Лекція 13. Довільні напрямки спуску

Для того щоб уникнути рискання при найшвидшому спуску, розглянемо послідовну мінімізацію φ вздовж деякої множини напрямків $\{p_1, p_2, \ldots\}$, які не зобов‘язані відповідати відхилам $\{r_1, r_2, \ldots\}$. Як уже показано, мінімум $\varphi(x_{k-1} + \alpha p_k)$ по α відповідає значенню

$$\alpha = \alpha_k = \frac{(r_{k-1}, p_k)}{(Ap_k, p_k)}.$$

Лема 1. При такому виборі отримуємо

$$\varphi(x_{k-1} + \alpha_k p_k) = \varphi(x_{k-1}) - \frac{1}{2}\alpha_k^2 (Ap_k, p_k).$$

Доведення.

1) $\varphi(x_{k-1} + \alpha_k p_k) = \frac{1}{2}\left(A(x_{k-1} + \alpha_k p_k), x_{k-1} + \alpha_k p_k\right) - (b, x_{k-1} + \alpha_k p_k) =$

$= \frac{1}{2}(Ax_{k-1} + \alpha_k Ap_k, x_{k-1} + \alpha_k p_k) - (b, x_{k-1}) - \alpha_k (b, p_k) =$

$= \frac{1}{2}(Ax_{k-1}, x_{k-1}) + \frac{1}{2}\alpha_k (Ax_{k-1}, p_k) + \frac{1}{2}\alpha_k (Ap_k, x_{k-1}) +$

$+ \frac{1}{2}\alpha_k^2 (Ap_k, p_k) - (b, x_{k-1}) - \alpha_k (b, p_k) =$

$= \varphi(x_{k-1}) + \frac{1}{2}\alpha_k (Ax_{k-1}, p_k) + \frac{1}{2}\alpha_k (Ap_k, x_{k-1}) + \frac{1}{2}\alpha_k^2 (Ap_k, p_k) - \alpha_k (b, p_k)$;

2) $\alpha_k (b, p_k) = \frac{(r_{k-1}, p_k)}{(Ap_k, p_k)}(r_{k-1} + Ax_{k-1}, p_k) = \frac{(r_{k-1}, p_k)}{(Ap_k, p_k)}\left((r_{k-1}, p_k) + (Ax_{k-1}, p_k)\right) =$

$= \frac{(r_{k-1}, p_k)^2}{(Ap_k, p_k)} + \frac{(r_{k-1}, p_k)}{(Ap_k, p_k)}(Ax_{k-1}, p_k) = \alpha_k^2 (Ap_k, p_k) + \alpha_k (Ax_{k-1}, p_k)$;

3) $\varphi(x_{k-1} + \alpha_k p_k) = \varphi(x_{k-1}) + \frac{1}{2}\alpha_k (Ax_{k-1}, p_k) + \frac{1}{2}\alpha_k (Ap_k, x_{k-1}) +$

$+ \frac{1}{2}\alpha_k^2 (Ap_k, p_k) - \alpha_k^2 (Ap_k, p_k) - \alpha_k (Ax_{k-1}, p_k) =$

$= \varphi(x_{k-1}) - \frac{1}{2}\alpha_k^2 (Ap_k, p_k) + \frac{1}{2}\alpha_k (Ap_k, x_{k-1}) - \frac{1}{2}\alpha_k (Ax_{k-1}, p_k)$;

4) Оскільки матриця A - симетрична, то $(Ax, y) = (x, Ay)$. Отже,

$$\frac{1}{2}\alpha_k (Ap_k, x_{k-1}) - \frac{1}{2}\alpha_k (Ax_{k-1}, p_k) = 0.$$

5) Таким чином,

$$\varphi(x_{k-1} + \alpha_k p_k) = \varphi(x_{k-1}) - \frac{1}{2}\alpha_k^2 (Ap_k, p_k).$$

Отже, для того щоб зменшити φ, ми повинні вимагати, щоб p_k не був ортогональним до r_{k-1}. Це приводить до такої стратегії мінімізації:

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

while $r_k \neq 0$

$k = k + 1;$

Обрати напрямок p_k так, щоб $\left(r_{k-1}, p_k\right) \neq 0$

$$\alpha_k = \frac{\left(r_{k-1}, p_k\right)}{\left(Ap_k, p_k\right)}; \tag{3}$$

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

$r_k = b - Ax_k;$

end

Зауважимо, що $x_k \in span\left\{p_1, p_2, \ldots, p_k\right\} = \left\{\sum_{i=1}^{k} \beta_i p_i,\ \beta_i \in R\right\}$. Проблема полягає в тому, як обрати ці вектори з тим, щоб гарантувати збіжність і в той же час обійти пастки найшвидшого спуску

А-спряжені напрямки спуску

Певно, ідеальний підхід був би пов‘язаний із вибором лінійно незалежних p_i з тим, щоб кожне x_i було розв‘язком задачі

$$\min_{x \in span\left\{p_1, p_2, \ldots, p_k\right\}} \varphi(x). \tag{2}$$

Це гарантувало б не тільки збіжність, але й скінченність процесу, оскільки повинно бути

$$Ax_n = b.$$

Зауважимо, що то, що нам потрібно, це такий вектор p_k, для якого розв‘язок одновимірної задачі мінімізації

$$\min_{\alpha} \varphi\left(x_{k-1} + \alpha p_k\right)$$

одночасно дає розв‘язок для k-вимірної задачі мінімізації (2). Це – серйозне обмеження, але, виявляється, цього можна домогтися.

Лема 2. Нехай $P_k = \left[p_1, p_2, \ldots, p_k\right] \in R^{n \times k}$ є матрицею напрямків спуску. Якщо $x \in range\, P_k = \left\{x \in R^{n \times k} :\ x = P_k y\ \ \textit{для деякого}\ y \in R^{n \times k}\right\}$ – області значень матриці P_k, то $x = P_{k-1} y + \alpha p_k$ для деяких $y \in R^{k-1}$ і $\alpha \in R$. Для вектора x такого виду має місце співвідношення:

$$\varphi(x) = \varphi\left(P_{k-1} y\right) + \alpha\left(Ap_k, P_{k-1} y\right) + \frac{\alpha^2}{2}\left(Ap_k, p_k\right) - \alpha\left(b, p_k\right).$$

Доведення.

$$\varphi\left(P_{k-1} y + \alpha p_k\right) = \frac{1}{2}\left(A\left(P_{k-1} y + \alpha p_k\right), P_{k-1} y + \alpha p_k\right) - \left(b, P_{k-1} y + \alpha p_k\right) =$$

$$= \frac{1}{2}\left(AP_{k-1} y + \alpha Ap_k, y + \alpha p_k\right) - \left(b, y + \alpha p_k\right) =$$

$$= \frac{1}{2}\left(AP_{k-1}y, P_{k-1}y\right) + \frac{\alpha}{2}\left(AP_{k-1}y, p_k\right) + \frac{\alpha}{2}\left(AP_k y, P_{k-1}y\right) +$$

$$+ \frac{\alpha^2}{2}\left(Ap_k, p_k\right) - \left(b, P_{k-1}y\right) - \alpha\left(b, p_k\right) =$$

$$= \varphi\left(P_{k-1}y\right) + \alpha\left(Ap_k, P_{k-1}y\right) + \frac{\alpha^2}{2}\left(Ap_k, p_k\right).$$

Наявність "змішаного" члена $\alpha\left(Ap_k, P_{k-1}y\right)$ ускладнює мінімізацію. Без нього задача мінімізації φ по $range\, P_k$ розщепилася би на задачу мінімізації по $range\, P_{k-1}$, розв‘язок якої x_{k-1} вважається відомим, і просту мінімізацію для скаляра α.

Один із способів добитися цього розщеплення – вимагати, щоб напрямок p_k був А-спряженим відносно до $p_1, p_2, \ldots, p_{k-1}$, тобто $\left(Ap_k, p_{k-1}\right) = 0$. Якщо це зробити і визначити $x_{k-1} \in range\, P_{k-1}$ і $\alpha_k \in R$ співвідношеннями

$$\varphi\left(x_{k-1}\right) = \min_y \varphi\left(P_{k-1}y\right),$$

$$\alpha_k = \frac{\left(b, p_k\right)}{\left(Ap_k, p_k\right)},$$

то задача

$$\min_{y,\alpha} \varphi\left(P_{k-1}y + \alpha_k p_k\right) = \min_{y,\alpha} \varphi\left(P_{k-1}y\right) + \min_{\alpha}\left\{\frac{\alpha^2}{2}\left(Ap_k, p_k\right) - \alpha\left(b, p_k\right)\right\}$$

буде розв‘язана, якщо взяти $P_{k-1}y = x_{k-1}$ і $\alpha = \alpha_k$. Оскільки $x_{k-1} \in span\left\{p_1, p_2, \ldots, p_{k-1}\right\}$, а p_k є А-спряженим до $p_1, p_2, \ldots, p_{k-1}$, то $\left(Ax_{k-1}, p_k\right) = 0$, і, отже,

$$\alpha_k = \frac{\left(r_{k-1}, p_k\right)}{\left(Ap_k, p_k\right)},$$

тобто рецепт для α_k точно такий же, як і в методі довільних напрямків спуску. Таким чином, отримуємо

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

while $r_k \neq 0$

$k = k + 1;$

обрати $p_k \in span\left\{Ap_1, Ap_2, \ldots, Ap_{k-1}\right\}^{\perp}$ так, щоби $\left(r_{k-1}, p_k\right) \neq 0$.

$$\alpha_k = \frac{\left(r_{k-1}, p_k\right)}{\left(Ar_{k-1}, r_{k-1}\right)};$$

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

$r_k = b - Ax_k;$

end

Тут вектор $x_k = x_{k-1} + \alpha_k p_k$ мінімізує φ на лінійній оболонці напрямків спуску, тобто на підпросторі $span\left\{p_1, p_2, \ldots, p_k\right\}$.

Метод спряжених градієнтів

Для того щоб задовольнити (3), ми повинні бути впевнені, що можливий такий вибір p_k, при якому $\left(Ap_k, p_{k-1} = 0\right)$ і $\left(r_{k-1}, p_k\right) \neq 0$. Припустимо, що вектори $p_1, p_2, \ldots, p_{k-1}$ ненульові і задовольняють умови $\left(Ap_i, p_j\right) = 0, i \neq j$. Оскільки $x_{k-1} \in span\left\{p_1, p_2, \ldots, p_{k-1}\right\}$, звідси випливає, що для будь-якого $p \in R^n$ ми маємо $\left(r_{k-1}, p\right) = \left(b, p\right) - \left(Ap_{k-1}z, p\right)$ для деякого $z \in R^{k-1}$. Якщо $\left(r_{k-1}, p\right) = 0$ для всіх p, які є A-спряженими з $p_1, p_2, \ldots, p_{k-1}$, то $b \in span\left\{Ap_1, Ap_2, \ldots, Ap_{k-1}\right\}$, а з цього випливає, що $x = A^{-1}b \in span\left\{p_1, p_2, \ldots, p_{k-1}\right\}$ і, значить, $x_{k-1} = x$, тобто $r_{k-1} = 0$. Отже, якщо $r_{k-1} \neq 0$, то можна знайти ненульовий вектор p_k, який є A-спряженим з $p_1, \ldots, p_{k-1}$ і задовольняє умову $\left(r_{k-1}, p_k\right) \neq 0$.

Оскільки наша мета – здійснити швидке зменшення величини відхилів, природно обрати як p_k вектор, який ближче за всіх до r_{k-1} серед векторів, які є A-спряженими до $p_1, p_2, \ldots, p_{k-1}$. Таким чином, ми отримуємо "нульову" версію 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
        нехай p_k мінімізує $\left\| p - r_{k-1} \right\|_2$
        серед усіх векторів $p_k \in span\left\{Ap_1, Ap_2, \ldots, Ap_{k-1}\right\}^{\perp}$
    endif
    $$\alpha_k = \frac{\left(r_{k-1}, p_k\right)}{\left(Ar_{k-1}, r_{k-1}\right)};$$
    $x_k = x_{k-1} + \alpha_k p_k;$
    $r_k = b - Ax_k;$
    end
$x = x_k$

Скінченність цього алгоритму забезпечується тим, що вектори p_k ненульові і, будучи ненульовими A-спряженими векторами, є лінійно незалежними. Більше того, або $r_{k-1} = 0$ для деякого $k \leq n$, або ми обчислимо x_n, що мінімізує φ на $span\left\{p_1, p_2, \ldots, p_n\right\} = R^n$.