Лекція 14. Метод спряжених градієнтів

Для того щоб зробити "нульову версію" методу спряжених градієнтів ефективною процедурою для розв‘язання розріджених систем, потрібен ефективний метод обчислення p_k . Це можливо зробити, вивчивши властивості p_k як розв‘язку певної задачі найменших квадратів.

Лема 1. *Для* $k \geq 2$ *вектори* p_k, *що отримані за алгоритмом "нульової версії" методу спряжених градієнтів, задовольняють співвідношення*

$$p_k = r_{k-1} - AP_{k-1}z_{k-1},$$

де z_{k-1} *є розв‘язком задачі*

$$\min_z \left\| r_{k-1} - AP_{k-1}z \right\|_2 .$$

Без доведення.

Використовуючи цей результат, ми можемо установити кілька важливих співвідношень між відхилами r_i і напрямками спуску p_i .

Теорема 1. *Після* k *ітерацій методу спряжених напрямків маємо*

$$r_j = r_{j-1} - \alpha_j Ap_j, \tag{1}$$

$$P_j^T r_j = 0, \tag{2}$$

$$span\left\{p_1, \dots, p_j\right\} = span\left\{r_0, \dots, r_{j-1}\right\} = span\left\{b, Ab, \dots, A^{j-1}b\right\}. \tag{3}$$

Доведення.

1) Подіємо матрицею A на обидві частини рівності

$$x_j = x_{j-1} + \alpha_j p_j \Rightarrow Ax_j = Ax_{j-1} + \alpha_j Ap_j \Rightarrow b - Ax_j = b - Ax_{j-1} - \alpha_j Ap_j$$
$$\Rightarrow r_j = r_{j-1} - \alpha_j Ap_j .$$

2) Для того щоб установити (2) зауважимо, що якщо y_j – розв‘язок системи

$$\left(P_j^T AP_j\right) y_j = P_j^T b,$$

то y_j мінімізує квадоатичний функціонал

$$\varphi\left(P_j y\right) = \frac{1}{2}\left(P_j^T AP_j y, y\right) - \left(P_j^T b, y\right).$$

Отже, подіявши на рівняння

$$P_j^T AP_j y = P_j^T b$$

матрицей P_j^{-T}, отримуємо

$$AP_j y = b .$$

Таким чином, $P_j y_j = x_j$. Тоді

$$P_j^T r_j = P_j^T\left(b - Ax_j\right) = P_j^T\left(b - AP_j y_j\right) = 0 .$$

3) Рівняння (3) доводиться за індукцією. При $j = 1$ воно очевидне. Припустимо, що воно виконується для деякого j, що задовольняє нерівність $1 \leq j \leq k$. Внаслідок (1) маємо

$$r_j \in span\left\{b, Ab, \dots, A^j b\right\}.$$

Отже,

$$p_{j+1} = r_j - AP_j z_j \in span\{b, Ab, \ldots, A^j b\}.$$

Обидва простори $span\{r_0, \ldots, r_j\}$ і $span\{p_1, \ldots, p_{j+1}\}$ мають розмірність $j+1$, внаслідок чого вони повинні співпадати із $span\{b, Ab, \ldots, A^j b\}$.

Теорема 2. *Після $k-1$ кроків алгоритма спряжених градієнтів за "нульовою версією" відхили $r_0, \ldots, r_{k-1}$ взаємно ортогональні.*

Доіведення. Із рівняння $x_i = x_{i-1} + \alpha_i p_i$ випливає, що $r_i = r_{i-1} - \alpha_i A p_i$. Таким чином, за лемою 1 маємо для $j = 1, \ldots, k$

$$p_j = r_{j-1} - \left[Ap_1, \ldots, Ap_{j-1}\right] z_{j-1} \in span\{r_0, \ldots, r_{j-1}\}.$$

Отже, $\left[p_1, \ldots, p_k\right] = \left[r_0, \ldots, r_{k-1}\right] T$ для деякої верхньої трикутної матриці $T \in R^{k \times k}$. Оскільки вектори p_i незалежні, матриця T невироджена і тому

$$\left[r_0, \ldots, r_{k-1}\right] = \left[p_1, \ldots, p_k\right] T^{-1}.$$

З огляду на це, $r_{j-1} \in span\{p_1, \ldots, p_j\}$ для $j = 1, \ldots, k$. Проте в силу (2)

$$P_{k-1}^T r_{k-1} = 0.$$

Оскільки $r_0, \ldots, r_{k-2} \in span\{p_1, \ldots, p_{k-1}\}$, маємо

$$\left(r_j, r_{k-1}\right) = 0 \text{ для } j = 0,1, \ldots k-2.$$

Тепер покажемо, яким чином p_k виражається у вигляді лінійної комбінації p_{k-1} і r_{k-1}. Подамо вектор z_{k-1} із леми 1 в блочному вигляді

$$z_{k-1} = \begin{bmatrix} w \\ \mu \end{bmatrix} \begin{matrix} {\scriptstyle k-2} \\ {\scriptstyle 1} \end{matrix}.$$

Скористаємося тотожністю

$$r_{k-1} = r_{k-2} - \alpha_{k-1} A p_{k-1}.$$

Мибачимо, що

$$p_k = r_{k-1} - AP_{k-1} z_{k-1} = r_{k-1} - AP_{k-2} w - \mu A p_k = \left(1 + \frac{\mu}{\alpha_k}\right) r_{k-1} + s_{k-1},$$

де

$$s_{k-1} = -\frac{\mu}{\alpha_{k-1}} r_{k-2} - AP_{k-2} w.$$

Внаслідок взаємної ортогональності відхилів r_i вектори s_{k-1} і r_{k-1} ортогональні. Таким чином, задача найменших квадратів із леми 1 зводиться до вибору μ і w таких, що величина

$$\left\| p_k \right\|_2^2 = \left(1 + \frac{\mu}{\alpha_k}\right)^2 \left\| r_{k-1} \right\|_2^2 + \left\| s_{k-1} \right\|_2^2$$

мінімальна. Оскільки 2-норма для $r_{k-2} - AP_{k-2}z$ має мінімум в точці z_{k-2}, що дає відхил p_{k-1},звідси випливає, що s_{k-1} є кратним вектора p_{k-1}. Отже, $p_k \in span\{r_{k-1}, p_{k-1}\}$. Не обмежуючи загальності, ми можемо вважати, що

$$p_k = r_{k-1} + \beta_k p_{k-1}.$$

Оскільки $(Ap_k, p_{k-1}) = 0$ і $(r_{k-1}, p_{k-1}) = 0$, звідси випливає, що

$$\beta_k = \frac{(Ar_{k-1}, p_{k-1})}{(Ap_{k-1}, p_{k-1})},\ \alpha_k = \frac{(r_{k-1}, r_{k-1})}{(Ap_k, p_k)}.$$

Це приводить до "версії 1" методу спряжених градієнтів:

$k = 0;\ x_0 = 0;\ r_0 = b;$

while $r_k \neq 0$

  $k = k + 1;$

  if $k = 1$

    $p_1 = r_0;$

  else

    $\beta_k = -\dfrac{(Ar_{k-1}, p_{k-1})}{(Ap_{k-1}, p_{k-1})};$

    $p_k = r_{k-1} + \beta_k p_{k-1};$

  endif

  $\alpha_k = \dfrac{(r_{k-1}, r_{k-1})}{(Ap_k, p_k)};$

  $x_k = x_{k-1} + \alpha_k p_k;$

  $r_k = b - Ax_k;$

  end

$x = x_k$

Як здається, ці ітерації - в тому вигляді, як вони записані - потребують трьох різних множень матриці на вектор на кожний прохід цикла. Проте, якщо обчислювати відхили за допомогою рекурсії

$$r_k = r_{k-1} - \alpha_k Ap_k$$

і підставити

$$(r_{k-1}, r_{k-1}) = -\alpha_k (Ap_{k-1}, r_{k-1}),$$

$$(r_{k-2}, r_{k-2}) = \alpha_{k-1} (Ap_{k-1}, p_{k-1})$$

в формулу для β_k, то у нас вийде більш ефективна реалізація методу спряжених градієнтів, яка і була запропонована Хестенсем і Штіфелем

$$
\begin{array}{l}
k = 0;\;\; x_0 = 0;\;\; r_0 = b; \\
\textit{while}\;\; r_k \neq 0 \\
\qquad k = k + 1; \\
\qquad \textit{if}\;\; k = 1 \\
\qquad\qquad p_1 = r_0\,; \\
\qquad \textit{else} \\
\qquad\qquad \beta_k = -\dfrac{(r_{k-1}, r_{k-1})}{(r_{k-2}, r_{k-2})}\,; \\
\qquad\qquad p_k = r_{k-1} + \beta_k\, p_{k-1}\,; \\
\qquad \textit{endif} \\
\qquad \alpha_k = \dfrac{(r_{k-1}, r_{k-1})}{(Ap_k\,, p_k)}\,; \\
\qquad x_k = x_{k-1} + \alpha_k\, p_k\,; \\
\qquad r_k = b - Ax_k\,; \\
\qquad \textit{end} \\
x = x_k
\end{array}
$$