Домашнее задание №2. Ускоренные градиентные методы. Стохастический градиентный спуск.

1 (2 балл)

Получите явные формулы для вычисления α_k, β_k для сильно выпуклой квадратичной функции

$$f(x) = \frac{1}{2}x^\top \mathrm{A}x + \mathrm{b}^\top x + \mathrm{c},$$

$$\mathrm{A} = \mathrm{A}^\top \succ 0.$$

Коэффициент β_k необходимо выразить через $\nabla f(x_{k+1})$, A и направление h_k.

2 (2 балл)

Для функции

$$f(x) = x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_1x_3 + x_2x_3 + x_1 + x_2,$$

$$x_0 = (0,0,1)^\top$$

сделайте три шага методом сопряженных градиентов. Проверьте, что для выпуклой квадратичной функции МСГ сходится не более чем за n шагов, где n – размерность пространства.

3 (2 балла)

Разберите и оформите доказательство сходимости любого варианта метода Нестерова. В доказательстве необходимо сделать акценты на тех вещах, которые были изучены в курсе методов оптимизации(сильная выпуклость, L-гладкость итд) и полученной оценке скорости сходимости, которую необходимо сравнить с нижней оценкой для заданного класса.

Нижние оценки можно найти в конце семинара 4: https://canvas.instructure.com/courses/1844771/pages/uskoriennyie-ghradiientnyie-mietody.

Литература: https://canvas.instructure.com/courses/1844771/pages/litieratura.

4

Разберите семинар, посвященный стохастическому градиентному спуску. https://canvas.instructure.com/courses/1844771/pages/stokhastichieskii-ghradiientnyi-spusk

5 (5 баллов)

1. Реализуйте метод Нестерова(из пункта 3), метод сопряженных градиентов и градиентный спуск.
2. Задайте три квадратичные функции ($n = 2, 10, 1000$) с разными числами обусловленности ($\varkappa = 1,\ 100,\ 10000$). Всего у вас получается 9 разных матриц A. Запустите на них методы следующим образом:

- для метода Нестерова рассмотрите разные начальные условия, а именно в одном случае $x_0 = y_0 = 0_n$, в другом $x_0 = 0_n,\ y_0 = 1_n$;
- рассмотрите разную точность для остановки;
- в методе градиентного спуска α необходимо выбирать по правилу Армихо.

Постарайтесь ответить на вопрос: как зависит поведение методов от числа обусловленности и от начальной точки?

6 Замечания

6.1 Критерий остановки

В этом задании используйте следующий критерий остановки:

$$\frac{\|\nabla f(x_k)\|_2^2}{\|\nabla f(x_0)\|_2^2} \le \varepsilon$$

Этот критерий задает относительную точность решения благодаря нормировке на $\|\nabla f(x_0)\|_2^2$.

6.2 Квадратичная функция

Рассмотрим матрицу $\mathrm{A} \in \mathbf{S}^n_{++}$ и вектор $\mathrm{b} \in \mathbf{R}^n$ и зададим функцию

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

Её число обсуловленности зависит от матрицы A и равняется $\varkappa = \frac{\lambda_{\max}(\mathrm{A})}{\lambda_{\min}(\mathrm{A})} = \frac{L}{\mu}$, где L– константа Липшица градиента, а μ - константа сильной выпуклости. В задании вам необходимо исследовать поведение методов в зависимости от числа обусловленности.

Сгенерировать случайную квадратичную задачу с заданным $\varkappa$ можно, например, так: взять случайные числа $\lambda_1,\ \lambda_2,\ \ldots, \lambda_n \in [1, \varkappa]$, так что $\min\limits_i \lambda_i = 1$, а $\max\limits_i \lambda_i = \varkappa$, и положить $\mathrm{A} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$. Элементы вектора b можно взять произвольными, они на обусловленность не влияют. В случае двумерной функции $x \in \mathbf{R}^2$ можно взять ортогональную матрицу вида

$$\mathrm{S} = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix}$$

и задать $\mathrm{A} = \mathrm{S} \cdot \mathrm{diag}(1, \varkappa) \cdot \mathrm{S}^\top$. Получится "повернутая"кадратичная функция.