ОПЫТ ИСПОЛЬЗОВАНИЯ И МОДИФИКАЦИИ r-АЛГОРИТМА ШОРА

Н.Г. ЖУРБЕНКО

Институт кибернетики им. В.М. Глушкова НАН Украины, Киев, Украина
zhurbnick@gmail.com
25 октября 2022 г.

К 85-летию академика Наума Зуселевича Шора.

ПЛАН

- Шор Наум Зуселевич.
- Идея методов. Общая схема.
- r-алгоритм.
- r-алгоритм. Численная эффективность.
- r-алгоритм. Теория и практика.
- Модификации r-алгоритма.
- Модификации r-алгоритма. Регуляризация матрицы преобразования.
- Модификации r-алгоритма. $r(\sigma)$-алгоритмы.
- $\alpha(\varepsilon)$-алгоритм.
- Квазиньютоновские методы
- Алгоритм Поляка с преобразованием пространства

Шор Наум Зуселевич

Академик Наум Зуселевич Шор (1937 - 2006)

Шор Наум Зуселевич. Цитата.

«Для творчества необходимо накопить положительные эмоции.» (Шор Н.З.)

Шор Наум Зуселевич

• ОГС – Методы обобщённого (субградиентного) градиентного спуска . Негладкая оптимизация. 1962 г.
• ОГСРП – Субградиентные методы с растяжением пространства в направлении субградиента. 1969 г. Метод эллипсоидов;
• r-алгоритм – Растяжение пространства в направлении разности двух последовательных (суб)градиентов. 1971 г.

Шор Н. З., Журбенко Н. Г. Метод оптимизации, использующий операцию растяжения пространства в направлении разности двух последовательных градиентов // Кибернетика. 1971. № 3. С.51 - 59.

$$\min\{f(x)\,|\,x \in R^n\}$$

Идея методов. Общая схема.

$X_0 \underset{R_0^{-1}}{\overset{R_0}{\longleftrightarrow}} X_1 \underset{R_1^{-1}}{\overset{R_1}{\longleftrightarrow}} X_2 \dots$

$A_{k+1} = R_k A_k;\ \ B_{k+1} = A_{k+1}^{-1} = B_k R_k^{-1}$

$X_{k+1} = R_k X_k;$

$g_{k+1} \in \partial f(x_{k+1}),$

$g_{k+1}^* = B_k^* g_{k+1}$ – субградиент функции $\varphi(y) = f(A^{-1}y)$ (функция $f(x)$ в пространстве $Y = AX$.).

$$x_k = x_{k-1} - h_k B_k B_k^* g_{k-1} / ||B_k^* g_{k-1}||, k = 1, 2, \dots, . \qquad (1)$$

$$B_{k+1} = B_k R_k^{-1}.$$

Итеративная процедура (1) порождает конкретные алгоритмы при указании последовательностей h_k, R_k, оператора B_0 и начальной точки x_0.

Идея методов. Общая схема.

H– форма (Скоков В.А.)
$x_k = x_{k-1} - h_k B_k B_k^* g_{k-1} / ||B_k^* g_{k-1}||, k = 1, 2, \ldots, .$
$B_{k+1} = B_k R_k^{-1}.$
$H_k = B_k B_k^*$
$x_k = x_{k-1} - h_k H_k g_{k-1}, k = 1, 2, \ldots, .$
$H_{k+1} = H_k + \triangle H_k.$

r-алгоритм

В r-алгоритме в качестве оператора R_k используется оператор растяжения пространства
$R_\alpha(\eta) = (\alpha - 1)\eta\eta^T + I.$
$\eta \in R^n, \alpha -$ направление и коэффициент растяжения пространства, $|\eta| = 1, \alpha \geq 0.$
$R^*_\alpha(\eta) = R_\alpha(\eta)$ ($R_\alpha(\eta)$ – самосопряженный оператор);
$R^{-1}_\alpha(\eta) = R_\beta(\eta)$, где $\beta = 1/\alpha.$
Направление и коэффициент растяжения пространства:
$\eta_{k+1} = (g^*_{k+1} - g^*_k)/||g^*_{k+1} - g^*_k||$
α_{k+1} – const >1 (≈ 2)
$\beta_{k+1} = 1/\alpha_{k+1}$
$R_{\beta_{k+1}}(\eta_{k+1})$ – оператор преобразования матрицы B_k.
h_{k+1} – процедура одномерной минимизации по направлению.

r-алгоритм. Численная эффективность

Функция Розенброка:

$$f(x) = (1 - x_1)^2 + 100(x_2 - {x_1}^2)^2$$

Iter = 0 Funct = 2.42000000000000E+001
Iter = 20 Funct = 1.88227899270044E-003
Iter = 40 Funct = 1.51681969684129E-012
Iter = 41 Funct = 7.24286252084236E-012
x(1) = 0.999999696
x(2) = 0.999999125
114 – количество вычислений функции (градиента).
Поляк, Б. Т. Введение в оптимизацию. М.: Наука, 1983. 384с.

r-алгоритм. Теория и практика.

Шор Н.З. Исследование сходимости метода градиентного типа с растяжением пространства в направлении разности двух последовательных субградиентов. "Кибернетика 4, 48-53, 1975.

Эпоха использования r-алгоритма.

Программные реализации.

Задачи больших размерностей – схемы декомпозиции.

Прикладные задачи оптимизации.

Пакеты прикладных программ («ПЛАНЕР», «ДИСПРО»).

Настройка алгоритма на определенный класс прикладных задач.

Модификации r-алгоритма. **Регуляризация матрицы преобразования**.

Численная устойчивость. $\alpha \approx 2$. $B_k^* \to 0$ В r-алгоритме используется операция деления на $|g_k^*|$: $g_k^* = B_k^* g_k$, Ошибка «деление на нуль».
Вместо $R_\beta(\eta)$ использовать оператор $\tilde{R}_\beta(\eta) = (1/\sqrt[n]{\beta})R_\beta(\eta)$.
$\det(\tilde{R}_\beta(\eta) = 1$, $\det(R_\beta(\eta)) = \beta < 1$.
Этот прием можно использовать и для других алгоритмов, использующих матрицы преобразования пространства (**алгоритм эллипсоидов Юдина-Немировского**).
Решение задач небольшой размерности с большой точностью.
Журбенко Н.Г. Лиховид А.П. Регуляризация матрицы преобразования r-алгоритма. Теорія оптимальних рішень. – Київ: Ін-т кибернетики им. В.М.Глушкова НАН УкраЇни, 2018. – С. 145–151.

Модификации r-алгоритма. $r(\sigma)$-**алгоритмы.**

Основная проблема – выбор коэффициентов растяжения α и шагового множителя h.
$R_\alpha(\eta) = (\alpha - 1)\eta\eta^T + 1,\ |\eta| = 1$
$\tilde{R}_\sigma(\tilde{\eta}) = \sigma\tilde{\eta}\tilde{\eta}^T + I.$
Здесь $\tilde{\eta}$ – вектор R^n, σ- нормирующий множитель, $\sigma \in R^1$, $\sigma > 0$. Условие $|\tilde{\eta}| = 1$ не требуется.
Свойства оператора $\tilde{R}_\sigma(\tilde{\eta})$.
1. $\tilde{R}_\sigma(0) = I$
2. Пусть $|\tilde{\eta}| > 0$ и $\eta = \tilde{\eta}/|\tilde{\eta}|$.
Тогда $\tilde{R}_\sigma(\tilde{\eta}) = R_\alpha(\eta)$, где $\alpha = 1 + \sigma|\tilde{\eta}|^2$.
$\tilde{\eta}_{k+1} = g^*_{k+1} - g^*_k$ (разность субградиентов в преобразованном пространстве).
$\sigma_{k+1} = \sigma(g^*_{k+1}, g^*_k)$. Условие на функцию $\sigma(g_1, g_2)$:
$\sigma(\mu g_1, \mu g_2) = \sigma(g_1, g_2)/\mu^2$,
где $\mu \in R^1, \mu > 0$ (это условие обеспечивает независимость работы алгоритма от множителя на целевую функцию).

Модификации r-алгоритма. $r(\sigma)$-**алгоритмы**.

Примеры нормирующих множителей.
1. $\sigma_0(g_1, g_2) = 1/|g_2 - g_1|^2$. Алгоритм $r(\sigma_0)$ является r-алгоритмом с коэффициентом растяжения равным 2.
2.$\sigma_2(g_1, g_2) = 1/|Nr\{g_1, g_2\}|^2$, $Nr\{g_1, g_2\}$ – кратчайший вектор выпуклой оболочки векторов g_1, g_2.

$$1 \le \alpha \le \infty.$$

3. $\sigma_1(g_1, g_2) = 1/(|g_1|^2 + |g_1|^2)$.

$$1 \le \alpha \le 3.$$

r_n-алгоритм.
$\tilde{\eta}_{k+1} = g^*_{k+1}/|g^*_{k+1}| - g^*_k/|g^*_k|$.
$\sigma = 1$.

$$1 \le \alpha \le 5.$$

Модификации r-алгоритма. $r(\sigma)$-**алгоритмы**.

Численные результаты для $\sigma_1(g_1, g_2) = 1/|Nr\{g_1, g_2\}|^2$.

Тестовая задача: минимизация функции: $f(x) = \sum_{i=1}^{n} \rho_n^{i-1}|x_i|$,

где $\rho_n = 10^{6/(n-1)}$. Начальная точка $x_i = 10.0,\ i = 1, 2, ...n$.

Критерий останова: $f_k \leq 10^{-6}$.

Результаты решения тестовой задачи

Алгоритм	n	k	k_g	$\alpha_{\max}$	α_{avrg}
$r^*(\sigma_1)$	1000	18617	18618	5.3	3.6
$r(\sigma_1)$	1000	19183	19199	7.5	3.7
$r(\alpha)$	1000	35186	35199	2.0	2.0

Модификации r-алгоритма. $r(\sigma)$-**алгоритмы**.

Журбенко Н.Г., Лиховид А.П. К численной эффективности одной модификации r-алгоритма. Компьютерная математика. Киев: Ин-т кибернетики им. В.М.Глушкова НАН Украины, 2019. №1. С. 124–131.

Модификации r-алгоритма. $\alpha(\varepsilon)$-**алгоритм**.

$f^* = \min\{f(x) | x \in R^n\}$.
ε-оптимальное множество $X^*(\bar{\varepsilon})$:
$X^*(\bar{\varepsilon}) = \{x \in R^n | f(x) \leq f^* + \bar{\varepsilon}\}$.
Шар начальной локализации.
$(\varepsilon, \tilde{f})-$субградиент.
Агрегированный ε-субградиент.
Для числа итераций k $\alpha(\varepsilon)$-алгоритма, за которые он обеспечивает решение $2\bar{\varepsilon}$–оптимизации, справедлива следующая оценка:
$k \leq \left] n \frac{\ln(1/\gamma)}{\ln(1/q)} \right[$
$q \approx \sqrt{2}/2 \approx 0.707$;
Метод центров тяжести: $q \approx 1 - 1/e \approx 0.632$.
Но n^2 одномерных спусков (теоретическая оценка)!

Модификации r-алгоритма. $\alpha(\varepsilon)$-алгоритм.

Численная эффективность $\alpha(\varepsilon)$-алгоритма.

Тестовая задача: $f_1(x) = \sum\limits_{i=1}^{n} \rho_n^{i-1} x_i^2$, где $\rho_n = 10^{6/(n-1)}$.

Начальная точка $x_i = 1.0, \; i = 1, 2, ...n$. Точность решения по функционалу $\bar{\varepsilon} = 10^{-6}$.

nVarbl	qVolum	nIter	nLStep	Alpha
100	0.99	711	1.136	1.592
100	0.7	243	4.407	2.744

Журбенко Н.Г. Субградиентный алгоритм минимизации с преобразованием пространства – $\alpha(\varepsilon)$-алгоритм. В кн. Стохастическое программирование и его приложения. П.С. Кнопов, В.И. Зоркальцев, Я.М. Иваньо и др. –Иркутск: Институт систем энергетики им. Л.А. Мелентьева СО РАН, 2012. – С. 36 – 51

Квазиньютоновские методы

$r(\infty)$ – **r-алгоритм с бесконечным коэффициентом растяжения является алгоритмом сопряженных градиентов.** ($\alpha = \infty \leftrightarrow \beta = 1/\alpha = 0$)

Квазиньютоновские методы

В. Давидон 1959 метод переменной метрики.

Davidon–Fletcher–Powell (DFP),

Broyden–Fletcher–Goldfarb–Shanno (BFGS)

и др.

A. Rodomanov, Y. Nesterov New Results on Superlinear Convergence of Classical Quasi-Newton Methods (2020), arXiv:2004.14866

Презентация А.О. Родоманова: https://www.youtube.com/watch?v=0j45FpsLKyI

Квазиньютоновские методы

Различные варианты квазиньютоновских методов можно представить в форме преобразования пространства. Примеры такого представления приведены в работе ***Стецюк П.И. Линейные операторы в квазиньютоновских методах. //Теория и приложения методов оптимизации. К. Ин-т кибернетики им. В.М.Глушкова НАН Украины, 1998, с.3-8.*** Достаточно просто записывается условие на выбор очередного оператора преобразования, которое эквивалентно так называемому «квазиньютонвскому условию» коррекции матрицы квазиньютонвского метода.

Квазиньютоновские методы

Применительно к квадратичной функции «квазиньютонвское условие» состоит в следующем.

Выполняем шаг по антиградиенту.

Для оператора очередного преобразования должно выполняться условие: образ градиента является собственным вектором матрицы квадратичной функции в преобразованном пространстве.

При этом собственное число равно 1.

Для квадратичной функции образ траектория метода в преобразованном пространстве является движением по (взаимно ортогональным) собственным векторам матрицы квадратичной функции в преобразованном пространстве.

На основе такой интерпретации достаточно просто строится семейство вариантов квазиньютоновских алгоритмов на основе простейшего оператора растяжения $R_\alpha(\eta)$.

Алгоритм Поляка с преобразованием пространства

ε-субградиент в точке z: $\forall x$:$f(x) \geq \tilde{f} + (g, x - z) - \varepsilon$, $\tilde{f} \geq f^*; \varepsilon \in R^1$.

$\tilde{f}$ на итерации $k \Rightarrow$ полученный рекорд значения $f(x)$. f^* известно $\Rightarrow \tilde{f} = f^*$.

Так определенный ε-субградиент – частный A-субградиент Нурминского Е.А. ***Нурминский Е.А. А-субградиент выпуклых функций. Определение и основные свойства. Дальневост. мат. сб. –1995. – Вып. 1. – С. 132-136.***

Агрегированный ε-субградиент $\Rightarrow$ ε-субградиент с максимальной величиной "сдвига"отсекающей плоскости.

$h = (\bar{\varepsilon} - \varepsilon)/|g|$;

Если $\tilde{f} = f^*$ то субградиент $g \in \partial f(z)$ является ε-субградиентом с $\varepsilon = -(f(z) - f^*)$. Для случая обычной задачи минимизации ($\bar{\varepsilon} = 0$) величина "сдвига"$h = (\bar{\varepsilon} - \varepsilon)/|g| = (f(z) - f^*)/|g| \Rightarrow$ шаговый множитель метода Поляка ($\gamma = 1$).

Алгоритм Поляка с преобразованием пространства

Алгоритм основан на использовании: агрегатных $\varepsilon-$субградиентов; локализации множества решения; преобразования пространства (эллипсоид локализации $\Rightarrow$ шар.)
Оценка трудоемкости алгоритма .
$k < O\left(n(1/\gamma)\ln(1/\gamma)\right)$ для $n \ll 1/\gamma$ («большая точность»).
$k < O\left((1/\gamma^2)\ln(1/\gamma)\right)$ для $n \gg 1/\gamma$ («большая размерность»).
Для задач большой размерности и относительно малой точности решения оценка соответствует оценке трудоемкости алгоритма Б.Т. Поляка (не улучшаемую равномерно по n).
Для задач высокой точности решения оценку целесообразно сравнить с трудоемкостью алгоритма эллипсоидов
$k < O\left(n^2\ln(1/\gamma)\right)$.

Журбенко Н.Г. Субградиентный алгоритм Поляка с преобразованием пространства. XV Всероссийская конференция «Математическое программирование и приложения» г. Екатеринбург, 2 - 6 марта 2015 года

Заключение

Борис Теодорович Поляк:
«...разработка методов минимизации негладких функций требует привлечения новых идей. Одна из них, принадлежащая Н. З. Шору выглядит несколько неожиданно....».
Аркадий Семенович Немировский:
«Я не понимаю, как можно было придти к таким алгоритмам на интуитивном уровне..».
Юрий Михайлович Ермольев (о Науме):
«в науке, как и в искусстве, есть исполнители-интерпретаторы и есть композиторы...».

СПАСИБО!

zhurbnick@gmail.com