

МЕТОДЫ ВЫПУКЛОЙ ОПТИМИЗАЦИИ

НЕДЕЛЯ 7

Даня Меркулов Пётр Остроухов



Градиентный спуск. Скорости сходимости Семинар

Оптимизация для всех! ЦУ



Воспоминания с лекции



Предположим, у нас есть задача минимизации гладкой функции $f(x):\mathbb{R}^n o \mathbb{R}$:

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



Предположим, у нас есть задача минимизации гладкой функции $f(x): \mathbb{R}^n \to \mathbb{R}$:

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

Один из методов решения этой задачи — градиентный спуск:

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



Предположим, у нас есть задача минимизации гладкой функции $f(x): \mathbb{R}^n \to \mathbb{R}$:

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

Один из методов решения этой задачи — градиентный спуск:

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

Бутылочным горлышком (почти для всех градиентных методов) является выбор шага, который может привести к драматическим различиямв поведении метода.



• Одно из теоретических предложений: выбирать шаг, обратно пропорциональный константе Липшица градиента

$$\eta_k = \frac{1}{L}$$



• Одно из теоретических предложений: выбирать шаг, обратно пропорциональный константе Липшица градиента

$$\eta_k = \frac{1}{L}$$



• Одно из теоретических предложений: выбирать шаг, обратно пропорциональный константе Липшица градиента

$$\eta_k = \frac{1}{L}$$

• Линейный поиск с возвратом (backtracking). Зафиксируем два параметра: $0<\beta<1$ и $0<\alpha\leq0.5$. На каждой итерации начинаем с t=1, и пока

$$f(x_k - t \nabla f(x_k)) > f(x_k) - \alpha t \|\nabla f(x_k)\|_2^2,$$

уменьшаем $t=\beta t$. Иначе выполняем обновление градиентного спуска $x_{k+1}=x_k-t\nabla f(x_k)$.



• Одно из теоретических предложений: выбирать шаг, обратно пропорциональный константе Липшица градиента

$$\eta_k = \frac{1}{L}$$

• Линейный поиск с возвратом (backtracking). Зафиксируем два параметра: $0<\beta<1$ и $0<\alpha\leq0.5$. На каждой итерации начинаем с t=1, и пока

$$f(x_k - t \nabla f(x_k)) > f(x_k) - \alpha t \|\nabla f(x_k)\|_2^2,$$

уменьшаем $t=\beta t$. Иначе выполняем обновление градиентного спуска $x_{k+1}=x_k-t\nabla f(x_k)$.



• Одно из теоретических предложений: выбирать шаг, обратно пропорциональный константе Липшица градиента

$$\eta_k = \frac{1}{L}$$

• Линейный поиск с возвратом (backtracking). Зафиксируем два параметра: $0<\beta<1$ и $0<\alpha\leq0.5$. На каждой итерации начинаем с t =1, и пока

$$f(x_k - t \nabla f(x_k)) > f(x_k) - \alpha t \|\nabla f(x_k)\|_2^2,$$

уменьшаем $t=\beta t$. Иначе выполняем обновление градиентного спуска $x_{k+1}=x_k-t\nabla f(x_k)$.

• Точный линейный поиск.

$$\eta_k = \underset{\eta \geq 0}{\arg\min} f(x_k - \eta \nabla f(x_k))$$







Рассмотрим линейное приближение дифференцируемой функции f вдоль некоторого направления $h,\|h\|_2=1$:



Рассмотрим линейное приближение дифференцируемой функции f вдоль некоторого направления $h,\|h\|_2=1$:

$$f(x+\alpha h) = f(x) + \alpha \langle f'(x), h \rangle + o(\alpha)$$



Рассмотрим линейное приближение дифференцируемой функции f вдоль некоторого направления $h,\|h\|_2=1$:

$$f(x+\alpha h) = f(x) + \alpha \langle f'(x), h \rangle + o(\alpha)$$

Мы хотим, чтобы h было убывающим направлением:

$$f(x + \alpha h) < f(x)$$

$$f(x) + \alpha \langle f'(x), h \rangle + o(\alpha) < f(x)$$



Рассмотрим линейное приближение дифференцируемой функции f вдоль некоторого направления $h,\|h\|_2=1$:

$$f(x + \alpha h) = f(x) + \alpha \langle f'(x), h \rangle + o(\alpha)$$

Мы хотим, чтобы h было убывающим направлением:

$$f(x + \alpha h) < f(x)$$

$$f(x) + \alpha \langle f'(x), h \rangle + o(\alpha) < f(x)$$

и переходя к пределу при lpha o 0:

$$\langle f'(x),h\rangle \leq 0$$



Рассмотрим линейное приближение дифференцируемой функции f вдоль некоторого направления $h,\|h\|_2=1$:

Также из неравенства Коши — Буняковского — Шварца:

$$f(x + \alpha h) = f(x) + \alpha \langle f'(x), h \rangle + o(\alpha)$$

$$\begin{split} |\langle f'(x), h \rangle| & \leq \|f'(x)\|_2 \|h\|_2 \\ \langle f'(x), h \rangle & \geq -\|f'(x)\|_2 \|h\|_2 = -\|f'(x)\|_2 \end{split}$$

Мы хотим, чтобы h было убывающим направлением:

$$f(x + \alpha h) < f(x)$$

$$f(x) + \alpha \langle f'(x), h \rangle + o(\alpha) < f(x)$$

и переходя к пределу при lpha o 0:

$$\langle f'(x),h\rangle \leq 0$$



Рассмотрим линейное приближение дифференцируемой функции f вдоль некоторого направления $h,\|h\|_2=1$:

$$f(x+\alpha h)=f(x)+\alpha \langle f'(x),h\rangle +o(\alpha)$$

Мы хотим, чтобы h было убывающим направлением:

$$f(x+\alpha h) < f(x)$$

$$f(x) + \alpha \langle f'(x), h \rangle + o(\alpha) < f(x)$$

и переходя к пределу при lpha o 0:

$$\langle f'(x),h\rangle \leq 0$$

Также из неравенства Коши — Буняковского — Шварца:

$$\begin{split} |\langle f'(x), h \rangle| &\leq \|f'(x)\|_2 \|h\|_2 \\ \langle f'(x), h \rangle &\geq -\|f'(x)\|_2 \|h\|_2 = -\|f'(x)\|_2 \end{split}$$

Таким образом, направление антиградиента

$$h = -\frac{f'(x)}{\|f'(x)\|_2}$$

дает направление **локального наискорейшего** убывания функции f.



Рассмотрим линейное приближение дифференцируемой функции f вдоль некоторого направления $h,\|h\|_2=1$:

$$f(x+\alpha h)=f(x)+\alpha \langle f'(x),h\rangle +o(\alpha)$$

Мы хотим, чтобы h было убывающим направлением:

$$f(x + \alpha h) < f(x)$$

$$f(x) + \alpha \langle f'(x), h \rangle + o(\alpha) < f(x)$$

и переходя к пределу при lpha o 0:

$$\langle f'(x),h\rangle \leq 0$$

Также из неравенства Коши — Буняковского — Шварца:

$$\begin{split} |\langle f'(x), h \rangle| &\leq \|f'(x)\|_2 \|h\|_2 \\ \langle f'(x), h \rangle &\geq -\|f'(x)\|_2 \|h\|_2 = -\|f'(x)\|_2 \end{split}$$

Таким образом, направление антиградиента

$$h=-\frac{f'(x)}{\|f'(x)\|_2}$$

дает направление **локального наискорейшего** убывания функции f. Результатом этого метода является

$$x_{k+1} = x_k - \alpha f'(x_k)$$



Если функция $f:\mathbb{R}^n o \mathbb{R}$ непрерывно дифференцируема и ее градиент удовлетворяет условиям Липшица с константой L, то $\forall x,y \in \mathbb{R}^n$:

$$|f(y)-f(x)-\langle\nabla f(x),y-x\rangle|\leq \frac{L}{2}\|y-x\|^2,$$



Если функция $f:\mathbb{R}^n \to \mathbb{R}$ непрерывно дифференцируема и ее градиент удовлетворяет условиям Липшица с константой L, то $\forall x,y \in \mathbb{R}^n$:

$$|f(y)-f(x)-\langle\nabla f(x),y-x\rangle|\leq \frac{L}{2}\|y-x\|^2,$$

что геометрически означает, что если мы зафиксируем некоторую точку $x_0 \in \mathbb{R}^n$ и определим две параболы:

$$\begin{split} \phi_1(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle - \frac{L}{2} \|x - x_0\|^2, \\ \phi_2(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{L}{2} \|x - x_0\|^2. \end{split}$$



Если функция $f:\mathbb{R}^n \to \mathbb{R}$ непрерывно дифференцируема и ее градиент удовлетворяет условиям Липшица с константой L, то $\forall x,y \in \mathbb{R}^n$:

$$|f(y)-f(x)-\langle\nabla f(x),y-x\rangle|\leq \frac{L}{2}\|y-x\|^2,$$

что геометрически означает, что если мы зафиксируем некоторую точку $x_0 \in \mathbb{R}^n$ и определим две параболы:

$$\begin{split} \phi_1(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle - \frac{L}{2} \|x - x_0\|^2, \\ \phi_2(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{L}{2} \|x - x_0\|^2. \end{split}$$

Тогда

$$\phi_1(x) \le f(x) \le \phi_2(x) \quad \forall x \in \mathbb{R}^n.$$



Если функция $f:\mathbb{R}^n o \mathbb{R}$ непрерывно дифференцируема и ее градиент удовлетворяет условиям Липшица с константой L, то $\forall x,y \in \mathbb{R}^n$:

$$|f(y)-f(x)-\langle\nabla f(x),y-x\rangle|\leq \frac{L}{2}\|y-x\|^2,$$

что геометрически означает, что если мы зафиксируем некоторую точку $x_0 \in \mathbb{R}^n$ и определим две параболы:

$$\begin{split} \phi_1(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle - \frac{L}{2} \|x - x_0\|^2, \\ \phi_2(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{L}{2} \|x - x_0\|^2. \end{split}$$

Тогда

$$\phi_1(x) \leq f(x) \leq \phi_2(x) \quad \forall x \in \mathbb{R}^n.$$

Теперь, если у нас есть глобальная верхняя граница функции в виде параболы, мы можем попробовать сразу попасть в ее минимум.



Если функция $f:\mathbb{R}^n o \mathbb{R}$ непрерывно дифференцируема и ее градиент удовлетворяет условиям Липшица с константой L, то $\forall x,y \in \mathbb{R}^n$:

$$|f(y)-f(x)-\langle\nabla f(x),y-x\rangle|\leq \frac{L}{2}\|y-x\|^2,$$

что геометрически означает, что если мы зафиксируем некоторую точку $x_0 \in \mathbb{R}^n$ и определим две параболы:

$$\begin{split} \phi_1(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle - \frac{L}{2} \|x - x_0\|^2, \\ \phi_2(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{L}{2} \|x - x_0\|^2. \end{split}$$

Тогда

$$\phi_1(x) \leq f(x) \leq \phi_2(x) \quad \forall x \in \mathbb{R}^n.$$

Теперь, если у нас есть глобальная верхняя граница функции в виде параболы, мы можем попробовать сразу попасть в ее минимум.

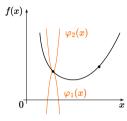


Рисунок 1. Иллюстрация



Если функция $f:\mathbb{R}^n o \mathbb{R}$ непрерывно дифференцируема и ее градиент удовлетворяет условиям Липшица с константой L, то $\forall x,y \in \mathbb{R}^n$:

$$|f(y)-f(x)-\langle\nabla f(x),y-x\rangle|\leq \frac{L}{2}\|y-x\|^2,$$

что геометрически означает, что если мы зафиксируем некоторую точку $x_0 \in \mathbb{R}^n$ и определим две параболы:

$$\begin{split} \phi_1(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle - \frac{L}{2} \|x - x_0\|^2, \\ \phi_2(x) &= f(x_0) + \langle \nabla f(x_0), x - x_0 \rangle + \frac{L}{2} \|x - x_0\|^2. \end{split}$$

Тогда

$$\phi_1(x) \leq f(x) \leq \phi_2(x) \quad \forall x \in \mathbb{R}^n.$$

Теперь, если у нас есть глобальная верхняя граница функции в виде параболы, мы можем попробовать сразу попасть в ее минимум.

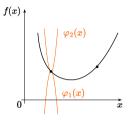
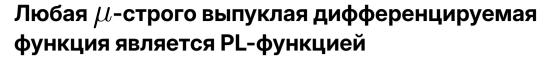


Рисунок 1. Иллюстрация

$$\begin{split} \nabla\phi_2(x) &= 0\\ \nabla f(x_0) + L(x^* - x_0) &= 0\\ x^* &= x_0 - \frac{1}{L} \nabla f(x_0)\\ x_{k+1} &= x_k - \frac{1}{L} \nabla f(x_k) \end{split}$$

Таким образом, мы получаем шаг $\frac{1}{L}$. Однако часто константа L неизвестна.





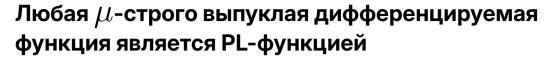
Если функция f(x) дифференцируема и μ -строго выпукла, то она является PL-функцией.

Доказательство

По критерию строгой выпуклости первого порядка:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

$$f(x^*) \geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2$$





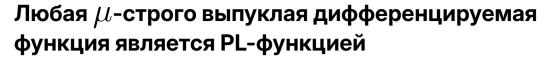
Если функция f(x) дифференцируема и μ -строго выпукла, то она является PL-функцией.

Доказательство

По критерию строгой выпуклости первого порядка:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

$$\begin{split} f(x^*) & \geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2 \\ f(x) - f(x^*) & \leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2} \|x^* - x\|_2^2 = \end{split}$$





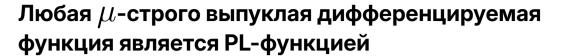
Если функция f(x) дифференцируема и μ -строго выпукла, то она является PL-функцией.

Доказательство

По критерию строгой выпуклости первого порядка:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

$$\begin{split} f(x^*) & \geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2 \\ f(x) - f(x^*) & \leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2} \|x^* - x\|_2^2 = \\ & = \left(\nabla f(x)^T - \frac{\mu}{2} (x^* - x) \right)^T (x - x^*) = \end{split}$$





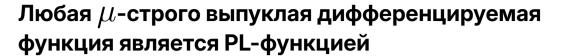
Если функция f(x) дифференцируема и μ -строго выпукла, то она является PL-функцией.

Доказательство

По критерию строгой выпуклости первого порядка:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

$$\begin{split} f(x^*) &\geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2 \\ f(x) - f(x^*) &\leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2} \|x^* - x\|_2^2 = \\ &= \left(\nabla f(x)^T - \frac{\mu}{2} (x^* - x)\right)^T (x - x^*) = \\ &= \frac{1}{2} \left(\frac{2}{\sqrt{\mu}} \nabla f(x)^T - \sqrt{\mu} (x^* - x)\right)^T \sqrt{\mu} (x - x^*) \end{split}$$





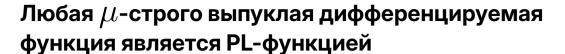
Если функция f(x) дифференцируема и μ -строго выпукла, то она является PL-функцией.

Доказательство

По критерию строгой выпуклости первого порядка:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

$$\begin{split} f(x^*) &\geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2 \\ f(x) - f(x^*) &\leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2} \|x^* - x\|_2^2 = \\ &= \left(\nabla f(x)^T - \frac{\mu}{2} (x^* - x)\right)^T (x - x^*) = \\ &= \frac{1}{2} \left(\frac{2}{\sqrt{\mu}} \nabla f(x)^T - \sqrt{\mu} (x^* - x)\right)^T \sqrt{\mu} (x - x^*) \end{split}$$





Если функция f(x) дифференцируема и μ -строго выпукла, то она является PL-функцией.

Доказательство

По критерию строгой выпуклости первого порядка:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

Пусть
$$a=\frac{1}{\sqrt{\mu}}\nabla f(x)$$
 \ и
$$b=\sqrt{\mu}(x-x^*)-\frac{1}{\sqrt{\mu}}\nabla f(x)$$

$$\begin{split} f(x^*) &\geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2 \\ f(x) - f(x^*) &\leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2} \|x^* - x\|_2^2 = \\ &= \left(\nabla f(x)^T - \frac{\mu}{2} (x^* - x)\right)^T (x - x^*) = \\ &= \frac{1}{2} \left(\frac{2}{\sqrt{\mu}} \nabla f(x)^T - \sqrt{\mu} (x^* - x)\right)^T \sqrt{\mu} (x - x^*) \end{split}$$



Любая μ -строго выпуклая дифференцируемая функция является PL-функцией

i Theorem

Если функция f(x) дифференцируема и μ -строго выпукла, то она является PL-функцией.

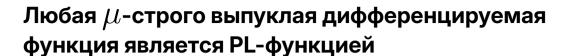
Доказательство

По критерию строгой выпуклости первого порядка:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\mu}{2} \|y-x\|_2^2$$

$$\begin{split} f(x^*) &\geq f(x) + \nabla f(x)^T (x^* - x) + \frac{\mu}{2} \|x^* - x\|_2^2 \\ f(x) - f(x^*) &\leq \nabla f(x)^T (x - x^*) - \frac{\mu}{2} \|x^* - x\|_2^2 = \\ &= \left(\nabla f(x)^T - \frac{\mu}{2} (x^* - x)\right)^T (x - x^*) = \\ &= \frac{1}{2} \left(\frac{2}{\sqrt{\mu}} \nabla f(x)^T - \sqrt{\mu} (x^* - x)\right)^T \sqrt{\mu} (x - x^*) \end{split}$$

Пусть
$$a=\frac{1}{\sqrt{\mu}}\nabla f(x)$$
 \ и
$$b=\sqrt{\mu}(x-x^*)-\frac{1}{\sqrt{\mu}}\nabla f(x)$$
 Тогда $a+b=\sqrt{\mu}(x-x^*)$ и $a-b=\frac{2}{\sqrt{\mu}}\nabla f(x)-\sqrt{\mu}(x-x^*)$





$$f(x) - f(x^*) \leq \frac{1}{2} \left(\frac{1}{\mu} \| \nabla f(x) \|_2^2 - \left\| \sqrt{\mu} (x - x^*) - \frac{1}{\sqrt{\mu}} \nabla f(x) \right\|_2^2 \right)$$



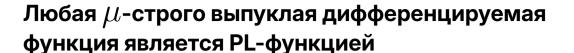
Любая μ -строго выпуклая дифференцируемая функция является PL-функцией

$$\begin{split} f(x) - f(x^*) &\leq \frac{1}{2} \left(\frac{1}{\mu} \|\nabla f(x)\|_2^2 - \left\| \sqrt{\mu} (x - x^*) - \frac{1}{\sqrt{\mu}} \nabla f(x) \right\|_2^2 \right) \\ f(x) - f(x^*) &\leq \frac{1}{2\mu} \|\nabla f(x)\|_2^2, \end{split}$$



Любая μ -строго выпуклая дифференцируемая функция является PL-функцией

$$\begin{split} f(x) - f(x^*) &\leq \frac{1}{2} \left(\frac{1}{\mu} \|\nabla f(x)\|_2^2 - \left\| \sqrt{\mu} (x - x^*) - \frac{1}{\sqrt{\mu}} \nabla f(x) \right\|_2^2 \right) \\ f(x) - f(x^*) &\leq \frac{1}{2\mu} \|\nabla f(x)\|_2^2, \end{split}$$





$$\begin{split} f(x) - f(x^*) &\leq \frac{1}{2} \left(\frac{1}{\mu} \| \nabla f(x) \|_2^2 - \left\| \sqrt{\mu} (x - x^*) - \frac{1}{\sqrt{\mu}} \nabla f(x) \right\|_2^2 \right) \\ f(x) - f(x^*) &\leq \frac{1}{2\mu} \| \nabla f(x) \|_2^2, \end{split}$$

что точно соответствует условию PL. Это означает, что мы уже имеем доказательство линейной сходимости для любой строго выпуклой функции.

Точный линейный поиск, или наискорейший спуск



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

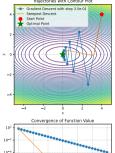
Более теоретический, чем практический подход. Он также позволяет анализировать сходимость, но часто точный линейный поиск может быть сложным, если вычисление функции занимает слишком много времени или стоит слишком дорого.

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

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

Условия оптимальности:

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



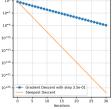


Рисунок 2. Наискорейший спуск

Открыть в Colab 🖺



Практика!



Рассмотрим задачу

$$\min_{x\in\mathbb{R}^n} f(x),$$

где f(x) выпукла и L-гладкая. Найдите скорость сходимости градиентного спуска с оптимальным теоретическим шагом $\eta_k=\frac{1}{L}$ для усредненной точки и для лучшей точки. Другими словами, получите верхние границы на

- + $f(\bar{x}_N) f^*$, where $\bar{x}_N = \frac{1}{N} \sum_{i=0}^{N-1} x_i$,
 - і Шаг градиентного спуска

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



Рассмотрим задачу

$$\min_{x \in \mathbb{R}^n} f(x),$$

где f(x) выпукла и L-гладкая. Найдите скорость сходимости градиентного спуска с оптимальным теоретическим шагом $\eta_k=\frac{1}{L}$ для усредненной точки и для лучшей точки. Другими словами, получите верхние границы на

- $f(\bar{x}_N) f^*$, where $\bar{x}_N = \frac{1}{N} \sum_{i=0}^{N-1} x_i$,
- $\bullet \ \min\nolimits_{0 \leq i \leq N-1} f(x_i) f^*.$
 - і Шаг градиентного спуска

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



Рассмотрим задачу

$$\min_{x \in \mathbb{R}^n} f(x),$$

где f(x) выпукла и L-гладкая. Найдите скорость сходимости градиентного спуска с оптимальным теоретическим шагом $\eta_k=\frac{1}{L}$ для усредненной точки и для лучшей точки. Другими словами, получите верхние границы на

- $f(\bar{x}_N) f^*$, where $\bar{x}_N = \frac{1}{N} \sum_{i=0}^{N-1} x_i$,
- $\bullet \ \min\nolimits_{0 \leq i \leq N-1} f(x_i) f^*.$
 - і Шаг градиентного спуска

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



Рассмотрим задачу

$$\min_{x\in\mathbb{R}^n} f(x),$$

где f(x) выпукла и L-гладкая. Найдите скорость сходимости градиентного спуска с оптимальным теоретическим шагом $\eta_k=\frac{1}{L}$ для усредненной точки и для лучшей точки. Другими словами, получите верхние границы на

- + $f(\bar{x}_N) f^*$, where $\bar{x}_N = \frac{1}{N} \sum_{i=0}^{N-1} x_i$,
- $\bullet \ \min\nolimits_{0 \leq i \leq N-1} f(x_i) f^*.$
 - і Шаг градиентного спуска

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

Совет

Используйте факт, что $\Psi_k(x)$ является L-строго выпуклой из-за квадратичного регуляризатора.

Код



Примеры: **@**code snippet.