# ПРО ОДНЕ СІМЕЙСТВО СУБГРАДІЄНТНИХ АЛГОРИТМІВ З ПЕРЕТВОРЕННЯМ ПРОСТОРУ

Микола Журбенко, Олексій Лиховид

Інститут кібернетики ім. В.М. Глушкова НАН України,
Відділ методів негладкої оптимізації
zhurbnick@gmail.com, o.lykhovyd@gmail.com

Питання оптимізації обчислень (ПОО-XLVIII) (ISCOPT 2023)

Львів, Вересень 19-22, 2023

# Шор Наум Зуселевич

**Академік Наум Зуселевич Шор (1937 - 2006)**

# План

- Вступ
- Обчислювальна схема r-алгоритму
- Обчислювальна схема $r(\sigma)$-алгоритмів
- Чисельна ефективність $r_n$-алгоритму
- Висновок
- Література

# ВСТУП

Понад 50 років тому було розроблено субградієнтний алгоритм мінімізації з розтягом простору у напрямку різниці двох послідовних градієнтів – r-алгоритм [1], [2].

Практика використання r-алгоритму показує, що і досі він є одним із найефективніших алгоритмів негладкої оптимізації.

Проте теоретичне дослідження ефективності алгоритму не закінчено. Основна проблема теоретичного обґрунтування r-алгоритму полягає у узгодженому виборі значень коефіцієнтів розтягу простору та крокових множників.

З метою подолання певною мірою цієї проблеми було розроблено сімейство модифікацій r-алгоритму з програмним вибором значень коефіцієнтів розтягу простору [3], [4].

У доповіді буде надано коротку характеристику $r(\sigma)$-алгоритмів.

# субградієнтні алгоритми з перетворенням простору

$minf(x), x \in R^n$.
$\partial f(x)$ – множина субгадєнтів.
$A$ – невироджений лінійний оператор: $y = Ax$.
$\varphi(y) = f(A^{-1}y)$ – функція $f(x)$ у просторі $Y = AX$.
$\partial\varphi(y) = A^{*-1}\partial f(x)$, – субградієнт в просторі $Y$.
**Ітерація** $k$ : точка $x_k$; оператор $A_k : Y_k = A_k X$.
**Ітерація** $k+1$ : $y_{k+1} = y_k - h_k g_\varphi(y_k)/||g_\varphi(y_k)||$.
У вихідному просторі $X$:
$$x_{k+1} = x_k - h_k B_k B_k^* g_f(x_k)/||B_k^* g_f(x_k)||, \qquad (*)$$
де $B_k = A_k^{-1}$.
$T_{k+1}$ – оператор перетворення простору $A_{k+1} = T_{k+1}A_k$
Обернене перетворення $B_{k+1} = B_k T_k^{-1}$.
Ітеративна процедура (*) породжує конкретні алгоритми при вказівці послідовностей $h_k, T_k$, оператора $B_0$ та початкової точки $x_0$.

# Обчислювальна схема r-алгоритму

$T_k = R_\alpha(\eta) = (\alpha - 1)\eta\eta^T + I$,
де $I$ – одинична матриця, $\eta \in R^n$, $\alpha$ – напрямок та коефіцієнт розтягу простору, $||\eta|| = 1$, $\alpha \geq 1$. Напрямок розтягу $\eta_{k+1} = (g^*_{k+1} - g^*_k)/||g^*_{k+1} - g^*_k||$, де $g^*_{k+1}, g^*_k$ – субградієнти на ітераціях $k+1$, $k$ у перетвореному просторі $Y_k$.
Значення коефіцієнтів розтягу простору $\alpha_k$ (параметр r-алгоритму) вибираються однаковими на всіх ітераціях: $\alpha_k = \alpha \geq 1$. На практиці рекомендується вибирати це значення порядку 2.
Розмір крокового множника $h_k$ визначається процедурою мінімізації за напрямком $p_k = -B_k g^*_k/||g^*_k||$. Основною вимогою при цьому є виконання умови: $(p_k, g(x_{k+1})) < 0$ (ця умова забезпечує, що $||g^*_{k+1} - g^*_k|| > 0$).

# Обчислювальна схема $r(\sigma)$-алгоритмів

Обчислювальна схема $r(\sigma)$-алгоритмів переважно відповідає r-алгоритму.

$R_\alpha(\eta) = (\alpha - 1)\eta\eta^T + I \quad \Rightarrow \quad \tilde{R}_\sigma(\tilde{\eta}) = \gamma_\sigma \sigma \tilde{\eta}\tilde{\eta}^T + I,$

де $\tilde{\eta} \in R^n$; $\sigma(g^*_{k+1}, g^*_k) \in R^1, \sigma > 0$; $\gamma_\sigma \in R^1$, $\gamma_\sigma > 0$.

Число $\gamma_\sigma \Rightarrow \alpha_k \leq \alpha_{\max}$ – параметр алгоритму.

На відміну від оператора $R_\alpha(\eta)$, вектор $\tilde{\eta}$ не нормовано, тобто виконання умови $||\tilde{\eta}|| = 1$ не вимагається.

$\tilde{\eta} = 0, \Rightarrow \alpha = 1, \ \tilde{R}_\sigma(\tilde{\eta}) = I.$

$\tilde{\eta} \neq 0, \Rightarrow \tilde{R}_\sigma(\tilde{\eta}) = R_\alpha(\tilde{\eta}/||\tilde{\eta}||)$, де $\alpha = 1 + \gamma_\sigma \sigma ||\tilde{\eta}||^2$.

Природня вимога на функцію $\sigma(g_1, g_2)$

$\sigma(\mu g_1, \mu g_2) = \sigma(g_1, g_2)/\mu^2$, де $\mu \in R^1, \mu > 0$. Ця умова забезпечує незалежність роботи алгоритму від множника на цільову функцію. Різні варіанти алгоритму будуть визначатися вибором множника $\sigma$ та значенням параметра $\alpha_{\max}$.

# Обчислювальна схема $r(\sigma)$-алгоритмів

Приклади нормуючих множників.
1. Алгоритм $r(\sigma_1) : \sigma_1(g_1, g_2) = 1/(||g_1||^2 + ||g_2||^2)$;
$\gamma_\sigma = (\alpha_{\max} - 1)/3$.
2. Алгоритм $r(\sigma_2) : \sigma_2(g_1, g_2) = 1/\max\{||g_1||^2, ||g_2||^2\}$;
$\gamma_\sigma = (\alpha_{\max} - 1)/4$.
Особливий інтерес становить алгоритм з розтягом простору за напрямком різниці нормованих субградієнтів – $r_n$-алгоритм. У $r_n$-алгоритмі вектор $\tilde{\eta}_{k+1} = g^*_{k+1}/||g^*_{k+1}|| - g^*_k/||g^*_k||$. Для цього алгоритму: $\sigma = 1$; $\gamma_\sigma = (\alpha_{\max} - 1)/4$.
$\alpha_{k+1} = 1 + (\alpha_{\max} - 1)(1 - \cos(\varphi_{k+1}))/2$, де $\varphi_{k+1}$ – кут між векторами $g^*_{k+1}, g^*_k$. Таким чином, значення коефіцієнта розтягу на ітерації $r_n$-алгоритму має тим більше значення, чим більший кут між векторами $g^*_{k+1}, g^*_k$. Максимальне значення коефіцієнта розтягу ($\alpha_{\max}$) досягається при куті, що дорівнює $\pi$.

# Обчислювальна схема $r(\sigma)$-алгоритмів

$h_k \quad \Rightarrow \quad$ використання процедури одномірного спуску (r-алгоритм). Однак, суттєво, що $r(\sigma)$-алгоритм можна використовувати з постійним кроком у перетвореному просторі. Обчислювальна схема таких варіантів $r(\sigma)$-алгоритмів значно простіша за схему r-алгоритму.

Для теоретичного обґрунтування збіжності $r(\sigma)$-алгоритмів пропонується наступний їх варіант: максимальне значення коефіцієнта розтягу простору змінюється на ітераціях, тобто визначається послідовність

$\alpha_{\max}(k) \geq 1$; $\alpha_{\max}(k) \to 1$ ; $\prod\limits_{j=1}^{k} \alpha_{\max}(j) \to \infty$ (загальний розтяг простору не обмежено).

Приклад: $\alpha_{\max}(k) = e^{c/k};\ c > 0$. Така послідовність асоціюється з послідовністю вибору крокових множників у класичному субградієнтному алгоритмі (крок прагне до нуля, сума кроків розходиться).

# Чисельна ефективність $r_n$-алгоритму

Тестова задача (для демонстрації).

$$f(x) = \sum_{i=1}^{n} \rho_n^{i-1} |x_i|, \text{ де } \rho_n = 10^{6/(n-1)}.$$

Ступінь витягнутості ліній рівня («ярів») функцій не залежить від розмірності і дорівнює $10^6$.

Початкова точка $x_i = 1.0, i = 1, \ldots n$. Параметр точності рішення по функціоналу $\bar{\varepsilon} = 10^{-6}$.

Позначення стовпців таблиці:

**nVarbl** — кількість змінних;

**nIter** — кількість ітерацій;

**dAlphaAvrg** — середнє значення коефіцієнта розтягу;

**r-agorithm** — кількість ітерацій $r$-алгоритму ($\alpha = 2.0$).

# Чисельна ефективність $r_n$-алгоритму

ТАБЛИЦЯ. Мінімізація $f(x)$

| nVarbl | dAlphaAvrg | nIter | r-agorithm |
|---|---|---|---|
| 10 | 3.800 | 166 | 298 |
| 50 | 3.784 | 808 | 1521 |
| 100 | 3.821 | 1610 | 3054 |

( r-алгоритм nVarbl = 100 и $\alpha = 4.0 \Rightarrow$ nIter = 1539)

# Висновок

$r(\sigma)$-алгоритми є модифікаціями r-алгоритму. Обчислювальна схема $r(\sigma)$-алгоритмів з постійним кроком суттєво простіше схеми r-алгоритму.

Величини коефіцієнтів розтягу простору на ітераціях $r(\sigma)$-алгоритмів не постійні, вони обчислюються у процесі його роботи.

Алгоритми можуть використовуватися з постійним кроковим множником у перетвореному просторі.

Чисельні експерименти показали досить високу ефективність $r(\sigma)$-алгоритмів. Їхня ефективність не поступається ефективності r-алгоритму.

Результати дослідження чисельної ефективності показують, що найбільш ефективним варіантом із сімейства $r(\sigma)$-алгоритмів є $r_n$-алгоритм.

# Література

[1] Шор Н.З., Журбенко Н.Г. Метод минимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов. Кибернетика. 1971. №3. С. 51–59.
[2] Shor N.Z. Minimization methods for non-differentiable functions. Berlin: Springer-Verlag, 1985. 178 p.
[3] Журбенко Н.Г. Численная эффективность одной модификации r-алгоритма. Теорія оптимальних рішень. Київ: Інститут кібернетики ім. В.М. Глушкова НАН України. 2017. С. 33–38.
[4] Журбенко Н.Г., Лиховид А.П. К численной эффективности одной модификации r-алгоритма. Комп'ютерна математика. К.: Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2019. № 1. С. 2–10.

# ACKNOWLEDGEMENTS

The work was supported by a grant from the Volkswagen Foundation (grant 97775).

# Thank you for attention!

zhurbnick@gmail.com