

Метод Ньютона. Квазиньютоновские методы. Матрица предобработки Методы оптимизации

Александр Безносиков

Московский физико-технический институт

10 октября 2024



Задача поиска нуля

- Рассмотрим задачу поиска «корня» функции:

Найти t^* , что $\varphi(t^*) = 0$,

где $\varphi : \mathbb{R} \rightarrow \mathbb{R}$.

Задача поиска нуля

- Рассмотрим задачу поиска «корня» функции:

Найти t^* , что $\varphi(t^*) = 0$,

где $\varphi : \mathbb{R} \rightarrow \mathbb{R}$.

- Пусть мы находимся в точке t^0 и хотим найти такую поправку Δt , что $t^0 + \Delta t \approx t^*$.

Задача поиска нуля

- Рассмотрим задачу поиска «корня» функции:

Найти t^* , что $\varphi(t^*) = 0$,

где $\varphi : \mathbb{R} \rightarrow \mathbb{R}$.

- Пусть мы находимся в точке t^0 и хотим найти такую поправку Δt , что $t^0 + \Delta t \approx t^*$.
- Разложим в ряд:

$$\varphi(t^0 + \Delta t) = \varphi(t^0) + \varphi'(t^0)\Delta t + o(\Delta t).$$

Задача поиска нуля

- Рассмотрим задачу поиска «корня» функции:

Найти t^* , что $\varphi(t^*) = 0$,

где $\varphi : \mathbb{R} \rightarrow \mathbb{R}$.

- Пусть мы находимся в точке t^0 и хотим найти такую поправку Δt , что $t^0 + \Delta t \approx t^*$.
- Разложим в ряд:

$$\varphi(t^0 + \Delta t) = \varphi(t^0) + \varphi'(t^0)\Delta t + o(\Delta t).$$

- Так как мы хотим $t^0 + \Delta t \approx t^*$, то

$$\varphi(t^0 + \Delta t) \approx \varphi(t^*) = 0 \Rightarrow \varphi(t^0) + \varphi'(t^0)\Delta t \approx 0.$$

Задача поиска нуля: метод Ньютона

- Из $\varphi(t^0) + \varphi'(t^0)\Delta t \approx 0$ получаем:

$$\Delta t \approx -\frac{\varphi(t^0)}{\varphi'(t^0)}.$$

Задача поиска нуля: метод Ньютона

- Из $\varphi(t^0) + \varphi'(t^0)\Delta t \approx 0$ получаем:

$$\Delta t \approx -\frac{\varphi(t^0)}{\varphi'(t^0)}.$$

- Значит получаем новую точку $t^1 = t^0 + \Delta t$. Откуда получается итеративный метод:

$$t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)}$$

- Этот метод называется методом Ньютона. Его предложил во второй половине 17го века тот самый Ньютон.

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона?

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .
- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение?

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .
- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение? $t^* = 0$.

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .
- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение? $t^* = 0$.

- Производная: $\varphi'(t) = \frac{1}{(1+t^2)^{3/2}}.$

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .
- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение? $t^* = 0$.

- Производная: $\varphi'(t) = \frac{1}{(1+t^2)^{3/2}}$. Откуда итерация метода Ньютона

$$t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)} = -(t^k)^3.$$

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .

- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение? $t^* = 0$.

- Производная: $\varphi'(t) = \frac{1}{(1+t^2)^{3/2}}$. Откуда итерация метода Ньютона

$$t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)} = -(t^k)^3.$$

- **Вопрос:** что можем сказать о сходимости к решению?

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .
- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение? $t^* = 0$.

- Производная: $\varphi'(t) = \frac{1}{(1+t^2)^{3/2}}$. Откуда итерация метода Ньютона

$$t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)} = -(t^k)^3.$$

- **Вопрос:** что можем сказать о сходимости к решению?
 - $|t^0| < 1$ — есть сходимость

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .
- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение? $t^* = 0$.

- Производная: $\varphi'(t) = \frac{1}{(1+t^2)^{3/2}}$. Откуда итерация метода Ньютона

$$t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)} = -(t^k)^3.$$

- **Вопрос:** что можем сказать о сходимости к решению?
 - $|t^0| < 1$ — есть сходимость
 - $|t^0| = 1$ — колеблемся в точках -1 и 1

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .
- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение? $t^* = 0$.

- Производная: $\varphi'(t) = \frac{1}{(1+t^2)^{3/2}}$. Откуда итерация метода Ньютона

$$t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)} = -(t^k)^3.$$

- **Вопрос:** что можем сказать о сходимости к решению?
 - $|t^0| < 1$ — есть сходимость
 - $|t^0| = 1$ — колеблемся в точках -1 и 1
 - $|t^0| > 1$ — расходимся

Метод Ньютона: локальная сходимость

- **Вопрос:** какие есть вопросы к интуиции получения итерации метода Ньютона? Важно, что t^0 из «хорошей окрестности» t^* .

- Рассмотрим

$$\varphi(t) = \frac{t}{\sqrt{1+t^2}}.$$

Вопрос: какое решение? $t^* = 0$.

- Производная: $\varphi'(t) = \frac{1}{(1+t^2)^{3/2}}$. Откуда итерация метода Ньютона

$$t^{k+1} = t^k - \frac{\varphi(t^k)}{\varphi'(t^k)} = -(t^k)^3.$$

- **Вопрос:** что можем сказать о сходимости к решению?
 - $|t^0| < 1$ — есть сходимость
 - $|t^0| = 1$ — колеблемся в точках -1 и 1
 - $|t^0| > 1$ — расходимся
- Ключевая особенность метода Ньютона — локальная сходимость (только в окрестности решения).

Метод Ньютона: оптимизация

- Рассмотрим задачу безусловной оптимизации с выпуклой дважды непрерывно дифференцируемой целевой функцией:

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

- Вопрос:** для такой задачи мы тоже ищем 0, но чего?

Метод Ньютона: оптимизация

- Рассмотрим задачу безусловной оптимизации с выпуклой дважды непрерывно дифференцируемой целевой функцией:

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

- Вопрос:** для такой задачи мы тоже ищем 0, но чего? $\nabla f(x^*) = 0$.

Метод Ньютона: оптимизация

- Рассмотрим задачу безусловной оптимизации с выпуклой дважды непрерывно дифференцируемой целевой функцией:

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

- Вопрос:** для такой задачи мы тоже ищем 0, но чего? $\nabla f(x^*) = 0$. Откуда метод Ньютона для задачи оптимизации

Алгоритм 3 Метод Ньютона

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

- for** $k = 0, 1, \dots, K - 1$ **do**
- Вычислить $\nabla f(x^k)$, $\nabla^2 f(x^k)$
- $x^{k+1} = x^k - (\nabla^2 f(x^k))^{-1} \nabla f(x^k)$
- end for**

Выход: x^K

Метод Ньютона и градиентный спуск

- Градиентный спуск работает с линейной аппроксимацией в текущей точке, метод Ньютона — с квадратичной:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle.$$

Метод Ньютона и градиентный спуск

- Градиентный спуск работает с линейной аппроксимацией в текущей точке, метод Ньютона — с квадратичной:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle.$$

Минимизируем квадратичную аппроксимацию по x :

Метод Ньютона и градиентный спуск

- Градиентный спуск работает с линейной аппроксимацией в текущей точке, метод Ньютона — с квадратичной:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle.$$

Минимизируем квадратичную аппроксимацию по x :

$$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0.$$

Метод Ньютона и градиентный спуск

- Градиентный спуск работает с линейной аппроксимацией в текущей точке, метод Ньютона — с квадратичной:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle.$$

Минимизируем квадратичную аппроксимацию по x :

$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$. Откуда получаем следующую

точку метода:

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

Метод Ньютона и градиентный спуск

- Градиентный спуск работает с линейной аппроксимацией в текущей точке, метод Ньютона — с квадратичной:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle.$$

Минимизируем квадратичную аппроксимацию по x :

$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$. Откуда получаем следующую точку метода:

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

- Метод Ньютона использует оракул второго порядка: требует вычисление гессиана.

Метод Ньютона и градиентный спуск

- Градиентный спуск работает с линейной аппроксимацией в текущей точке, метод Ньютона — с квадратичной:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle.$$

Минимизируем квадратичную аппроксимацию по x :

$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$. Откуда получаем следующую точку метода:

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

- Метод Ньютона использует оракул второго порядка: требует вычисление гессиана.
- Стоимость итерации значительно возрастает (по сравнению с градиентным спуском) не только из-за гессиана, но и его обращения.

Метод Ньютона и градиентный спуск

- Градиентный спуск работает с линейной аппроксимацией в текущей точке, метод Ньютона — с квадратичной:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle.$$

Минимизируем квадратичную аппроксимацию по x :

$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$. Откуда получаем следующую точку метода:

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

- Метод Ньютона использует оракул второго порядка: требует вычисление гессиана.
- Стоимость итерации значительно возрастает (по сравнению с градиентным спуском) не только из-за гессиана, но и его обращения. **Вопрос:** за сколько итераций метод Ньютона сойдется для квадратичной задачи с положительно определенной матрицей?

Метод Ньютона и градиентный спуск

- Градиентный спуск работает с линейной аппроксимацией в текущей точке, метод Ньютона — с квадратичной:

$$f(x) \approx f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle.$$

Минимизируем квадратичную аппроксимацию по x :

$\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$. Откуда получаем следующую точку метода:

$$x^{k+1} = x^k - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

- Метод Ньютона использует оракул второго порядка: требует вычисление гессиана.
- Стоимость итерации значительно возрастает (по сравнению с градиентным спуском) не только из-за гессиана, но и его обращения. **Вопрос:** за сколько итераций метод Ньютона сойдется для квадратичной задачи с положительно определенной матрицей? за 1 (но дорогую).

Метод Ньютона: сходимость

- То, что для квадратичной задачи метод Ньютона сходится за 1 итерацию, наталкивает на мысль о том, что при всех своих минусах (локальная сходимость, дороговизна итерации) ключевым плюсом является скорость сходимости.

Метод Ньютона: сходимость

- То, что для квадратичной задачи метод Ньютона сходится за 1 итерацию, наталкивает на мысль о том, что при всех своих минусах (локальная сходимость, дороговизна итерации) ключевым плюсом является скорость сходимости.
- Пусть целевая функция в задаче безусловной минимизации является дважды непрерывно дифференцируемой, μ -сильно выпуклой и имеет M -Липшицев гессиан, т.е. для любых $x, y \in \mathbb{R}^d$ справедливо:

$$\nabla^2 f(x) \succeq \mu I, \quad \|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq M\|x - y\|_2.$$

В случае матрицы $\|\cdot\|_2$ – спектральная норма (согласованная норма с евклидовой для векторов).

Метод Ньютона: сходимость

- Доказываем сходимость.

Метод Ньютона: сходимость

- Доказываем сходимость. Будем изучать, как меняется расстояние до решения:

$$x^{k+1} - x^* = x^k - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k) - x^*.$$

Метод Ньютона: сходимость

- Доказываем сходимость. Будем изучать, как меняется расстояние до решения:

$$x^{k+1} - x^* = x^k - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k) - x^*.$$

- Снова вспомним формулу Ньютона-Лейбница для интеграла вдоль кривой:

$$\nabla f(x^k) - \nabla f(x^*) = \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)) (x^k - x^*) d\tau$$

Метод Ньютона: сходимость

- Доказываем сходимость. Будем изучать, как меняется расстояние до решения:

$$x^{k+1} - x^* = x^k - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k) - x^*.$$

- Снова вспомним формулу Ньютона-Лейбница для интеграла вдоль кривой:

$$\nabla f(x^k) - \nabla f(x^*) = \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)) (x^k - x^*) d\tau$$

Зная, что $\nabla f(x^*) = 0$, получим

$$x^{k+1} - x^* = x^k - x^* - \left(\nabla^2 f(x^k) \right)^{-1} \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)) (x^k - x^*) d\tau.$$

Метод Ньютона: сходимость

- Продолжаем и используем «умную единицу»:

$$\begin{aligned}x^{k+1} - x^* &= x^k - x^* - \left(\nabla^2 f(x^k)\right)^{-1} \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*))(x^k - x^*) d\tau \\&= \left(\nabla^2 f(x^k)\right)^{-1} \nabla^2 f(x^k)(x^k - x^*) \\&\quad - \left(\nabla^2 f(x^k)\right)^{-1} \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*))(x^k - x^*) d\tau.\end{aligned}$$

Метод Ньютона: сходимость

- Продолжаем и используем «умную единицу»:

$$\begin{aligned}x^{k+1} - x^* &= x^k - x^* - \left(\nabla^2 f(x^k)\right)^{-1} \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*))(x^k - x^*) d\tau \\&= \left(\nabla^2 f(x^k)\right)^{-1} \nabla^2 f(x^k)(x^k - x^*) \\&\quad - \left(\nabla^2 f(x^k)\right)^{-1} \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*))(x^k - x^*) d\tau.\end{aligned}$$

- Заметим, что $x^k - x^*$ можно вынести за пределы интеграла:

$$\begin{aligned}x^{k+1} - x^* &= \left(\nabla^2 f(x^k)\right)^{-1} \nabla^2 f(x^k)(x^k - x^*) \\&\quad - \left(\nabla^2 f(x^k)\right)^{-1} \left(\int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)) d\tau\right) (x^k - x^*).\end{aligned}$$

Метод Ньютона: сходимость

- Введем обозначение $G_k = \nabla^2 f(x^k) - \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)) d\tau$:

$$x^{k+1} - x^* = \left(\nabla^2 f(x^k) \right)^{-1} G_k (x^k - x^*).$$

Метод Ньютона: сходимость

- Введем обозначение $G_k = \nabla^2 f(x^k) - \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)) d\tau$:

$$x^{k+1} - x^* = \left(\nabla^2 f(x^k) \right)^{-1} G_k (x^k - x^*).$$

- Перейдем к оценке нормы расстояния:

$$\|x^{k+1} - x^*\|_2 = \left\| \left(\nabla^2 f(x^k) \right)^{-1} G_k (x^k - x^*) \right\|_2$$

Метод Ньютона: сходимость

- Введем обозначение $G_k = \nabla^2 f(x^k) - \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)) d\tau$:

$$x^{k+1} - x^* = \left(\nabla^2 f(x^k) \right)^{-1} G_k (x^k - x^*).$$

- Перейдем к оценке нормы расстояния:

$$\|x^{k+1} - x^*\|_2 = \left\| \left(\nabla^2 f(x^k) \right)^{-1} G_k (x^k - x^*) \right\|_2$$

- Пользуемся, что спектральная норма матрицы согласована с евклидовой вектора:

$$\begin{aligned} \|x^{k+1} - x^*\|_2 &\leq \left\| \left(\nabla^2 f(x^k) \right)^{-1} G_k \right\|_2 \|x^k - x^*\|_2 \\ &\leq \left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2 \|G_k\|_2 \|x^k - x^*\|_2. \end{aligned}$$

Метод Ньютона: сходимость

- С предыдущего слайда:

$$\|x^{k+1} - x^*\|_2 \leq \left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2 \|G_k\|_2 \|x^k - x^*\|_2.$$

Метод Ньютона: сходимость

- С предыдущего слайда:

$$\|x^{k+1} - x^*\|_2 \leq \left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2 \|G_k\|_2 \|x^k - x^*\|_2.$$

- Вопрос:** как оценить $\left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2$?

Метод Ньютона: сходимость

- С предыдущего слайда:

$$\|x^{k+1} - x^*\|_2 \leq \left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2 \|G_k\|_2 \|x^k - x^*\|_2.$$

- Вопрос:** как оценить $\left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2$? Мы знаем, что $\nabla^2 f(x) \succeq \mu I$, а значит $\frac{1}{\mu} I \succeq \left(\nabla^2 f(x^k) \right)^{-1}$,

Метод Ньютона: сходимость

- С предыдущего слайда:

$$\|x^{k+1} - x^*\|_2 \leq \left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2 \|G_k\|_2 \|x^k - x^*\|_2.$$

- Вопрос:** как оценить $\left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2$? Мы знаем, что $\nabla^2 f(x) \succeq \mu I$, а значит $\frac{1}{\mu} I \succeq \left(\nabla^2 f(x^k) \right)^{-1}$, откуда $\left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2 \leq \frac{1}{\mu}$ и

$$\|x^{k+1} - x^*\|_2 \leq \frac{1}{\mu} \|G_k\|_2 \|x^k - x^*\|_2.$$

Метод Ньютона: сходимость

- С предыдущего слайда:

$$\|x^{k+1} - x^*\|_2 \leq \left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2 \|G_k\|_2 \|x^k - x^*\|_2.$$

- Вопрос:** как оценить $\left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2$? Мы знаем, что $\nabla^2 f(x) \succeq \mu I$, а значит $\frac{1}{\mu} I \succeq \left(\nabla^2 f(x^k) \right)^{-1}$, откуда $\left\| \left(\nabla^2 f(x^k) \right)^{-1} \right\|_2 \leq \frac{1}{\mu}$ и

$$\|x^{k+1} - x^*\|_2 \leq \frac{1}{\mu} \|G_k\|_2 \|x^k - x^*\|_2.$$

- Осталось оценить $\|G_k\|_2$.

Метод Ньютона: сходимость

- Оцениваем $\|G_k\|_2$:

$$\begin{aligned}\|G_k\|_2 &= \left\| \nabla^2 f(x^k) - \int_0^1 \nabla^2 f(x^* + \tau(x^k - x^*)) d\tau \right\|_2 \\ &= \left\| \int_0^1 \left(\nabla^2 f(x^k) - \nabla^2 f(x^* + \tau(x^k - x^*)) \right) d\tau \right\|_2 \\ &\leq \int_0^1 \left\| \nabla^2 f(x^k) - \nabla^2 f(x^* + \tau(x^k - x^*)) \right\|_2 d\tau \\ &\leq \int_0^1 M(1 - \tau) \|x^k - x^*\|_2 d\tau \\ &= M \|x^k - x^*\|_2 \int_0^1 (1 - \tau) d\tau = \frac{M}{2} \|x^k - x^*\|_2.\end{aligned}$$

Метод Ньютона: сходимость

- Подставляем оценку на $\|G_k\|_2$:

$$\|x^{k+1} - x^*\|_2 \leq \frac{M}{2\mu} \|x^k - x^*\|_2^2.$$

Метод Ньютона: сходимость

- Подставляем оценку на $\|G_k\|_2$:

$$\|x^{k+1} - x^*\|_2 \leq \frac{M}{2\mu} \|x^k - x^*\|_2^2.$$

Теорема об оценке сходимости метода Ньютона для μ -сильно выпуклых функций с M -Липшецевым гессианом

Пусть задача безусловной оптимизации с μ -сильно выпуклой целевой функцией f с M -Липшецевыми гессианом решается методом Ньютона. Тогда справедлива следующая оценка сходимости за 1 итерацию

$$\|x^{k+1} - x^*\|_2 \leq \frac{M}{2\mu} \|x^k - x^*\|_2^2.$$

Метод Ньютона: сходимость

- Подставляем оценку на $\|G_k\|_2$:

$$\|x^{k+1} - x^*\|_2 \leq \frac{M}{2\mu} \|x^k - x^*\|_2^2.$$

Теорема об оценке сходимости метода Ньютона для μ -сильно выпуклых функций с M -Липшецевым гессианом

Пусть задача безусловной оптимизации с μ -сильно выпуклой целевой функцией f с M -Липшецевыми гессианом решается методом Ньютона. Тогда справедлива следующая оценка сходимости за 1 итерацию

$$\|x^{k+1} - x^*\|_2 \leq \frac{M}{2\mu} \|x^k - x^*\|_2^2.$$

Мы уже знаем, что такого рода оценки дают квадратичную скорость сходимости.

Метод Ньютона: сходимость

- Сходимость, как и в случае первоначального метода Ньютона, является локальной.

Метод Ньютона: сходимость

- Сходимость, как и в случае первоначального метода Ньютона, является локальной. А именно, чтобы гарантировать $\|x^1 - x^*\|_2 < \|x^0 - x^*\|_2$, нужно предположить, что

$$\|x^0 - x^*\|_2 < \frac{2\mu}{M}.$$

Метод Ньютона: сходимость

- Сходимость, как и в случае первоначального метода Ньютона, является локальной. А именно, чтобы гарантировать $\|x^1 - x^*\|_2 < \|x^0 - x^*\|_2$, нужно предположить, что

$$\|x^0 - x^*\|_2 < \frac{2\mu}{M}.$$

- Поймем насколько быстро сходится метод. Пусть $M = 2$, $\mu = 1$, а $\|x^0 - x^*\|_2 = \frac{1}{2}$.

Метод Ньютона: сходимость

- Сходимость, как и в случае первородного метода Ньютона, является локальной. А именно, чтобы гарантировать $\|x^1 - x^*\|_2 < \|x^0 - x^*\|_2$, нужно предположить, что

$$\|x^0 - x^*\|_2 < \frac{2\mu}{M}.$$

- Поймем насколько быстро сходится метод. Пусть $M = 2$, $\mu = 1$, а $\|x^0 - x^*\|_2 = \frac{1}{2}$. Тогда мы можем гарантировать, что $\|x^1 - x^*\|_2 \leq \frac{1}{2^2}$,

Метод Ньютона: сходимость

- Сходимость, как и в случае первородного метода Ньютона, является локальной. А именно, чтобы гарантировать $\|x^1 - x^*\|_2 < \|x^0 - x^*\|_2$, нужно предположить, что

$$\|x^0 - x^*\|_2 < \frac{2\mu}{M}.$$

- Поймем насколько быстро сходится метод. Пусть $M = 2$, $\mu = 1$, а $\|x^0 - x^*\|_2 = \frac{1}{2}$. Тогда мы можем гарантировать, что $\|x^1 - x^*\|_2 \leq \frac{1}{2^2}$, $\|x^2 - x^*\|_2 \leq \frac{1}{(2^2)^2}$ и так далее.

Метод Ньютона: модификации

- Пытаемся решить проблему локальной сходимости. Действуем по аналогии с градиентным спуском. **Вопрос:** идеи?

Метод Ньютона: модификации

- Пытаемся решить проблему локальной сходимости. Действуем по аналогии с градиентным спуском. **Вопрос:** идеи?
- Идея первая – шаг:

$$x^{k+1} = x^k - \gamma_k \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

Такой метод называется демпфированный метод Ньютона.

Метод Ньютона: модификации

- Пытаемся решить проблему локальной сходимости. Действуем по аналогии с градиентным спуском. **Вопрос:** идеи?
- Идея первая – шаг:

$$x^{k+1} = x^k - \gamma_k \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

Такой метод называется демпфированный метод Ньютона.
Вопрос: как выбирать шаг?

Метод Ньютона: модификации

- Пытаемся решить проблему локальной сходимости. Действуем по аналогии с градиентным спуском. **Вопрос:** идеи?
- Идея первая – шаг:

$$x^{k+1} = x^k - \gamma_k \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k).$$

Такой метод называется демпфированный метод Ньютона.

Вопрос: как выбирать шаг? Много разных способов, например, на прошлой лекции обсуждали линейный поиск:

$\arg \min_{\gamma} f(x^k + \gamma p_k)$, где $p_k = - \left(\nabla^2 f(x^k) \right)^{-1} \nabla f(x^k)$.

Метод Ньютона: модификации

- Идея вторая – «оценки сверху». В основе анализа градиентного спуска лежала оптимизация «оценки сверху» на функцию:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^d} \left(f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2} \|x - x^k\|_2^2 \right).$$

Метод Ньютона: модификации

- Идея вторая – «оценки сверху». В основе анализа градиентного спуска лежала оптимизация «оценки сверху» на функцию:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^d} \left(f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2} \|x - x^k\|_2^2 \right).$$

Вопрос: чему равно x^{k+1} ?

Метод Ньютона: модификации

- Идея вторая – «оценки сверху». В основе анализа градиентного спуска лежала оптимизация «оценки сверху» на функцию:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^d} \left(f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2} \|x - x^k\|_2^2 \right).$$

Вопрос: чему равно x^{k+1} ? $x^{k+1} = x^k - \frac{1}{L} \nabla f(x^k)$.

Метод Ньютона: модификации

- Идея вторая – «оценки сверху». В основе анализа градиентного спуска лежала оптимизация «оценки сверху» на функцию:

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^d} \left(f(x^k) + \langle \nabla f(x^k), x - x^k \rangle + \frac{L}{2} \|x - x^k\|_2^2 \right).$$

Вопрос: чему равно x^{k+1} ? $x^{k+1} = x^k - \frac{1}{L} \nabla f(x^k)$. Запишем, похожее для аппроксимации 2-го порядка:

$$\begin{aligned} x^{k+1} = \arg \min_{x \in \mathbb{R}^d} & \left(f(x^k) + \langle \nabla f(x^k), x - x^k \rangle \right. \\ & \left. + \frac{1}{2} \langle x - x^k, \nabla^2 f(x^k)(x - x^k) \rangle + \frac{M}{6} \|x - x^k\|_2^3 \right). \end{aligned}$$

Здесь M – константа Липшица гессиана. Такой метод называется кубический метод Ньютона.

Квазиньютоновское уравнение

- Запишем метод Ньютона следующим образом:

$$x^{k+1} = x^k - H_k \nabla f(x^k).$$

Квазиньютоновское уравнение

- Запишем метод Ньютона следующим образом:

$$x^{k+1} = x^k - H_k \nabla f(x^k).$$

В случае метода Ньютона вместо H_k стоит $(\nabla^2 f(x^k))^{-1}$.

- Хочется заменить $(\nabla^2 f(x^k))^{-1}$ на что-то более дешевое с точки зрения вычислений.

Квазиньютоновское уравнение

- Запишем метод Ньютона следующим образом:

$$x^{k+1} = x^k - H_k \nabla f(x^k).$$

В случае метода Ньютона вместо H_k стоит $(\nabla^2 f(x^k))^{-1}$.

- Хочется заменить $(\nabla^2 f(x^k))^{-1}$ на что-то более дешевое с точки зрения вычислений.
- Идея – выудить какие-то свойства, присущие гессиану.

Квазиньютоновское уравнение

- Запишем метод Ньютона следующим образом:

$$x^{k+1} = x^k - H_k \nabla f(x^k).$$

В случае метода Ньютона вместо H_k стоит $(\nabla^2 f(x^k))^{-1}$.

- Хочется заменить $(\nabla^2 f(x^k))^{-1}$ на что-то более дешевое с точки зрения вычислений.
- Идея – выудить какие-то свойства, присущие гессиану.
- Связь градиента и гессиана:

$$\nabla f(x^k) = \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})(x^k - x^{k+1}) + o(\|x^{k+1} - x^k\|_2)$$

или $\nabla f(x^k) - \nabla f(x^{k+1}) \approx \nabla^2 f(x^{k+1})(x^k - x^{k+1})$. Откуда
 $x^{k+1} - x^k \approx (\nabla^2 f(x^{k+1}))^{-1}(\nabla f(x^{k+1}) - \nabla f(x^k))$.

Квазиньютоновское уравнение

- Запишем метод Ньютона следующим образом:

$$x^{k+1} = x^k - H_k \nabla f(x^k).$$

В случае метода Ньютона вместо H_k стоит $(\nabla^2 f(x^k))^{-1}$.

- Хочется заменить $(\nabla^2 f(x^k))^{-1}$ на что-то более дешевое с точки зрения вычислений.
- Идея – выудить какие-то свойства, присущие гессиану.
- Связь градиента и гессиана:

$$\nabla f(x^k) = \nabla f(x^{k+1}) + \nabla^2 f(x^{k+1})(x^k - x^{k+1}) + o(\|x^{k+1} - x^k\|_2)$$

или $\nabla f(x^k) - \nabla f(x^{k+1}) \approx \nabla^2 f(x^{k+1})(x^k - x^{k+1})$. Откуда $x^{k+1} - x^k \approx (\nabla^2 f(x^{k+1}))^{-1}(\nabla f(x^{k+1}) - \nabla f(x^k))$. Заменим $(\nabla^2 f(x^{k+1}))^{-1}$ на H_{k+1} , введем обозначения $s^k = x^{k+1} - x^k$ и $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$:

$$s^k = H_{k+1} y^k$$

Квазиньютоновское уравнение

- Квазиньютоновское уравнение:

$$s^k = H_{k+1}y^k$$

Квазиньютоновское уравнение

- Квазиньютоновское уравнение:

$$s^k = H_{k+1} y^k$$

- Еще потребуем, чтобы H_{k+1} была симметричной: $H_{k+1}^T = H_{k+1}$.

Квазиньютоновское уравнение

- Квазиньютоновское уравнение:

$$s^k = H_{k+1} y^k$$

- Еще потребуем, чтобы H_{k+1} была симметричной: $H_{k+1}^T = H_{k+1}$.
- **Вопрос:** сколько решений имеет система уравнений $s^k = H_{k+1} y^k$ относительно H_{k+1} при условии, что $H_{k+1}^T = H_{k+1}$?

Квазиньютоновское уравнение

- Квазиньютоновское уравнение:

$$s^k = H_{k+1}y^k$$

- Еще потребуем, чтобы H_{k+1} была симметричной: $H_{k+1}^T = H_{k+1}$.
- Вопрос:** сколько решений имеет система уравнений $s^k = H_{k+1}y^k$ относительно H_{k+1} при условии, что $H_{k+1}^T = H_{k+1}$? d уравнений, $d + d(d - 1)/2$ уравнений. Можно урешаться.

Квазиньютоновское уравнение

- Квазиньютоновское уравнение:

$$s^k = H_{k+1}y^k$$

- Еще потребуем, чтобы H_{k+1} была симметричной: $H_{k+1}^T = H_{k+1}$.
- Вопрос:** сколько решений имеет система уравнений $s^k = H_{k+1}y^k$ относительно H_{k+1} при условии, что $H_{k+1}^T = H_{k+1}$? d уравнений, $d + d(d - 1)/2$ уравнений. Можно урешаться. Нужно еще сузить правила поиска H_{k+1} .

Квазиньютоновские методы: SR1/Broyden

- Идея первая – одно-ранговая (дешевая с точки зрения вычислений) добавка:

$$H_{k+1} = H_k + \mu_k q^k (q^k)^T,$$

где $\mu_k \in \mathbb{R}$ и $q^k \in \mathbb{R}^d$ нужно подобрать.

Квазиньютоновские методы: SR1/Broyden

- Идея первая – одно-ранговая (дешевая с точки зрения вычислений) добавка:

$$H_{k+1} = H_k + \mu_k q^k (q^k)^T,$$

где $\mu_k \in \mathbb{R}$ и $q^k \in \mathbb{R}^d$ нужно подобрать.

- Подбираем исходя из квазиньютоновского уравнения:

$$\begin{aligned} s^k &= H_{k+1} y^k = H_k y^k + \mu_k q^k (q^k)^T y^k \\ &= H_k y^k + \mu_k \left((q^k)^T y^k \right) q^k \end{aligned}$$

Квазиньютоновские методы: SR1/Broyden

- Идея первая – одно-ранговая (дешевая с точки зрения вычислений) добавка:

$$H_{k+1} = H_k + \mu_k q^k (q^k)^T,$$

где $\mu_k \in \mathbb{R}$ и $q^k \in \mathbb{R}^d$ нужно подобрать.

- Подбираем исходя из квазиньютоновского уравнения:

$$\begin{aligned} s^k &= H_{k+1} y^k = H_k y^k + \mu_k q^k (q^k)^T y^k \\ &= H_k y^k + \mu_k \left((q^k)^T y^k \right) q^k \end{aligned}$$

Откуда

$$\mu_k \left((q^k)^T y^k \right) q^k = s^k - H_k y^k$$

Квазиньютоновские методы: SR1/Broyden

- С предыдущего слайда:

$$\mu_k \left((q^k)^T y^k \right) q^k = s^k - H_k y^k$$

- **Вопрос:** что можно сказать про вектор q^k ?

Квазиньютоновские методы: SR1/Broyden

- С предыдущего слайда:

$$\mu_k \left((q^k)^T y^k \right) q^k = s^k - H_k y^k$$

- **Вопрос:** что можно сказать про вектор q^k ? Коллинеарен $s^k - H_k y^k$.

Квазиньютоновские методы: SR1/Broyden

- С предыдущего слайда:

$$\mu_k \left((q^k)^T y^k \right) q^k = s^k - H_k y^k$$

- Вопрос:** что можно сказать про вектор q^k ? Коллинеарен $s^k - H_k y^k$. Пусть

$$q^k = s^k - H_k y^k,$$

тогда

$$\mu_k = \frac{1}{(q^k)^T y^k}.$$

- Получаем SR1 способ подсчета матриц H :

$$H_{k+1} = H_k + \frac{(s^k - H_k y^k)(s^k - H_k y^k)^T}{(s^k - H_k y^k)^T y^k}$$

Квазиньютоновские методы: BFGS

- Посмотрим на задачу поиска H_{k+1} , как на задачу поиска «близкой» к H_k матрицы с точки зрения оптимизации:

$$\begin{aligned} H_{k+1} &= \arg \min_{H \in \mathbb{R}^{d \times d}} \|H - H_k\|^2 \\ \text{s.t. } s^k &= Hy^k \\ H^T &= H \end{aligned}$$

Квазиньютоновские методы: BFGS

- Посмотрим на задачу поиска H_{k+1} , как на задачу поиска «близкой» к H_k матрицы с точки зрения оптимизации:

$$\begin{aligned} H_{k+1} &= \arg \min_{H \in \mathbb{R}^{d \times d}} \|H - H_k\|^2 \\ \text{s.t. } s^k &= Hy^k \\ H^T &= H \end{aligned}$$

- Норма в задаче оптимизации может быть любая. В зависимости от нормы будут получаться разные квазиньютоновские методы.

Квазиньютоновские методы: BFGS

- Посмотрим на задачу поиска H_{k+1} , как на задачу поиска «близкой» к H_k матрицы с точки зрения оптимизации:

$$\begin{aligned} H_{k+1} &= \arg \min_{H \in \mathbb{R}^{d \times d}} \|H - H_k\|^2 \\ \text{s.t. } s^k &= Hy^k \\ H^T &= H \end{aligned}$$

- Норма в задаче оптимизации может быть любая. В зависимости от нормы будут получаться разные квазиньютоновские методы.
- Рассмотрим взвешенную норму Фробениуса $\|A\|_W = \|W^{1/2}AW^{1/2}\|_F$, где должно выполняться $Wy^k = s^k$. Такой выбор дает метод BFGS:

$$H_{k+1} = (I - \rho_k s^k (y^k)^T) H_k (I - \rho_k y^k (s^k)^T) + \rho_k s^k (s^k)^T, \text{ где } \rho_k = \frac{1}{(y^k)^T s^k}$$

Квазиньютоновские методы: BFGS

- До такой формулы можно дойти по-другому.

Квазиньютоновские методы: BFGS

- До такой формулы можно прийти по-другому. Рассмотрим $B_{k+1} = H_{k+1}^{-1}$. Для B квазиньютоновское уравнение выглядит как
$$B_{k+1}s^k = y^k$$

Квазиньютоновские методы: BFGS

- До такой формулы можно прийти по-другому. Рассмотрим $B_{k+1} = H_{k+1}^{-1}$. Для B квазиньютоновское уравнение выглядит как

$$B_{k+1}s^k = y^k$$

- Для B_{k+1} можно написать SR1 пересчет матрицы:

$$B_{k+1} = B_k + \frac{(y^k - B_k s^k)(y^k - B_k s^k)^T}{(y^k - B_k s^k)^T s^k}$$

Квазиньютоновские методы: BFGS

- До такой формулы можно прийти по-другому. Рассмотрим $B_{k+1} = H_{k+1}^{-1}$. Для B квазиньютоновское уравнение выглядит как

$$B_{k+1}s^k = y^k$$

- Для B_{k+1} можно написать SR1 пересчет матрицы:

$$B_{k+1} = B_k + \frac{(y^k - B_k s^k)(y^k - B_k s^k)^T}{(y^k - B_k s^k)^T s^k}$$

- Смотрим на вид B_{k+1} и делаем из нее двухранговое изменение:

$$B_{k+1} = B_k + \mu_{k,1} y^k (y^k)^T + \mu_{k,2} B_k y^k (B_k y^k)^T$$

Квазиньютоновские методы: BFGS

- До такой формулы можно прийти по-другому. Рассмотрим $B_{k+1} = H_{k+1}^{-1}$. Для B квазиньютоновское уравнение выглядит как

$$B_{k+1}s^k = y^k$$

- Для B_{k+1} можно написать SR1 пересчет матрицы:

$$B_{k+1} = B_k + \frac{(y^k - B_k s^k)(y^k - B_k s^k)^T}{(y^k - B_k s^k)^T s^k}$$

- Смотрим на вид B_{k+1} и делаем из нее двухранговое изменение:

$$B_{k+1} = B_k + \mu_{k,1} y^k (y^k)^T + \mu_{k,2} B_k y^k (B_k y^k)^T$$

- Как и в SR1 можно подогнать $\mu_{k,1}$ и $\mu_{k,2}$:

$$B_{k+1} = B_k + \frac{y^k (y^k)^T}{(y^k)^T s^k} + \frac{B_k s^k (B_k s^k)^T}{(s^k)^T B_k s^k}$$

Квазиньютоновские методы: BFGS

- До такой формулы можно прийти по-другому. Рассмотрим $B_{k+1} = H_{k+1}^{-1}$. Для B квазиньютоновское уравнение выглядит как

$$B_{k+1}s^k = y^k$$

- Для B_{k+1} можно написать SR1 пересчет матрицы:

$$B_{k+1} = B_k + \frac{(y^k - B_k s^k)(y^k - B_k s^k)^T}{(y^k - B_k s^k)^T s^k}$$

- Смотрим на вид B_{k+1} и делаем из нее двухранговое изменение:

$$B_{k+1} = B_k + \mu_{k,1} y^k (y^k)^T + \mu_{k,2} B_k y^k (B_k y^k)^T$$

- Как и в SR1 можно подогнать $\mu_{k,1}$ и $\mu_{k,2}$:

$$B_{k+1} = B_k + \frac{y^k (y^k)^T}{(y^k)^T s^k} + \frac{B_k s^k (B_k s^k)^T}{(s^k)^T B_k s^k}$$

- Если теперь обратить B_{k+1} (формула Шермана-Маррисона-Вудберри), то получится выражение для H_{k+1}

Квазиньютоновские методы: BFGS

- **Вопрос:** чтобы посчитать новую матрицу нужно $O(d^2)$ операций (не учитывая подсчет градиентов). Кажется, что BFGS подсчет дороже (есть перемножение трех матриц). Так ли это?
$$H_{k+1} = (I - \rho_k s^k (y^k)^T) H_k (I - \rho_k y^k (s^k)^T) + \rho_k s^k (s^k)^T$$

Квазиньютоновские методы: BFGS

- **Вопрос:** чтобы посчитать новую матрицу нужно $O(d^2)$ операций (не учитывая подсчет градиентов). Кажется, что BFGS подсчет дороже (есть перемножение трех матриц). Так ли это?
$$H_{k+1} = (I - \rho_k s^k (y^k)^T) H_k (I - \rho_k y^k (s^k)^T) + \rho_k s^k (s^k)^T$$
- Нужно раскрыть скобки в матричном умножении. В подсчете $s^k (y^k)^T H_k$ нужно сначала умножить $(y^k)^T H_k$, а потом вектор на вектор. Аналогично для $H_k y^k (s^k)^T$. Получается, что сложность BFGS есть $O(d^2)$ операций (не учитывая подсчет градиентов).

Квазиньютоновские методы: BFGS

- **Вопрос:** чтобы посчитать новую матрицу нужно $O(d^2)$ операций (не учитывая подсчет градиентов). Кажется, что BFGS подсчет дороже (есть перемножение трех матриц). Так ли это?
$$H_{k+1} = (I - \rho_k s^k (y^k)^T) H_k (I - \rho_k y^k (s^k)^T) + \rho_k s^k (s^k)^T$$
- Нужно раскрыть скобки в матричном умножении. В подсчете $s^k (y^k)^T H_k$ нужно сначала умножить $(y^k)^T H_k$, а потом вектор на вектор. Аналогично для $H_k y^k (s^k)^T$. Получается, что сложность BFGS есть $O(d^2)$ операций (не учитывая подсчет градиентов).
- При инициализации матрицу H_0 достаточно брать равно единичной. Есть и более хитрые способы, но особо разницы не чувствует все работает хорошо.

Метод Ньютона и квазиньютоновские методы

- Квазиньютоновские методы не требуют подсчет гессиана и его обращение. Сложность всех арифметических операций на одной итерации $O(d^2)$, что дешевле обращения гессиана за $O(d^3)$.

Метод Ньютона и квазиньютоновские методы

- Квазиньютоновские методы не требуют подсчет гессиана и его обращение. Сложность всех арифметических операций на одной итерации $O(d^2)$, что дешевле обращения гессиана за $O(d^3)$.
- Квазиньютоновские методы имеют глобальную сверхлинейную скорость сходимости. Это медленнее, чем метод Ньютона, но не нужна «хорошая» окрестность решения.

Метод Ньютона и квазиньютоновские методы

- Квазиньютоновские методы не требуют подсчет гессиана и его обращение. Сложность всех арифметических операций на одной итерации $O(d^2)$, что дешевле обращения гессиана за $O(d^3)$.
- Квазиньютоновские методы имеют глобальную сверхлинейную скорость сходимости. Это медленнее, чем метод Ньютона, но не нужна «хорошая» окрестность решения.
- Квазиньютоновские методы используют только градиент, но в теории сходятся быстрее ускоренного градиентного метода.
Вопрос: почему так, ведь метод Нестерова оптимальный?

Метод Ньютона и квазиньютоновские методы

- Квазиньютоновские методы не требуют подсчет гессиана и его обращение. Сложность всех арифметических операций на одной итерации $O(d^2)$, что дешевле обращения гессиана за $O(d^3)$.
- Квазиньютоновские методы имеют глобальную сверхлинейную скорость сходимости. Это медленнее, чем метод Ньютона, но не нужна «хорошая» окрестность решения.
- Квазиньютоновские методы используют только градиент, но в теории сходятся быстрее ускоренного градиентного метода.

Вопрос: почему так, ведь метод Нестерова оптимальный?

Смотри определения класса задач, для которого метод Нестерова оптимальный: не разрешены векторные произведения.

Метод Ньютона и квазиньютоновские методы

- Квазиньютоновские методы не требуют подсчет гессиана и его обращение. Сложность всех арифметических операций на одной итерации $O(d^2)$, что дешевле обращения гессиана за $O(d^3)$.
- Квазиньютоновские методы имеют глобальную сверхлинейную скорость сходимости. Это медленнее, чем метод Ньютона, но не нужна «хорошая» окрестность решения.
- Квазиньютоновские методы используют только градиент, но в теории сходятся быстрее ускоренного градиентного метода.
Вопрос: почему так, ведь метод Нестерова оптимальный?
Смотри определения класса задач, для которого метод Нестерова оптимальный: не разрешены векторные произведения.
- Метод Ньютона и квазиньютоновские методы на практике быстро находят точный локальный минимум. Их спокойно можно использовать в качестве «дорешивателей». Квазиньютоновские методы и как «стартовый» метод.