Лабораторная работа 3. Ускорение град. спуска, метод сопряженных градиентов

Квадратичные функции

Для многих методов нам очень интересны квадратичные функции вида

$$f(x) = \frac{1}{2}x^T Ax - bx, \text{ где } A = A^T > 0$$

Научимся генерировать случайные матрицы с помощью следующего алгоритма:

1. Сгенерируем матрицу A_{rand}, наполненную случайными значениями от 0 до 1 *np.random.random((n, n))*

2. Проведем процедуру симметризации:

$$A_{\text{symm}} = 0.5(A_{\text{rand}} + A_{\text{rand}}^T)$$

3. Гарантируем, что матрица будет положительно определенной

$$A_{\text{spd}} = A_{\text{symm}} + nI_n$$

4. По необходимости, можно делать матрицу более или менее обусловленной, добавляя или вычитая матрицы вида aI_n из A.

5. Остается сгенерировать матрицу b любым удобным вам способом.

В рамках этого пункта работы:

1. Реализуйте алгоритм генерации квадратичной функции произвольной размерности;

2. Реализуйте метод тяжелого шара, метод Нестерова (с постоянными коэффициентами) и метод сопряженных градиентов;

3. Сравните (графики!) поведение методов между собой и с методом градиентного спуска (стратегия с $\alpha = \frac{1}{L}$), реализованным в первой лабораторной. Для этого проведите эксперементы на случайных начальных условиях, для функций разной размерности (не менее четырех размерностей от 2 до 1000);

4. Сравните поведение методов с аналитическими ожиданиями, сделайте выводы.

Почти квадратичная функция

Для функции $f(x) = \frac{1}{2}x_1^2 + \frac{1}{4}x_2^4 - \frac{1}{2}x_2^2$:

1. Постройте график функции, найдите точки минимума;

2. Модифицируйте метод сопряженных градиентов до метода Флетчера-Ривса. В качестве линейного поиска можно использовать как метод точной оптимизации, так и неточной;

3. Проведите эксперименты с методами тяжелого шара, Нестерова и Флетчера-Ривса (см подсказку в конце). Имперически оцените области, в которой методы применимы и сходятся.

4. Сравните поведение методов между собой и с градиентным спуском.

Функция Розенброка

Для функции Розенброка $f(x, y) = (1-x)^2 + 100(y-x^2)^2$

1. Модифицируйте метод Нестерова до версии с пересчитываемыми коэффициентами;

2. Примените метод с постоянными коэффициентами и с пересчитываемыми. Сравните результат с аналитическими ожиданиями. Постройте графики сходимости.

3. Реализуйте хотя бы 3 модификации метода сопряженных градиентов (см. подсказку в конце). Сравните между собой модификации и метод градиентного спуска. В какой области модификации дают результат?

Отчет

Отчет должен состоять из:

1. Кода для каждого из разделов;

2. Предложенных к построению графиков;

3. Аналитических выкладок;

4. Выводов к каждому разделу.

Подсказка с модификациями метода сопряженных градиентов

Нелинейный метод CG. Версия Флетчера–Ривса

Обычный CG:

$g_0 \leftarrow Ax_0 - b$
$d_0 \leftarrow -g_0$
$k \leftarrow 0$
while $\|g_k\| > \varepsilon\|b\|$ **do**
  $\alpha_k \leftarrow \frac{\|g_k\|^2}{\langle Ad_k, d_k\rangle}$
  $x_{k+1} \leftarrow x_k + \alpha_k d_k$
  $g_{k+1} \leftarrow g_k + \alpha_k A d_k$
  $\beta_k \leftarrow \frac{\|g_{k+1}\|^2}{\|g_k\|^2}$
  $d_{k+1} \leftarrow -g_{k+1} + \beta_k d_k$
  $k \leftarrow k+1$
end while

Метод Флетчера–Ривса:

$g_0 \leftarrow \nabla f(x_0)$
$d_0 \leftarrow -g_0$
$k \leftarrow 0$
while $\|g_k\| > \varepsilon\|\nabla f(x_0)\|$ **do**
  $\alpha_k \leftarrow$ {линейный поиск}
  $x_{k+1} \leftarrow x_k + \alpha_k d_k$
  $g_{k+1} \leftarrow \nabla f(x_{k+1})$
  $\beta_k \leftarrow \frac{\|g_{k+1}\|^2}{\|g_k\|^2}$
  $d_{k+1} \leftarrow -g_{k+1} + \beta_k d_k$
  $k \leftarrow k+1$
end while

Другие схемы нелинейного метода CG

Существует много версий нелинейного CG. Все они отличаются лишь выбором коэффициента β_k:

Полак–Рибье: $$\beta_k^{\mathrm{PR}} := \frac{\langle g_{k+1}, y_k\rangle}{\|g_k\|^2}$$

Хестинс–Штифель: $$\beta_k^{\mathrm{HS}} := \frac{\langle g_{k+1}, y_k\rangle}{\langle d_k, y_k\rangle}$$

Полак–Рибье+: $$\beta_k^{\mathrm{PR+}} := \max\{0, \beta_k^{\mathrm{PR}}\}$$

Гильберт–Ноусидаль: $$\beta_k^{\mathrm{GN}} := \max\{-\beta_k^{\mathrm{FR}}, \min\{\beta_k^{\mathrm{PR}}, \beta_k^{\mathrm{FR}}\}\}$$

Дай–Юань: $$\beta_k^{\mathrm{DY}} := \frac{\|g_{k+1}\|^2}{\langle d_k, y_k\rangle}$$

Агер–Джан: $$\beta_k^{\mathrm{HZ}} := \frac{\langle g_{k+1}, y_k - \frac{2\|y_k\|^2}{\langle d_k, y_k\rangle} d_k\rangle}{\langle d_k, y_k\rangle}$$

Во всех формулах $y_k := g_{k+1} - g_k$.

- Все эти схемы совпадают на строго выпуклой квадратичной функции.
- На неквадратичной функции их поведение различается.