ПРО ОДНУ ПРОЦЕДУРУ ГЕНЕРАЦІЇ АГРЕГАТНИХ ε-СУБГРАДІЄНТІВ

М.Г. Журбенко

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

XXVII Міжнародна конференція «Автоматизація 2024»

Дніпро, листопад 20-22, 2024

Шор Наум Зуселевич

Академік Наум Зуселевич Шор (1937 - 2006)

ПЛАН

- ВСТУП
- ЗАДАЧА ε-ОПТИМІЗАЦІЇ
- ε-SUBUBGRADIENT
- АГРЕГАТНИЙ ε-СУБГРАДІЄНТ
- ε-СУБГАДІЄНТНИЙ АЛГОРИТМ
- МОДИФИКАЦИЯ ε-СУБГАДІЄНТНОГО АЛГОРИТМУ
- ВИСНОВОК
- ACKNOWLEDGEMENTS
- REFERENCES

ВСТУП

1962 р. Алгоритм узагальненого градієнтного узвозу. Перший приклад чисельного алгоритму негладкої оптимізації.
1969 р. Субградієнтні методи з розтягом простору в напрямку субградієнта (1976 Метод еліпсоїдів).
1971 р. – Субградієнтні методи з розтягом простору в напрямку різності двох послідовних субградієнтів (r-алгоритм).
До теперішнього часу розроблено низку модифікацій rалгоритму: алгоритми з програмним вибором значень коефіцієнтів розтягування ($r(\sigma)$-алгоритмы.); алгоритми з різними правилами вибору крокових множників та інші.
Розроблено варіанти алгоритмів негладкої оптимізації на основі використання ε-субградієнтів.
Основна мета при цьому полягала у розробці алгоритму з порівнянною з r-алгоритмом ефективністю та з теоретично обґрунтованою трудомісткістю. Ці алгоритми ґрунтуються на суттєвому використанні агрегатних ε-субградієнтів.

ЗАДАЧА ε-МІНІМІЗАЦІЇ

$f(x)$ – опукла функція в R^n.
$f^* = \min\{f(x)|x \in R^n\}$.
X^* – множина точок мінімуму функції $f(x)$.
Для даного числа $\bar{\varepsilon} \geq 0$ введемо
$X^*(\bar{\varepsilon})$ – $\bar{\varepsilon}$-оптимальну множену:
$X^*(\bar{\varepsilon}) = \{x \in R^n|f(x) \leq f^* + \bar{\varepsilon}\}$.
Будь-яка точка $\tilde{x} \in X^*(\bar{\varepsilon})$ і значення функції $\tilde{f} = f(\tilde{x})$ будуть називатися розв'язками задачі $\bar{\varepsilon}$-оптимізації ($\bar{\varepsilon}$-рішення).
Нехай задано кулю $D(z, R)$ початкової локалізації $\bar{\varepsilon}$-розв'язків, де $z \in R^n$ – її центр, а $R > 0$ – його радіус. Припустимо, що куля $D(z, R)$ містить принаймні одну точку мінімуму $x^* \in X^*$.

ε-СУБГРАДІЄНТ

Ми будемо використовувати визначення ε-субградієнта, дещо відмінне від класичного (C. Lemarechal, K. Mifflin).

Визначення. Вектор $g \in R^n$ називається $(\varepsilon, \tilde{f})-$субградієнтом функції $f(x)$ у точці z, якщо для $\forall x$ виконується нерівність:

$$f(x) \geq \tilde{f} + (g, x - z) - \varepsilon \tag{1}$$

де $\tilde{f} \geq f^*; \varepsilon \in R^1$.

Значення $\tilde{f}$ на ітерації k алгоритму зазвичай дорівнює отриманому рекордному значенню функції $f(x)$. Якщо значення f^* відоме апріорі, то $\tilde{f} = f^*$. Класичне визначення ε-субградієнта збігається з визначенням $(\varepsilon, \tilde{f})$-субградієнта (1) при $\tilde{f} = f(z)$, $\varepsilon \geq 0$,

Позначення. $\partial f(x)$ – множена субградієнтів функції $f(x)$ у точці x.

$G(x)$ – множена $(\varepsilon, \tilde{f})$-субградієнтів функції $f(x)$ у точці x.

ε-СУБГРАДІЄНТ

Нехай $g(\varepsilon_1, \tilde{f}) \in G(z_1)$ тоді $g(\varepsilon_2, \tilde{f}) \in G(z_2)$ де

$$\varepsilon_2 = \varepsilon_1 - (g, z_2 - z_1). \tag{2}$$

Таким чином, $(\varepsilon, \tilde{f})$–субградієнт у певній точці z_1 є $(\varepsilon, \tilde{f})-$субградієнтом у будь-якій іншій точці z_2. Формула (2) визначає перерахунок параметра ε-субградієнта.
Нехай $g(\varepsilon_1, \tilde{f}_1) \in G(z)$, тоді $g(\varepsilon_1, \tilde{f}_1) \in G(z)$, де

$$\varepsilon_2 = \varepsilon_1 - (\tilde{f}_1 - \tilde{f}_2). \tag{3}$$

У алгоритмі перерахунок параметра $(\varepsilon, \tilde{f})$ виконується у ситуації, коли отримано нове рекордне значення функції $(\tilde{f}_1 > \tilde{f}_2)$.
Наведене визначення ε-субградієнта є зручним способом відстеження локалізації ε-рішення на основі відсікаючих площин.

АГРЕГАТНИЙ ε-СУБГРАДІЄНТ

Позначення.

$$\tilde{X}(\bar{\varepsilon}, \tilde{f}) = \{x \in R^n | f(x) \leq \tilde{f} - \bar{\varepsilon}\}.$$

В алгоритмі всі точки, що не належать множені $\tilde{X}(\bar{\varepsilon}, \tilde{f})$ відсікаються! Зазвичай відсікаються точки, що не належать оптимальній множині X^*.
Якщо на ітерації k множена $\tilde{X}(\bar{\varepsilon}, \tilde{f}_k)$ порожня, то $\tilde{f}_k$ є вирішенням задачі $2\bar{\varepsilon}$-оптимізації!
Нехай $g_i \in G(z), i = 1, \ldots, m$; $\tilde{\varepsilon}_i = \bar{\varepsilon} - \varepsilon_i$.
Тоді $\tilde{X}(\bar{\varepsilon}, \tilde{f})$ міститься в перетині підпросторів:
$(x - z, g_i) \leq -\tilde{\varepsilon}_i$, (не виключено, що цей перетин порожній).
y– вектор зсуву відсікаючої площини від точки z визначається рішенням задачі:

$$\min ||y||^2 \tag{4}$$

$$(y, g_i) \leq -\tilde{\varepsilon}_i; i = 1, \ldots, m. \tag{5}$$

Якщо система (5) несумісна, то вектор y невизначений.

AGGREGATED ε-SUBGRADIENT

Твердження 1. Нехай система обмежень (5) несумісна. Тоді $\tilde{f}$ є розв'язком задачі $2\bar{\varepsilon}$-оптимізації та опукла оболонка множини ε-субградієнтів $G(z)$ містить 0.

Твердження 2. Припустімо:

a). $\max\{\varepsilon_i; i = 1, \ldots, m\} \leq \bar{\varepsilon}$;

b). система (5) узгоджена (рішення існує);

c). λ оптимальні значення двоїстих змінних задачі (4)-(5);

d). $g = \sum \lambda_i g_i / \sum \lambda_i$;

e). $\varepsilon = \sum \lambda_i \varepsilon_i / \sum \lambda_i$.

Тоді: $g \in G(z)$,

$$y = -hg/||g||, h = (\bar{\varepsilon} - \varepsilon)/||g||. \quad (6)$$

Якщо система (5) несумісна, то встановлюємо $g = 0$.

Вектор g, визначений у твердженні 2, буде називатися **агрегатним ε-субградієнтом** множени субградієнтів $G(z)$.

AGGREGATED ε-SUBGRADIENT

Агрегатний ε-субградієнт забезпечує максимальний зсув площини відсікання множени $\tilde{X}(\bar{\varepsilon}, \tilde{f})$.

Коментар.

1. Якщо всі $\varepsilon_i = \varepsilon$ – константа, тоді $g = Nr\{g_i\} \Rightarrow$ класичний агрегатний субградієнт.
(На оновленні цієї властивості розроблено оригінальний алгоритм проектування нуля на опуклу оболонку векторів: $g = Nr\{g_i\}$)
2. Нехай $g \in \partial f(z)$, тоді g є $(\varepsilon, \tilde{f})$ -субградієнтн с $\varepsilon = -(f(z) - \tilde{f}) \leq 0$.
3. Нехай $\tilde{f} = f^*$, тоді $g \in \partial f(z)$ є (ε, f^*) -субградієнтн з $\varepsilon = -(f(z) - f^*)$. Для $\bar{\varepsilon} = 0$ значення кроку "зрушення"$h = (\bar{\varepsilon} - \varepsilon)/||g|| = (f(z) - f^*)/||g||$, що точно відповідає множнику кроку методу Поляка (Б. Т. Поляк).

ε-СУБГАДІЄНТНИЙ АЛГОРИТМ

Нехай на ітерації k отримано: точка x_k; множена ε-субградієнтів Ω_k.

$k+1$-я **ітерація.**

Обчислюємо:

g_k – ε-субградієнт у точці x_k;

$\Omega_{k+1} = g_k \bigcup \Omega_k$;

G_k – агрегатний ε-субградієнт множена Ω_{k+1}

$x_{k+1} = x_k - h_k G_k/||G_k||$;

Корекція множини Ω_{k+1} для точки x_{k+1} з урахуванням отримання нового рекордного значення функції та заданого параметра m (максимальне число ε-субградієнтів множини Ω_{k+1}).

Зупинити або перейти до ітерації $k+2$.

МОДИФИКАЦИЯ ε-СУБГАДІЄНТНОГО АЛГОРИТМУ

$k+1$-я **ітерація.**
Обчислюємо:
g_k – ε-субградієнт у точці x_k;
$y = x_k : \Omega = \Omega_k$;
M1:
$g(y)$ – ε-субградієнт у точці y;
$g(x)$ – ε-субградієнт, що відповідає $g(y)$ субградієнту для точки x_k;
$\Omega := g(x) \bigcup \Omega$;
G – агрегатний ε-субградієнт множени Ω
$y := x_k - h_k G/||G||$;
Якщо g_k є активним для множени Ω, **goto M1**.
$x_{k+1} = y$; $\Omega_{k+1} = \Omega$;
Зупинити або перейти до ітерації $k+2$.

ВИСНОВОК

Представлену стратегію генерації ε-субградієнтів можна використовувати для різних алгоритмів мінімізації. Зокрема, її можна використовувати для модифікації субградієнтного алгоритму Поляка з апріорі відомим оптимальним значенням функції. Для такого алгоритму (з використанням операторів перетворення простору) справедлива наступна оцінка числа ітерацій для вирішення задачі з відносною точністю γ :
Для задач високої точності ($n \ll 1/\gamma$):
$k < O\left(n(1/\gamma)\ln(1/\gamma)\right)$
(Алгоритм еліпсоїдів: $k < O\left(n^2 \ln(1/\gamma)\right)$.)
Для задач великої розмірності ($n \gg 1/\gamma$) :
$k < O\left((1/\gamma^2)\ln(1/\gamma)\right)$
(Агоритм Поляка: $k < O\left(1/\gamma^2\right)$.)

ВИСНОВОК

Процедура генерації ε-субградієнтів розроблена з орієнтуванням на розв'язання задач великої розмірності. Для алгоритму з $m = 2$ обчислення агрегатного субградієнта (вирішення задачі квадратичного програмування) виконується аналітично.
На цей час проводяться дослідження чисельної ефективності алгоритму.

ACKNOWLEDGEMENTS

Робота підтримана грантом Volkswagen Foundation
(грант № 97775).

REFERENCES

1. Н. З. Шор, "Методы минимизации недифференцируемых функций и их применение", Киев, Наук.думка, 1979, 200 с.
2. Б. Т. Поляк, "Минимизация негладких функционалов", Журнал выч. мат. и мат. физики, 1969, т.9, №3, с. 507–521 .
3. Н. Г. Журбенко, "Об алгоритме минимизации ε-субградиента", Теория оптимальных решений. Ин-т кибернетики им. В. М. Глушкова НАН Украины, 2002, с. 111–118.
4. Н.Г. Журбенко, "Алгоритм Поляка на основе агрегатных ε-субградиентов", Теория оптимальных решений. Киев, : Ин-т кибернетики им. В.М. Глушкова НАН Украины, 2015, с. 146–153.
5. Н.Г. Журбенко, 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.

ДЯКУЮ ЗА УВАГУ

zhurbnick@gmail.com