя $f(x_k + \alpha_k d_k)$ по формуле
$$\alpha_k = -\frac{d_k^\top (A x_k - b)}{d_k^\top A d_k} \tag{3}$$
3. Выполняем шаг алгоритма:
$$x_{k+1} = x_k + \alpha_k d_k$$
4. Обновляем направление: $d_{k+1} = -\nabla f(x_{k+1}) + \beta_k d_k$ в целях сохранения $d_{k+1} \perp_A d_k$, где β_k вычисляется по формуле:
$$\beta_k = \frac{\nabla f(x_{k+1})^\top A d_k}{d_k^\top A d_k}.$$
5. Повторяем шаги 2-4, пока не построим n направлений, где n - размерность пространства (x).

Метод сопряженных направлений (CD)

Лемма 1. Линейная независимость A-ортогональных векторов.

Если множество векторов $\boxed{d_1, \ldots, d_n}$ - попарно A-ортогональны (каждая пара векторов A-ортогональна), то эти векторы линейно независимы. $A \in \mathbb{S}^n_{++}$.

Метод сопряженных направлений (CD)

Лемма 1. Линейная независимость A-ортогональных векторов.

Если множество векторов $d_1, \ldots, d_n$ - попарно A-ортогональны (каждая пара векторов A-ортогональна), то эти векторы линейно независимы. $A \in \mathbb{S}^n_{++}$.

Доказательство

Покажем, что если $\sum\limits_{i=1}^{n} \alpha_i d_i = 0$, то все коэффициенты должны быть равны нулю:

Метод сопряженных направлений (CD)

Лемма 1. Линейная независимость A-ортогональных векторов.

Если множество векторов $d_1, \ldots, d_n$ - попарно A-ортогональны (каждая пара векторов A-ортогональна), то эти векторы линейно независимы. $A \in \mathbb{S}^n_{++}$.

Доказательство предположим $\exists\, \alpha \neq 0$

Покажем, что если $\boxed{\sum_{i=1}^n \alpha_i d_i = 0,}$ то все коэффициенты должны быть равны нулю:

$$d_k^T A \cdot \Big| \qquad 0 = \sum_{i=1}^n \alpha_i d_i$$

$$0 = \sum_{i=1}^n \alpha_i d_k^T A d_i = \alpha_k d_k^T A d_k$$

$$\alpha_k = 0$$

Метод сопряженных направлений (CD)

Лемма 1. Линейная независимость A-ортогональных векторов.

Если множество векторов $d_1, \ldots, d_n$ - попарно A-ортогональны (каждая пара векторов A-ортогональна), то эти векторы линейно независимы. $A \in \mathbb{S}^n_{++}$.

Доказательство

Покажем, что если $\sum\limits_{i=1}^n \alpha_i d_i = 0$, то все коэффициенты должны быть равны нулю:

$$0 = \sum_{i=1}^n \alpha_i d_i$$

$$\text{Умножаем на } d_j^T A \cdot \qquad = d_j^\top A \left(\sum_{i=1}^n \alpha_i d_i \right)$$

Метод сопряженных направлений (CD)

Лемма 1. Линейная независимость A-ортогональных векторов.

Если множество векторов $d_1, \ldots, d_n$ - попарно A-ортогональны (каждая пара векторов A-ортогональна), то эти векторы линейно независимы. $A \in \mathbb{S}^n_{++}$.

Доказательство

Покажем, что если $\sum\limits_{i=1}^n \alpha_i d_i = 0$, то все коэффициенты должны быть равны нулю:

$$0 = \sum_{i=1}^n \alpha_i d_i$$

$$\text{Умножаем на } d_j^T A \cdot \qquad = d_j^\top A \left(\sum_{i=1}^n \alpha_i d_i \right) = \sum_{i=1}^n \alpha_i d_j^\top A d_i$$

Метод сопряженных направлений (CD)

Лемма 1. Линейная независимость A-ортогональных векторов.

Если множество векторов $d_1, \ldots, d_n$ - попарно A-ортогональны (каждая пара векторов A-ортогональна), то эти векторы линейно независимы. $A \in \mathbb{S}^n_{++}$.

Доказательство

Покажем, что если $\sum\limits_{i=1}^n \alpha_i d_i = 0$, то все коэффициенты должны быть равны нулю:

$$
\begin{aligned}
0 &= \sum_{i=1}^n \alpha_i d_i \\
\text{Умножаем на } d_j^T A \cdot \qquad &= d_j^\top A \left(\sum_{i=1}^n \alpha_i d_i \right) = \sum_{i=1}^n \alpha_i d_j^\top A d_i \\
&= \alpha_j d_j^\top A d_j + 0 + \ldots + 0
\end{aligned}
$$

Метод сопряженных направлений (CD)

Лемма 1. Линейная независимость A-ортогональных векторов.

Если множество векторов $d_1, \ldots, d_n$ - попарно A-ортогональны (каждая пара векторов A-ортогональна), то эти векторы линейно независимы. $A \in \mathbb{S}^n_{++}$.

Доказательство

Покажем, что если $\sum\limits_{i=1}^n \alpha_i d_i = 0$, то все коэффициенты должны быть равны нулю:

$$
\begin{aligned}
0 &= \sum_{i=1}^n \alpha_i d_i \\
\text{Умножаем на } d_j^T A \cdot \qquad &= d_j^\top A \left(\sum_{i=1}^n \alpha_i d_i \right) = \sum_{i=1}^n \alpha_i d_j^\top A d_i \\
&= \alpha_j d_j^\top A d_j + 0 + \ldots + 0
\end{aligned}
$$

Метод сопряженных направлений (CD)

Лемма 1. Линейная независимость A-ортогональных векторов.

Если множество векторов $d_1, \ldots, d_n$ - попарно A-ортогональны (каждая пара векторов A-ортогональна), то эти векторы линейно независимы. $A \in \mathbb{S}^n_{++}$.

Доказательство

Покажем, что если $\sum\limits_{i=1}^n \alpha_i d_i = 0$, то все коэффициенты должны быть равны нулю:

$$0 = \sum_{i=1}^n \alpha_i d_i$$

$$\text{Умножаем на } d_j^T A \cdot \qquad = d_j^\top A \left(\sum_{i=1}^n \alpha_i d_i \right) = \sum_{i=1}^n \alpha_i d_j^\top A d_i$$

$$= \alpha_j d_j^\top A d_j + 0 + \ldots + 0$$

Таким образом, $\alpha_j = 0$, для всех остальных индексов нужно проделать тот же процесс

Доказательство сходимости

Введем следующие обозначения:

- $\boxed{r_k = b - Ax_k}$ - невязка

$r_k = -\nabla f(x_k)$

Доказательство сходимости

Введем следующие обозначения:

- $r_k = b - Ax_k$ - невязка
- $\boxed{e_k = x_k - x^* \text{ - ошибка}}$

$\|e_k\|$

Доказательство сходимости

residual

Введем следующие обозначения:

- $r_k = b - Ax_k$ - невязка
- $e_k = x_k - x^*$ - ошибка
- Поскольку $\boxed{Ax^* = b}$, имеем $\boxed{r_k = b - Ax_k} = Ax^* - Ax_k = -A(x_k - x^*)$

$$\boxed{r_k = -Ae_k.} \tag{4}$$

Доказательство сходимости

Введем следующие обозначения:

- $r_k = b - Ax_k$ - невязка
- $e_k = x_k - x^*$ - ошибка
- Поскольку $Ax^* = b$, имеем $r_k = b - Ax_k = Ax^* - Ax_k = -A(x_k - x^*)$

$$r_k = -Ae_k. \tag{4}$$

- Также заметим, что поскольку $\boxed{x_k = x_0 + \sum\limits_{i=1}^{k} \alpha_i d_i,}$ имеем

$$e_k = e_0 + \sum_{i=1}^{k} \alpha_i d_i. \tag{5}$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$
\boxed{x_{k\cancel{+1}} = x_0 + \sum_{i=0}^{k} \alpha_i d_i}
$$

$X_n = x^*$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, A d_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

наиск. спуск в напр. d_i

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$\boxed{e_0 = x_0 - x^*} = \sum_{i=0}^{n-1} \delta_i d_i$$

коэф. разл. e_0 в ЛНЗ систему d_i

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, A d_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

2. Умножаем обе части слева на $d_k^T A$:

$$d_k^T A e_0 = \underline{\sum_{i=0}^{n-1} \delta_i \cdot d_k^T A d_i} =$$

$$d_k^T \underbrace{A e_0} \qquad = \delta_k d_k^T A d_k$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

2. Умножаем обе части слева на $d_k^T A$:

$$d_k^T A e_0 = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$
x_{k} = x_0 + \sum_{i=0}^{k-1} \alpha_i d_i
$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$
e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i
$$

2. Умножаем обе части слева на $d_k^T A$:

$$
d_k^T A e_0 = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i = \delta_k d_k^T A d_k
$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

2. Умножаем обе части слева на $d_k^T A$:

$$d_k^T A \left(\sum_{i=0}^{k-1} \alpha_i d_i \right) = 0$$

$$d_k^T A e_0 = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i = \delta_k d_k^T A d_k$$

$$=$$

$$\underbrace{d_k^T A}\left(e_0 + \underbrace{\sum_{i=0}^{k-1} \alpha_i d_i} \right)$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$
x_{k} = x_0 + \sum_{i=0}^{k-1} \alpha_i d_i \qquad e_k = e_0 + \sum_{i=0}^{k-1} \alpha_i d_i
$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$
e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i
$$

$$
X_1 = X_0 + \alpha_0 d_0
$$

$$
X_2 = X_1 + \alpha_1 d_1
$$

2. Умножаем обе части слева на $d_k^T A$:

Итог

$$
d_k^T A e_0 = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i = \delta_k d_k^T A d_k
$$

$$
d_k^T A \underbrace{\left(e_0 + \sum_{i=0}^{k-1} \alpha_i d_i \right)} = d_k^T A \underbrace{e_k}
$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\boxed{\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\boxed{\delta_i = -\alpha_i}$:

$$\boxed{e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i}$$

2. Умножаем обе части слева на $d_k^T A$:

$$d_k^T A e_0 = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i = \delta_k d_k^T A d_k$$

$$\underline{d_k^T A \left(e_0 + \sum_{i=0}^{k-1} \alpha_i d_i \right)} = \boxed{d_k^T A e_k = \delta_k d_k^T A d_k} \quad (A - \text{ ортогональность})$$

$$\delta_k = \frac{d_k^T A e_k}{d_k^T A d_k} = -\frac{d_k^T r_k}{d_k^T A d_k}$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

2. Умножаем обе части слева на $d_k^T A$:

$$d_k^T A e_0 = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i = \delta_k d_k^T A d_k$$

$$d_k^T A \left(e_0 + \sum_{i=0}^{k-1} \alpha_i d_i \right) = d_k^T A e_k = \delta_k d_k^T A d_k \quad (A - \text{ ортогональность})$$

$$\delta_k = \frac{d_k^T A e_k}{d_k^T A d_k}$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

2. Умножаем обе части слева на $d_k^T A$:

$$d_k^T A e_0 = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i = \delta_k d_k^T A d_k$$

$$d_k^T A \left(e_0 + \sum_{i=0}^{k-1} \alpha_i d_i \right) = d_k^T A e_k = \delta_k d_k^T A d_k \quad (A- \text{ ортогональность})$$

$$\delta_k = \frac{d_k^T A e_k}{d_k^T A d_k} = -\frac{d_k^T r_k}{d_k^T A d_k}$$

Доказательство сходимости

Лемма 2. Сходимость метода сопряженных направлений.

Предположим, мы решаем n-мерную квадратичную сильно выпуклую задачу оптимизации (1). Метод сопряженных направлений

$$x_{k+1} = x_0 + \sum_{i=0}^{k} \alpha_i d_i$$

с $\alpha_i = \frac{\langle d_i, r_i \rangle}{\langle d_i, Ad_i \rangle}$ взятым из точного линейного поиска, сходится за не более n шагов алгоритма.

Доказательство

1. Нужно доказать, что $\delta_i = -\alpha_i$:

$$e_0 = x_0 - x^* = \sum_{i=0}^{n-1} \delta_i d_i$$

2. Умножаем обе части слева на $d_k^T A$:

$$d_k^T A e_0 = \sum_{i=0}^{n-1} \delta_i d_k^T A d_i = \delta_k d_k^T A d_k$$

$$d_k^T A \left(e_0 + \sum_{i=0}^{k-1} \alpha_i d_i \right) = d_k^T A e_k = \delta_k d_k^T A d_k \quad (A- \text{ ортогональность})$$

$$\delta_k = \frac{d_k^T A e_k}{d_k^T A d_k} = -\frac{d_k^T r_k}{d_k^T A d_k} \Leftrightarrow \delta_k = -\alpha_k$$

Леммы для сходимости

$x_0 + \alpha_0 d_0 + \alpha_1 d_1 + \ldots \alpha_{n-1} d_{n-1} = x^*$

$\alpha_0 d_0 + \ldots + \alpha_{n-1} d_{n-1} = x^* - x_0 = -e_0$

Лемма 3. Разложение ошибки.

$$\boxed{e_i = \sum_{j=i}^{n-1} -\alpha_j d_j} \tag{6}$$

Леммы для сходимости

Лемма 3. Разложение ошибки.

$$e_i = \sum_{j=i}^{n-1} -\alpha_j d_j \tag{6}$$

Доказательство

По определению

$$\boxed{e_i = e_0 + \sum_{j=0}^{i-1} \alpha_j d_j}$$

$$x_i = x_0 + \sum_{j=0}^{i-1} \alpha_j d_j$$

Леммы для сходимости

Лемма 3. Разложение ошибки.

$$e_i = \sum_{j=i}^{n-1} -\alpha_j d_j \tag{6}$$

Доказательство

По определению

$$e_i = e_0 + \sum_{j=0}^{i-1} \alpha_j d_j = x_0 - x^* + \sum_{j=0}^{i-1} \alpha_j d_j$$

Леммы для сходимости

$$x_0 + \sum_{j=0}^{n-1} \alpha_j d_j = x^*$$

Лемма 3. Разложение ошибки.

$$e_i = \sum_{j=i}^{n-1} -\alpha_j d_j \tag{6}$$

Доказательство

По определению

$$e_i = e_0 + \sum_{j=0}^{i-1} \alpha_j d_j = \boxed{x_0 - x^*} + \sum_{j=0}^{i-1} \alpha_j d_j = \boxed{-\sum_{j=0}^{n-1} \alpha_j d_j} + \sum_{j=0}^{i-1} \alpha_j d_j$$

Леммы для сходимости

Лемма 3. Разложение ошибки.

$$
e_i = \sum_{j=i}^{n-1} -\alpha_j d_j \tag{6}
$$

Доказательство

По определению

$$
e_i = e_0 + \sum_{j=0}^{i-1} \alpha_j d_j = x_0 - x^* + \sum_{j=0}^{i-1} \alpha_j d_j = -\sum_{j=0}^{n-1} \alpha_j d_j + \sum_{j=0}^{i-1} \alpha_j d_j = \sum_{j=i}^{n-1} -\alpha_j d_j
$$

Леммы для сходимости

conjugate direction (CD)

Лемма 4. Невязка ортогональна всем предыдущим направлениям для CD.

Рассмотрим невязку метода сопряженных направлений на k итерации r_k, тогда для любого $i < k$:

$$d_i^T r_k = 0 \tag{7}$$

Леммы для сходимости

Лемма 4. Невязка ортогональна всем предыдущим направлениям для CD.

Рассмотрим невязку метода сопряженных направлений на k итерации r_k, тогда для любого $i < k$:

$$d_i^T r_k = 0 \tag{7}$$

Доказательство

Запишем (6) для некоторого фиксированного индекса k:

Леммы для сходимости

Лемма 4. Невязка ортогональна всем предыдущим направлениям для CD.

Рассмотрим невязку метода сопряженных направлений на k итерации r_k, тогда для любого $i < k$:

$$d_i^T r_k = 0 \tag{7}$$

Доказательство

Запишем (6) для некоторого фиксированного индекса k:

$$\boxed{e_k = \sum_{j=k}^{n-1} -\alpha_j d_j}$$

Леммы для сходимости

Лемма 4. Невязка ортогональна всем предыдущим направлениям для CD.

Рассмотрим невязку метода сопряженных направлений на k итерации r_k, тогда для любого $i < k$:

$$d_i^T r_k = 0 \tag{7}$$

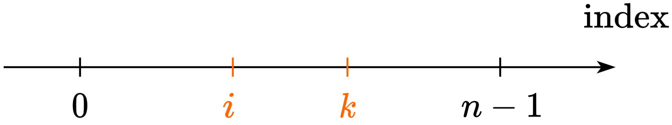
Доказательство

Запишем (6) для некоторого фиксированного индекса k:

$$e_k = \sum_{j=k}^{n-1} -\alpha_j d_j$$

Умножаем обе части на $\boxed{-d_i^T A \cdot}$

$$\underset{\displaystyle r_k}{\underset{\|}{-d_i^T A e_k}} = \underbrace{\sum_{j=k}^{n-1} \alpha_j d_i^T A d_j = 0}$$



Таким образом, $d_i^T r_k = 0$ и невязка r_k ортогональна всем предыдущим направлениям d_i для метода CD.

ОТКУДА d_k?

Из модифиц. процедуры Грама Шмидта (CD)

Метод сопряженных градиентов (CD)

Пусть есть $d_0 \ldots\ldots d_{n-1}$ - n A сопр. векторов

$$x_{k+1} = x_k + \alpha_k \cdot d_k, \quad \text{где } \alpha_k = \frac{\langle d_k, r_k \rangle}{\langle d_k, A d_k \rangle}$$

Метод сопряженных градиентов (CG)

Идея метода сопряженных градиентов (CG)

- Это буквально метод сопряженных направлений, в котором мы выбираем специальный набор $d_0, \ldots, d_{n-1}$, позволяющий значительно ускорить процесс Грама-Шмидта.

Идея метода сопряженных градиентов (CG)

- Это буквально метод сопряженных направлений, в котором мы выбираем специальный набор $d_0, \ldots, d_{n-1}$, позволяющий значительно ускорить процесс Грама-Шмидта.
- Используется процесс Грама-Шмидта с A-ортогональностью вместо Евклидовой ортогональности, чтобы получить их из набора начальных векторов.

Идея метода сопряженных градиентов (CG)

- Это буквально метод сопряженных направлений, в котором мы выбираем специальный набор $d_0, \ldots, d_{n-1}$, позволяющий значительно ускорить процесс Грама-Шмидта.
- Используется процесс Грама-Шмидта с A-ортогональностью вместо Евклидовой ортогональности, чтобы получить их из набора начальных векторов.
- На каждой итерации $r_0, \ldots, r_{n-1}$ используются в качестве начальных векторов для процесса Грама-Шмидта.

Идея метода сопряженных градиентов (CG)

- Это буквально метод сопряженных направлений, в котором мы выбираем специальный набор $d_0, \ldots, d_{n-1}$, позволяющий значительно ускорить процесс Грама-Шмидта.
- Используется процесс Грама-Шмидта с A-ортогональностью вместо Евклидовой ортогональности, чтобы получить их из набора начальных векторов.
- На каждой итерации $r_0, \ldots, r_{n-1}$ используются в качестве начальных векторов для процесса Грама-Шмидта.
- Основная идея заключается в том, что для произвольного метода CD процесс Грама-Шмидта вычислительно дорогой и требует квадратичного числа операций сложения векторов и скалярных произведений $\mathcal{O}\left(n^2\right)$, в то время как в случае CG мы покажем, что сложность этой процедуры может быть уменьшена до линейной $\mathcal{O}\left(n\right)$.

Идея метода сопряженных градиентов (CG)

- Это буквально метод сопряженных направлений, в котором мы выбираем специальный набор $d_0, \ldots, d_{n-1}$, позволяющий значительно ускорить процесс Грама-Шмидта.
- Используется процесс Грама-Шмидта с A-ортогональностью вместо Евклидовой ортогональности, чтобы получить их из набора начальных векторов.
- На каждой итерации $r_0, \ldots, r_{n-1}$ используются в качестве начальных векторов для процесса Грама-Шмидта.
- Основная идея заключается в том, что для произвольного метода CD процесс Грама-Шмидта вычислительно дорогой и требует квадратичного числа операций сложения векторов и скалярных произведений $\mathcal{O}\left(n^2\right)$, в то время как в случае CG мы покажем, что сложность этой процедуры может быть уменьшена до линейной $\mathcal{O}\left(n\right)$.

Идея метода сопряженных градиентов (CG)

- Это буквально метод сопряженных направлений, в котором мы выбираем специальный набор $d_0, \ldots, d_{n-1}$, позволяющий значительно ускорить процесс Грама-Шмидта.
- Используется процесс Грама-Шмидта с A-ортогональностью вместо Евклидовой ортогональности, чтобы получить их из набора начальных векторов.
- На каждой итерации $r_0, \ldots, r_{n-1}$ используются в качестве начальных векторов для процесса Грама-Шмидта.
- ~~Основная идея заключается в том, что для произвольного метода CD процесс Грама-Шмидта вычислительно дорогой и требует квадратичного числа операций сложения векторов и скалярных произведений $\mathcal{O}\left(n^2\right)$, в то время как в случае CG мы покажем, что сложность этой процедуры может быть уменьшена до линейной $\mathcal{O}\left(n\right)$.~~

$$\text{CG} = \text{CD} + \underline{r_0, \ldots, r_{n-1}} \text{ как начальные векторы для процесса Грама-Шмидта} + A\text{-ортогональность.}$$

Метод сопряженных градиентов (CG)

Леммы для сходимости

Лемма 5. Невязки ортогональны друг другу в методе CG

Все невязки в методе CG ортогональны друг другу:

$$\boxed{r_i^T r_k = 0 \qquad \forall i \neq k} \tag{8}$$

Леммы для сходимости

Лемма 5. Невязки ортогональны друг другу в методе CG

Все невязки в методе CG ортогональны друг другу:

$$r_i^T r_k = 0 \qquad \forall i \neq k \tag{8}$$

Доказательство

Запишем процесс Грама-Шмидта (2) с $\langle \cdot, \cdot \rangle$ замененным на $\langle x, y \rangle_A = x^T A y$

Леммы для сходимости

Лемма 5. Невязки ортогональны друг другу в методе CG

Все невязки в методе CG ортогональны друг другу:

$$r_i^T r_k = 0 \qquad \forall i \neq k \tag{8}$$

Доказательство

Запишем процесс Грама-Шмидта (2) с $\langle \cdot, \cdot \rangle$ замененным на $\langle \cdot, \cdot \rangle_A = x^T A y$

$$\boxed{d_i = u_i + \sum_{j=0}^{k-1} \beta_{ji} d_j} \boxed{\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A}} \tag{9}$$

Леммы для сходимости

Лемма 5. Невязки ортогональны друг другу в методе CG

Все невязки в методе CG ортогональны друг другу:

$$r_i^T r_k = 0 \qquad \forall i \neq k \tag{8}$$

Доказательство

Запишем процесс Грама-Шмидта (2) с $\langle \cdot, \cdot \rangle$ замененным на $\langle \cdot, \cdot \rangle_A = x^T A y$

$$d_i = u_i + \sum_{j=0}^{k-1} \beta_{ji} d_j \quad \beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} \tag{9}$$

CG

Тогда, мы используем невязки в качестве начальных векторов для процесса и $\boxed{u_i = r_i.}$

Леммы для сходимости

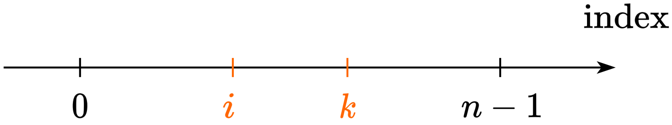
Лемма 5. Невязки ортогональны друг другу в методе CG

Все невязки в методе CG ортогональны друг другу:

$$
r_i^T r_k = 0 \qquad \forall i \neq k \tag{8}
$$

Доказательство

Запишем процесс Грама-Шмидта (2) с $\langle \cdot, \cdot \rangle$ замененным на $\langle \cdot, \cdot \rangle_A = x^T A y$



$$
d_i = u_i + \sum_{j=0}^{k-1} \beta_{ji} d_j \quad \beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} \tag{9}
$$

Тогда, мы используем невязки в качестве начальных векторов для процесса и $u_i = r_i$.

$$
\boxed{d_i = r_i + \sum_{j=0}^{k-1} \beta_{ji} d_j} \quad \beta_{ji} = -\frac{\langle d_j, r_i \rangle_A}{\langle d_j, d_j \rangle_A} \tag{10}
$$

Умножаем обе части (9) на $r_k^T \cdot$ для некоторого индекса k:

$$
r_k^T d_i = r_k^T u_i + \sum_{j=0}^{k-1} \beta_{ji} r_k^T d_j
$$

$$
r_k^T r_i + \sum_{j=0}^{k-1} \beta_{ji} r_k^T d_j
$$

Леммы для сходимости

Лемма 5. Невязки ортогональны друг другу в методе CG

Все невязки в методе CG ортогональны друг другу:

$$
r_i^T r_k = 0 \qquad \forall i \neq k \tag{8}
$$

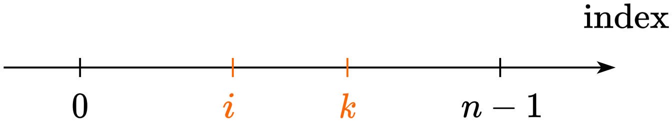
Доказательство

Запишем процесс Грама-Шмидта (2) с $\langle \cdot, \cdot \rangle$ замененным на $\langle \cdot, \cdot \rangle_A = x^T A y$

$$
d_i = u_i + \sum_{j=0}^{k-1} \beta_{ji} d_j \quad \beta_{ji} = - \frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} \tag{9}
$$

Тогда, мы используем невязки в качестве начальных векторов для процесса и $u_i = r_i$.

$$
d_i = r_i + \sum_{j=0}^{k-1} \beta_{ji} d_j \quad \beta_{ji} = - \frac{\langle d_j, r_i \rangle_A}{\langle d_j, d_j \rangle_A} \tag{10}
$$



Умножаем обе части (9) на $r_k^T \cdot$ для некоторого индекса k:

$$
\boxed{r_k^T d_i = r_k^T u_i + \sum_{j=0}^{k-1} \beta_{ji} r_k^T d_j}
$$

Если $j < i < k$, то имеем лемму 4 с $d_i^T r_k = 0$ и $d_j^T r_k = 0$. Имеем:

$\boxed{r_k^T u_i = 0}$ для CD $\boxed{r_k^T r_i = 0 \text{ для CG}}$

Леммы для сходимости

CD

Более того, если $k = i$:

$$r_k^T d_k = r_k^T u_k + \sum_{j=0}^{k-1} \beta_{jk} r_k^T d_j$$

Леммы для сходимости

Более того, если $k = i$:

$$r_k^T d_k = r_k^T u_k + \sum_{j=0}^{k-1} \beta_{jk} r_k^T d_j = r_k^T u_k + 0,$$

Леммы для сходимости

Более того, если $k = i$:

$$r_k^T d_k = r_k^T u_k + \sum_{j=0}^{k-1} \beta_{jk} r_k^T d_j = r_k^T u_k + 0,$$

Леммы для сходимости

Более того, если $k = i$:

$$r_k^T d_k = r_k^T u_k + \sum_{j=0}^{k-1} \beta_{jk} r_k^T d_j = r_k^T u_k + 0,$$

и мы имеем для любого k (из-за произвольного выбора i):

$$\boxed{r_k^T d_k = r_k^T u_k.} \qquad r_k^T d_k = r_k^T r_k \tag{11}$$

Леммы для сходимости

Более того, если $k = i$:

$$r_k^T d_k = r_k^T u_k + \sum_{j=0}^{k-1} \beta_{jk} r_k^T d_j = r_k^T u_k + 0,$$

и мы имеем для любого k (из-за произвольного выбора i):

$$r_k^T d_k = r_k^T u_k. \tag{11}$$

Лемма 6. Пересчет невязки

$$\boxed{r_{k+1} = r_k - \alpha_k A d_k} \tag{12}$$

Леммы для сходимости

Более того, если $k = i$:

$$r_k^T d_k = r_k^T u_k + \sum_{j=0}^{k-1} \beta_{jk} r_k^T d_j = r_k^T u_k + 0,$$

и мы имеем для любого k (из-за произвольного выбора i):

$$r_k^T d_k = r_k^T u_k. \tag{11}$$

Лемма 6. Пересчет невязки

$$\boxed{r_{k+1} = r_k - \alpha_k A d_k} \tag{12}$$

$$\boxed{r_{k+1} = -Ae_{k+1}} = -A\left(e_k + \alpha_k d_k\right) = -Ae_k - \alpha_k A d_k = r_k - \alpha_k A d_k$$

Наконец, все эти вышеуказанные леммы достаточны для доказательства, что $\beta_{ji} = 0$ для всех i, j, кроме соседних.

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\boxed{\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A}}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$u_i = r_i$ CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$
\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} \overset{A^T = A}{=} -\frac{r_i^T A d_j}{d_j^T A d_j}.
$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\boxed{\langle r_i, r_{j+1} \rangle}$ используя (12):

$$\langle r_i, r_{j+1} \rangle$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$
\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.
$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$
\langle r_i, r_{j+1} \rangle = \underline{\langle r_i, r_j - \alpha_j A d_j \rangle}
$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \underbrace{\langle r_i, r_j \rangle} - \underbrace{\alpha_j \langle r_i, A d_j \rangle}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$

$$\alpha_j \langle r_i, A d_j \rangle$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

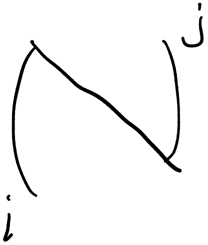
$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$

$$\boxed{\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle}$$

Грам-Шмидт в методе CG



Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$

$$\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle$$

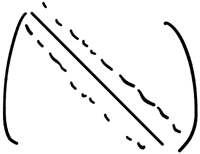
1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \overset{0}{\langle r_i, r_{i+1} \rangle} = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.

$$\alpha_i = \frac{\langle r_i, r_i \rangle}{\langle r_i, A d_i \rangle}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$



Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$

$$\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $\underline{i = j + 1}$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$
\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.
$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$
\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle
$$

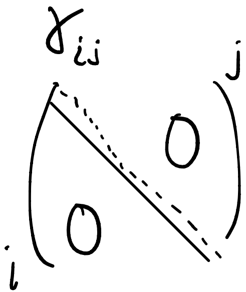
$$
\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle
$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j + 1$: $\boxed{\alpha_j \langle r_i, A d_j \rangle = \overset{0}{\langle r_i, r_{i-1} \rangle} - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle}$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$



Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$

$$\gamma_{ij} \longrightarrow \alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j + 1$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

$$-\frac{r_{ij}}{d_j^T A d_j}$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$

$$\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j + 1$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Наконец, мы имеем формулу для $i = j + 1$:

$$\boxed{\beta_{ji} = -\frac{r_i^T A d_j}{d_j^T A d_j}}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$

$$\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j + 1$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Наконец, мы имеем формулу для $i = j + 1$:

$$\beta_{ji} = -\frac{r_i^T A d_j}{d_j^T A d_j} = \frac{1}{\alpha_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$
$$\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j + 1$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Наконец, мы имеем формулу для $i = j + 1$:

$$\beta_{ji} = -\frac{r_i^T A d_j}{d_j^T A d_j} = \frac{1}{\alpha_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{d_j^T A d_j}{d_j^T r_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$
\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.
$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$
\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle
$$

$$
\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle
$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j + 1$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Наконец, мы имеем формулу для $i = j + 1$:

$$
\beta_{ji} = -\frac{r_i^T A d_j}{d_j^T A d_j} = \frac{1}{\alpha_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{d_j^T A d_j}{\boxed{d_j^T r_j}} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{\langle r_i, r_i \rangle}{\langle r_j, r_j \rangle}
$$

$= r_j^T r_j$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$

$$\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j + 1$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Наконец, мы имеем формулу для $\boxed{i = j + 1}$:

$$\beta_{ji} = -\frac{r_i^T A d_j}{d_j^T A d_j} = \frac{1}{\alpha_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{d_j^T A d_j}{d_j^T r_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{\langle r_i, r_i \rangle}{\langle r_j, r_j \rangle} = \frac{\langle r_i, r_i \rangle}{\langle r_{i-1}, r_{i-1} \rangle}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\langle r_i, r_{j+1} \rangle = \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle$$
$$\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j + 1$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Наконец, мы имеем формулу для $i = j + 1$:

$$\beta_{ji} = -\frac{r_i^T A d_j}{d_j^T A d_j} = \frac{1}{\alpha_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{d_j^T A d_j}{d_j^T r_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{\langle r_i, r_i \rangle}{\langle r_j, r_j \rangle} = \frac{\langle r_i, r_i \rangle}{\langle r_{i-1}, r_{i-1} \rangle}$$

Грам-Шмидт в методе CG

Рассмотрим процесс Грам-Шмидта в методе CG

$$\beta_{ji} = -\frac{\langle d_j, u_i \rangle_A}{\langle d_j, d_j \rangle_A} = -\frac{d_j^T A u_i}{d_j^T A d_j} = -\frac{d_j^T A r_i}{d_j^T A d_j} = -\frac{r_i^T A d_j}{d_j^T A d_j}.$$

Рассмотрим скалярное произведение $\langle r_i, r_{j+1} \rangle$ используя (12):

$$\begin{aligned}\langle r_i, r_{j+1} \rangle &= \langle r_i, r_j - \alpha_j A d_j \rangle = \langle r_i, r_j \rangle - \alpha_j \langle r_i, A d_j \rangle \\ \alpha_j \langle r_i, A d_j \rangle &= \langle r_i, r_j \rangle - \langle r_i, r_{j+1} \rangle\end{aligned}$$

1. Если $i = j$: $\alpha_i \langle r_i, A d_i \rangle = \langle r_i, r_i \rangle - \langle r_i, r_{i+1} \rangle = \langle r_i, r_i \rangle$. Этот случай не интересен по построению процесса Грам-Шмидта.
2. Соседний случай $i = j+1$: $\alpha_j \langle r_i, A d_j \rangle = \langle r_i, r_{i-1} \rangle - \langle r_i, r_i \rangle = -\langle r_i, r_i \rangle$
3. Для любого другого случая: $\alpha_j \langle r_i, A d_j \rangle = 0$, потому что все невязки ортогональны друг другу.

Наконец, мы имеем формулу для $i = j+1$:

$$\beta_{ji} = -\frac{r_i^T A d_j}{d_j^T A d_j} = \frac{1}{\alpha_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{d_j^T A d_j}{d_j^T r_j} \frac{\langle r_i, r_i \rangle}{d_j^T A d_j} = \frac{\langle r_i, r_i \rangle}{\langle r_j, r_j \rangle} = \frac{\langle r_i, r_i \rangle}{\langle r_{i-1}, r_{i-1} \rangle}$$

И для направления $d_{k+1} = r_{k+1} + \beta_{k,k+1} d_k, \qquad \beta_{k,k+1} = \beta_k = \dfrac{\langle r_{k+1}, r_{k+1} \rangle}{\langle r_k, r_k \rangle}.$

Метод сопряженных градиентов (CG)

$\mathbf{r}_0 := \mathbf{b} - \mathbf{A}\mathbf{x}_0$ $\quad r_0 = -\nabla f(x_0)$

if $\mathbf{r}_0$ is sufficiently small, then return $\mathbf{x}_0$ as the result

$\mathbf{d}_0 := \mathbf{r}_0$

$k := 0$

repeat

$$\alpha_k := \frac{\mathbf{r}_k^\top \mathbf{r}_k}{\mathbf{d}_k^\top \mathbf{A}\mathbf{d}_k} \qquad \alpha_k = \text{linesearch}(f(x), d_k)$$

$$\mathbf{x}_{k+1} := \mathbf{x}_k + \alpha_k \mathbf{d}_k$$

$$\mathbf{r}_{k+1} := \mathbf{r}_k - \alpha_k \mathbf{A}\mathbf{d}_k \qquad r_{k+1} = -\nabla f(x_{k+1})$$

if $\mathbf{r}_{k+1}$ is sufficiently small, then exit loop

$$\beta_k := \frac{\mathbf{r}_{k+1}^\top \mathbf{r}_{k+1}}{\mathbf{r}_k^\top \mathbf{r}_k} = \frac{\|\nabla f(x_{k+1})\|_2^2}{\|\nabla f(x_k)\|_2^2}$$

$$\mathbf{d}_{k+1} := \mathbf{r}_{k+1} + \beta_k \mathbf{d}_k$$

$$k := k + 1$$

end repeat

return $\mathbf{x}_{k+1}$ as the result

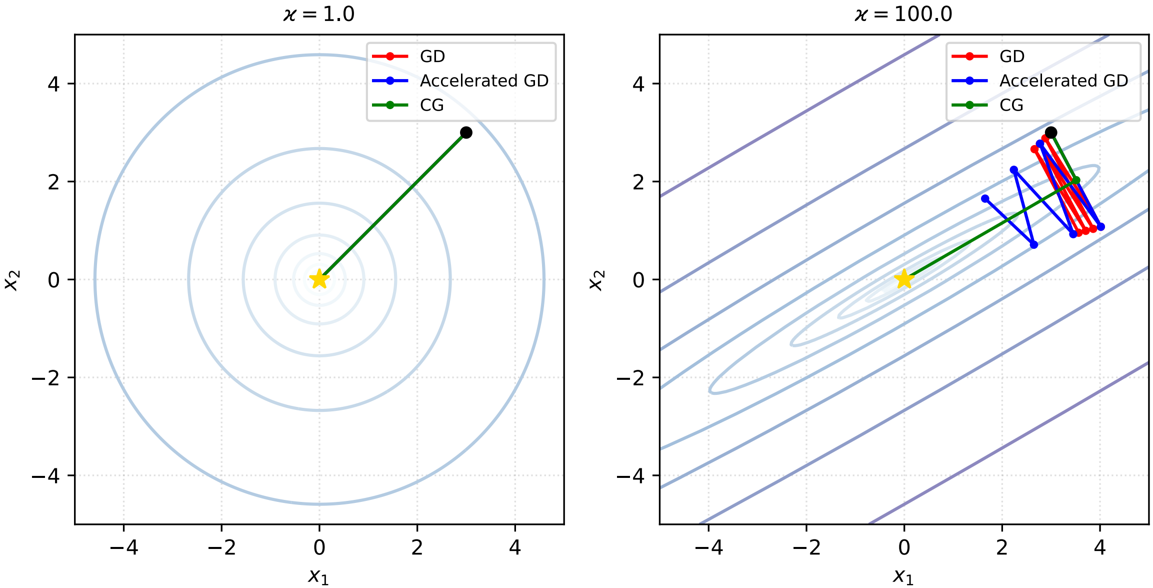
$\min_{x \in \mathbb{R}^n} f(x)$ $\quad A \in \mathbb{S}^n_{++}$

$f(x) = \frac{1}{2} x^\top A x - b^\top x$

$f(x)$ – любая

Метод Флетчера-Ривса

Закрываем квадратичный вопрос



Сходимость

Теорема 1. Если матрица A имеет только r различных собственных значений, то метод сопряженных градиентов сходится за r итераций.

Теорема 2. Следующая оценка сходимости выполняется для метода сопряженных градиентов, как для итерационного метода в сильно выпуклой задаче:

$$\|x_k - x^*\|_A \leq 2 \left(\frac{\sqrt{\kappa(A)} - 1}{\sqrt{\kappa(A)} + 1} \right)^k \|x_0 - x^*\|_A,$$

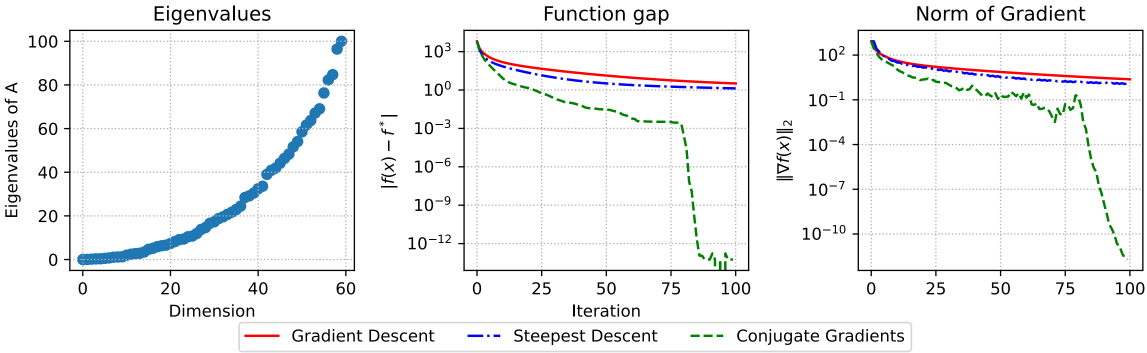
где $\|x\|^2_A = x^\top A x$ и $\varkappa(A) = \frac{\lambda_1(A)}{\lambda_n(A)}$ - это число обусловленности матрицы A, $\lambda_1(A) \geq ... \geq \lambda_n(A)$ - собственные значения матрицы A

Примечание: Сравните коэффициент геометрической прогрессии с его аналогом в методе градиентного спуска.

Численные эксперименты

$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$

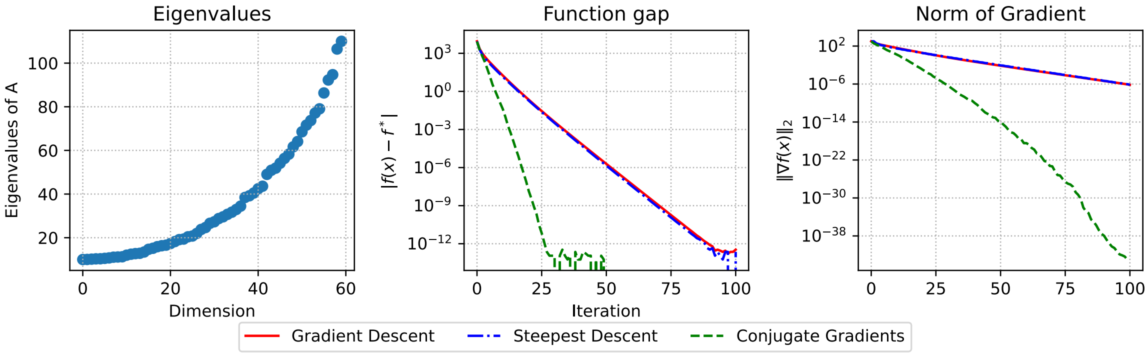
Convex quadratics. n=60, random matrix.



Численные эксперименты

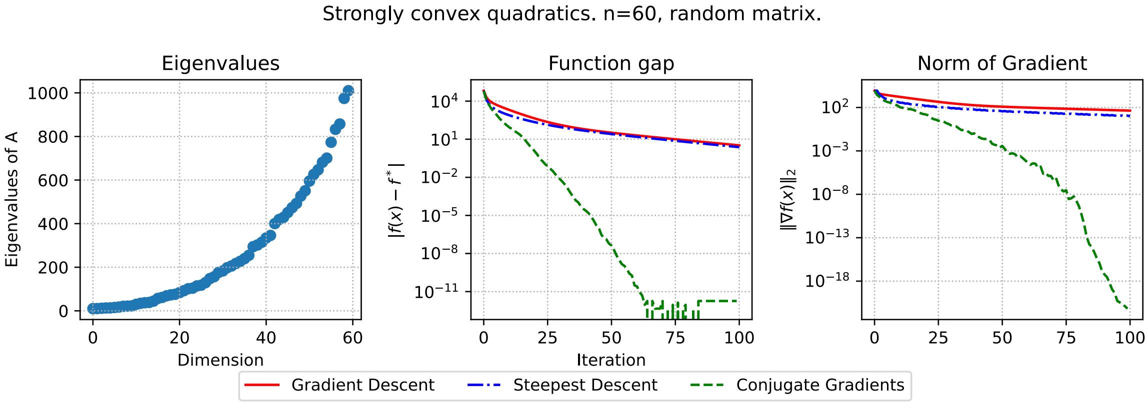
$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$

Strongly convex quadratics. n=60, random matrix.



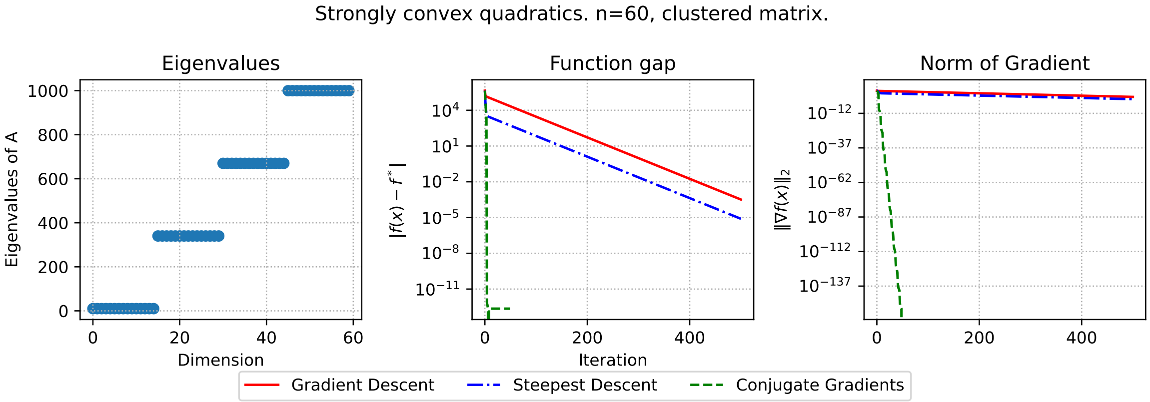
Численные эксперименты

$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$



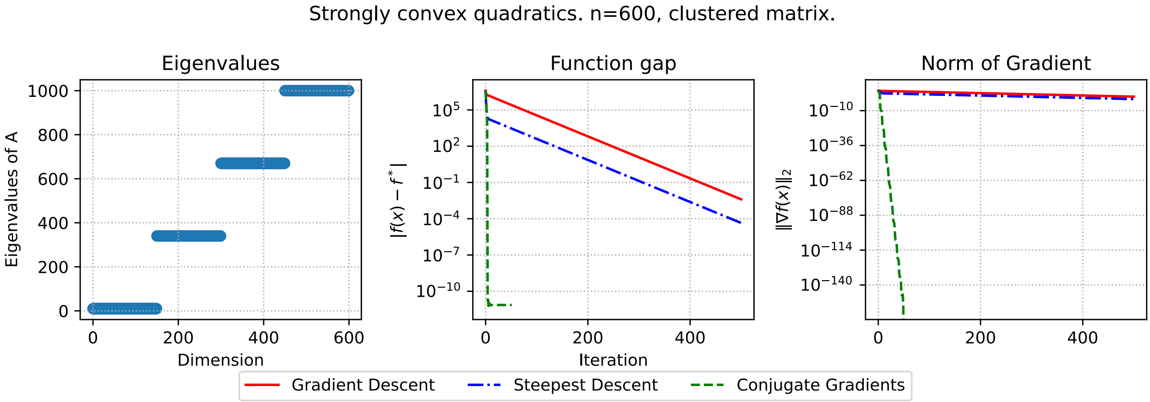
Численные эксперименты

$$
f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}
$$



Численные эксперименты

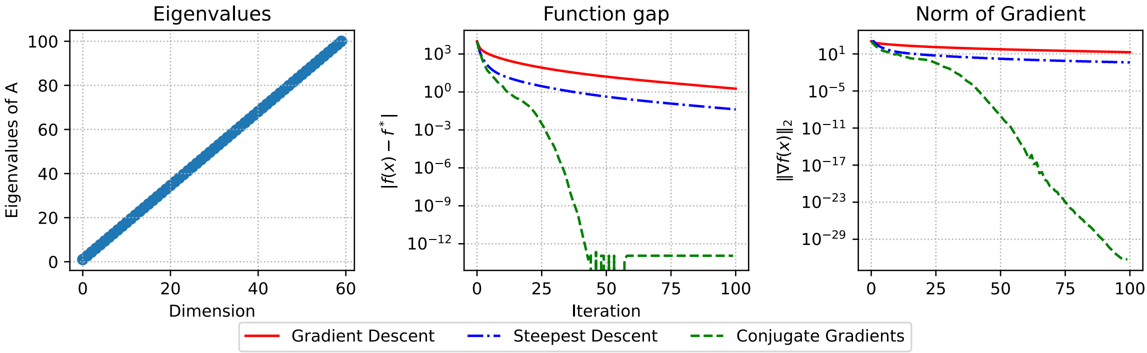
$$
f(x) = \frac{1}{2}x^T Ax - b^T x \to \min_{x \in \mathbb{R}^n}
$$



Численные эксперименты

$$f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}$$

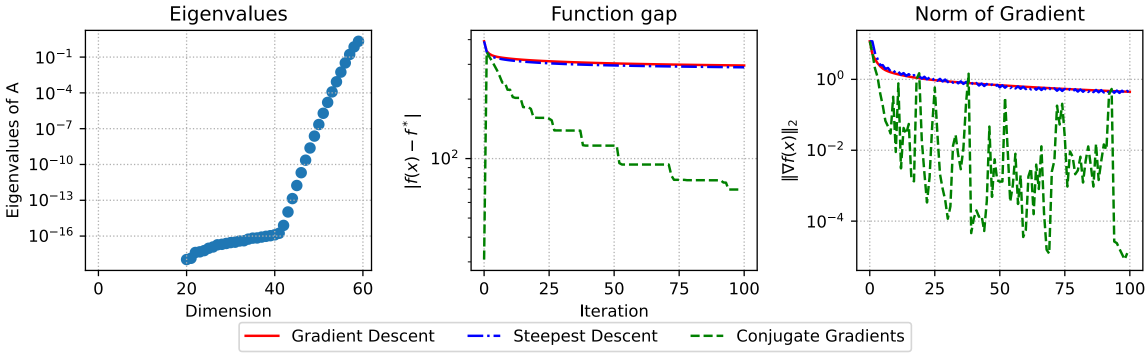
Strongly convex quadratics. n=60, uniform spectrum matrix.



Численные эксперименты

$$f(x) = \frac{1}{2}x^T A x - b^T x \to \min_{x \in \mathbb{R}^n}$$

Strongly convex quadratics. n=60, Hilbert matrix.



Патологический пример

Пусть $t \in (0,1)$ и

$$W = \begin{bmatrix} t & \sqrt{t} & & & \\ \sqrt{t} & 1+t & \sqrt{t} & & \\ & \sqrt{t} & 1+t & \sqrt{t} & \\ & & \ddots & \ddots & \ddots \\ & & & \sqrt{t} & 1+t \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

- Так как W невырожденна, существует единственное решение $Wx = b$. Решение методом сопряжённых градиентов даёт довольно плохую сходимость.
- Во время работы CG ошибка растёт экспоненциально (!), пока внезапно не становится нулевой, когда находится единственное решение.
- Невязка $\|Wx_k - b\|^2$ растёт экспоненциально как $(1/t)^k$ до n-й итерации, после чего резко падает к нулю.
- См. эксперимент здесь.

Метод сопряженных градиентов для неквадратичных задач (Non-linear CG)

Метод Полака-Рибьера

Polack Polyak

В случае, когда нет аналитического выражения для функции или ее градиента, мы, скорее всего, не сможем решить одномерную задачу минимизации аналитически. Поэтому шаг 2 алгоритма заменяется обычной процедурой линейного поиска. Но есть следующий математический трюк для четвертого шага:

Для двух итераций справедливо:

$$\boxed{x_{k+1} - x_k = cd_k,}$$

Метод Полака-Рибьера

В случае, когда нет аналитического выражения для функции или ее градиента, мы, скорее всего, не сможем решить одномерную задачу минимизации аналитически. Поэтому шаг 2 алгоритма заменяется обычной процедурой линейного поиска. Но есть следующий математический трюк для четвертого шага:

Для двух итераций справедливо:

$$
x_{k+1} - x_k = cd_k,
$$

где c - некоторая константа. Тогда для квадратичного случая мы имеем:

$$
\nabla f(x_{k+1}) - \nabla f(x_k) = (Ax_{k+1} - b) - (Ax_k - b) = A(x_{k+1} - x_k) = cAd_k
$$

$$
Ad_k = \frac{1}{c}\left(\nabla f(x_{k+1}) - \nabla f(x_k)\right)
$$

Метод Полака-Рибьера

В случае, когда нет аналитического выражения для функции или ее градиента, мы, скорее всего, не сможем решить одномерную задачу минимизации аналитически. Поэтому шаг 2 алгоритма заменяется обычной процедурой линейного поиска. Но есть следующий математический трюк для четвертого шага:

Для двух итераций справедливо:

$$
x_{k+1} - x_k = c d_k,
$$

где c - некоторая константа. Тогда для квадратичного случая мы имеем:

$$
\nabla f(x_{k+1}) - \nabla f(x_k) = (A x_{k+1} - b) - (A x_k - b) = A(x_{k+1} - x_k) = c A d_k
$$

Выражая из этого уравнения величину $A d_k = \dfrac{1}{c}\left(\nabla f(x_{k+1}) - \nabla f(x_k)\right)$, мы избавляемся от знания функции в определении β_k, тогда пункт 4 будет переписан как:

FR

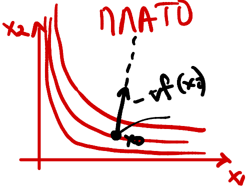
$$
\frac{\|\nabla f(x_{k+1})\|_2^2}{\|\nabla f(x_k)\|_2^2}
$$

PR

$$
\beta_k = \frac{\nabla f(x_{k+1})^\top \left(\nabla f(x_{k+1}) - \nabla f(x_k)\right)}{d_k^\top \left(\nabla f(x_{k+1}) - \nabla f(x_k)\right)}.
$$

Этот метод называется методом Полака-Рибьера.

Метод Флетчера-Ривза



Вернемся к формуле для β_k из линейного случая (для квадратичной задачи):

$$\beta_k = \frac{\mathbf{r}_{k+1}^\top \mathbf{r}_{k+1}}{\mathbf{r}_k^\top \mathbf{r}_k}$$

Так как невязка $\mathbf{r}_k = b - Ax_k = -\nabla f(x_k)$, то прямая подстановка дает формулу Флетчера-Ривза:

$$\beta_k^{FR} = \frac{\|\nabla f(x_{k+1})\|^2}{\|\nabla f(x_k)\|^2}$$

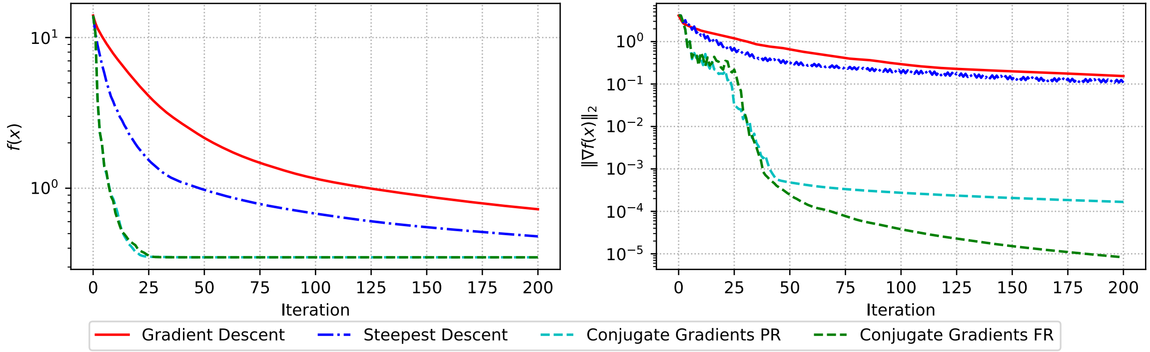
Очевидно, что в случае квадратичной функции этот метод совпадает с исходным методом сопряженных градиентов. Однако, для неквадратичных функций часто требуются рестарты.

Численные эксперименты

Бин. лог. регрессия

$$
f(x) = \underbrace{\frac{\mu}{2}\|x\|_2^2}_{\ell_2} + \frac{1}{m}\sum_{i=1}^{m}\log(1+\exp(-y_i\langle a_i, x\rangle)) \to \min_{x\in\mathbb{R}^n}
$$

Regularized binary logistic regression. n=300. m=1000. μ=0

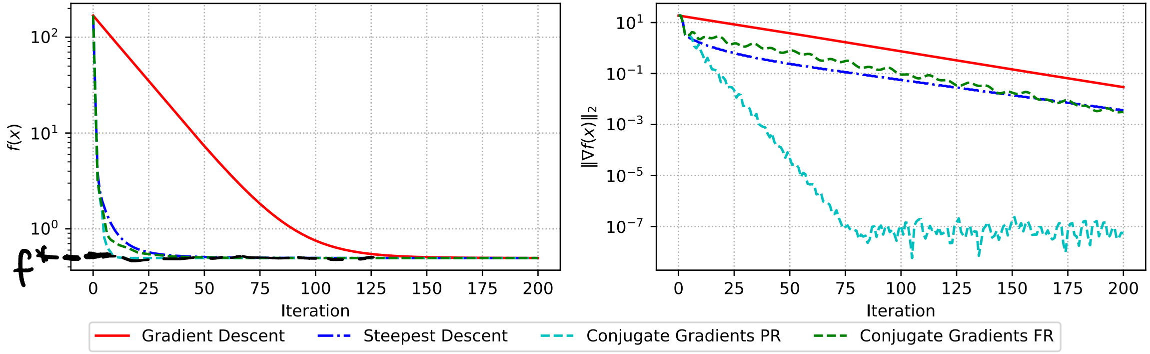


Численные эксперименты

$(f(x) - f^*)$

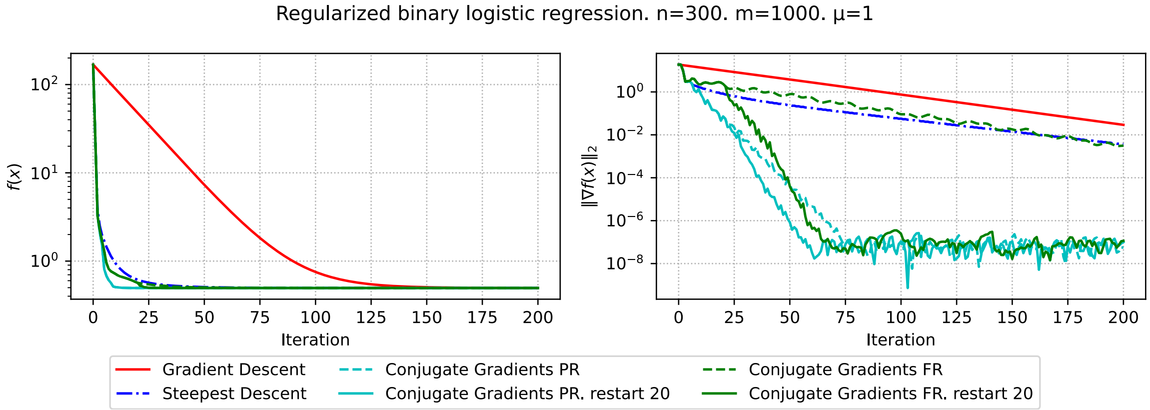
$$f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\log(1+\exp(-y_i\langle a_i, x\rangle)) \to \min_{x\in\mathbb{R}^n}$$

Regularized binary logistic regression. n=300. m=1000. μ=1



Численные эксперименты

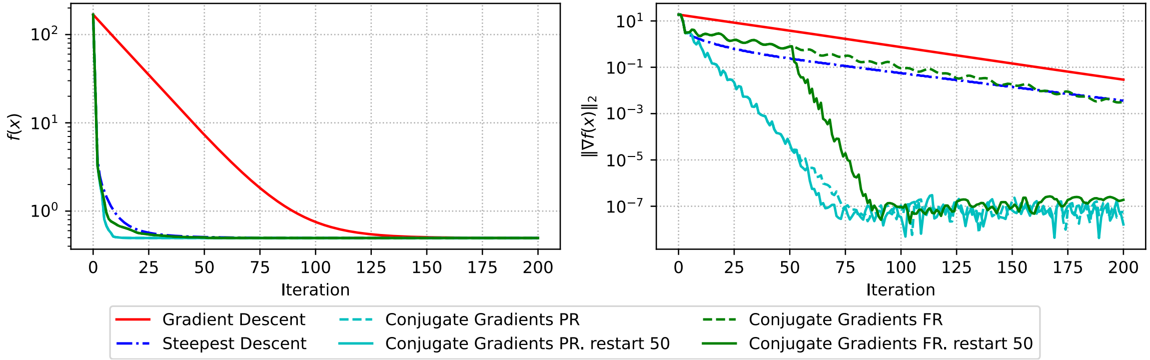
$$
f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\log(1+\exp(-y_i\langle a_i, x\rangle)) \to \min_{x\in\mathbb{R}^n}
$$



Численные эксперименты

$$
f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\log(1+\exp(-y_i\langle a_i, x\rangle)) \to \min_{x\in\mathbb{R}^n}
$$

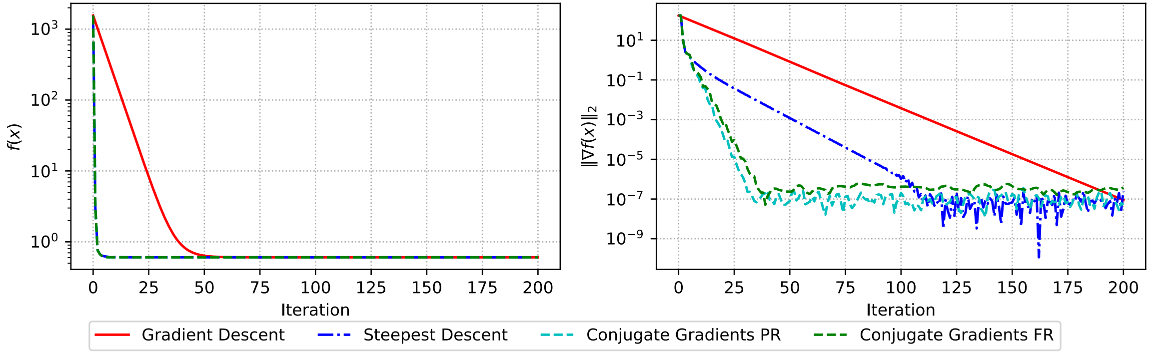
Regularized binary logistic regression. n=300. m=1000. μ=1



Численные эксперименты

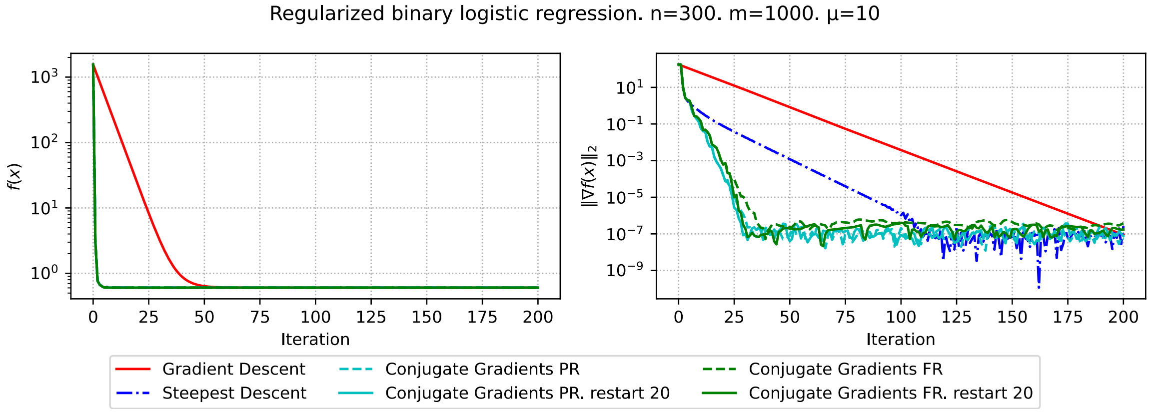
$$f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m}\log(1+\exp(-y_i\langle a_i, x\rangle)) \to \min_{x\in\mathbb{R}^n}$$

Regularized binary logistic regression. n=300. m=1000. μ=10



Численные эксперименты

$$f(x) = \frac{\mu}{2}\|x\|_2^2 + \frac{1}{m}\sum_{i=1}^{m} \log(1 + \exp(-y_i\langle a_i, x\rangle)) \to \min_{x\in\mathbb{R}^n}$$



Другие численные эксперименты

Посмотрим другие примеры здесь. Код взят из.