Теорія оптимальних рішень

УДК 519.8

Л.А. СОБОЛЕНКО, С.Г. НЕНАХОВА, И.А. ШУБЕНКОВА

МЕТОДЫ ГЛАДКОГО ШТРАФА НА ОСНОВЕ КОМБИНИРОВАННОЙ ШТРАФНОЙ ФУНКЦИИ

С помощью разработанной общей схемы аппроксимирующих методов последовательного квадратичного программирования, в основе которой лежит релаксация штрафной функции, содержащей гладкие и негладкие штрафы, описаны методы гладких штрафов для решения задач условной оптимизации с нелинейными ограничениями в форме неравенств.

Введение. Настоящая работа является продолжением работы*, в которой для решения гладких задач оптимизации с нелинейными ограничениями в форме неравенств была предложена общая схема аппроксимирующих методов последовательного квадратичного программирования на основе штрафной функции, объединяющей гладкие и негладкие штрафы, а также описаны алгоритмы негладкого штрафа.

В данной работе предлагается алгоритм гладкого штрафа на основе той же комбинированной штрафной функции.

Постановка задачи. По-прежнему рассматривается задача нелинейной оптимизации: найти

$$f_* = \min_{x\in\Omega} f(x),$$

$$\Omega = \left\{ x \in R^n \,\middle|\, g_i(x) \le 0,\, i = 1,\dots,m \right\}, \qquad (1)$$

где $f(x)$, $g_i(x)$, $i=1,\dots,m$ – дифференцируемые функции, градиенты $f'(x)$, $g_i'(x)$, $i=1,\dots,m,$ удовлетворяющие условию Липшица.

Под решением этой задачи понимается отыскание стационарных точек x_*, λ_*, удовлетворяющих необходимым условиям экстремума:

$$f'(x_*) + \sum_{i=1}^{m} \lambda_*^i g_i'(x_*) = 0,$$

$$\lambda_*^i g_i(x_*) = 0,\ \ \lambda_*^i \ge 0, \quad i = 1,\dots,m. \qquad (2)$$

---

* *Соболенко Л.А., Ненахова С.Г., Шубенкова И.А.* Комбинированная штрафная функция для построения различных методов решения нелинейных задач условной оптимизации // Теорія оптимальних рішень. – 2013. – С. 42–49.

Далее X_*, Λ_* – множества соответствующих точек x_*, λ_*, $I(x_*) = \{i = 1,...,$ $,..., m | g_i(x_*) = 0\}$, $I_* = \bigcup I(x_*),\, x_* \in X_*$.

Векторы рассматриваются как вектор-столбцы, $\|\cdot\|$ – евклидова норма вектора и согласованная с ней норма матриц: если $y \in R^r$, то $\|y\|_1 = \sum_{i=1}^{r} |y^i|$, $z(x,\lambda)$ – вектор с компонентами x и λ, $L(x,\lambda)$ – функция Лагранжа задачи (1).

Предлагаемая схема аппроксимирующих методов базируется на решении в каждой точке x_k итерационного процесса вспомогательной задачи квадратичного программирования

$$\min_x \left[\langle f'(x_k), x - x_k \rangle + \frac{1}{2} \langle A_k(x - x_k), x - x_k \rangle \right],$$

$$g_i(x_k) + \left\langle g_i{}'(x_k), x - x_k \right\rangle \le 0,\, i = 1,...,m, \tag{3}$$

где $A_k = A_k^T$ – сильно положительно определенная матрица

$$\underline{a}\|x\|^2 \le \langle A_k x, x \rangle \le \overline{a}\|x\|^2,\, \underline{a} > 0,\, \forall x \in R^n . \tag{4}$$

А. Относительно задачи (3) предполагается, что она разрешима (ее ограничения совместны) в любой точке итерационной последовательности $\{x_k\}$, которая предполагается ограниченной, и существует такая константа N и такие множители Лагранжа $\overline{\lambda}_k$, что

$$\|\overline{\lambda}_k\|_1 \le N,\ k = 0,1,... . \tag{5}$$

Пара Куна – Таккера $\{\overline{x}_k, \overline{\lambda}_k\}$ задачи (3) используется для поиска направления спуска, а величину спуска характеризует шаговый множитель α_k, определяемый из условия релаксации (убывания) некоторой штрафной функции.

Штрафная функция. В качестве таковой в [*] предложена смешанная (комбинированная) штрафная функция $F_{\mu\gamma}(z)$.

$$F_{\mu\gamma}(z) = f(x) + \sum_{i=1}^{m} \left(\mu - \lambda^i\right) g_i^+(x) + \frac{\gamma}{2} \sum_{i=1}^{m} \left(g_i^+(x)\right)^2, \quad \lambda \in \Lambda^+, \tag{6}$$

где $\Lambda^+ = \left\{\lambda \in R^m \middle| \lambda^i \ge 0, i = 1,...,m\right\}$, $\mu \ge 0$, $\gamma \ge 0$ – коэффициенты штрафа: μ – негладкий (недифференцируемый) штраф, γ – гладкий штраф, штраф $g^+(x)$ – вектор из R^m с компонентами $g_i^+(x) = \max\{0, g_i(x)\}$.

Будем называть алгоритм, основанный на стратегии увеличения коэффициента штрафа μ (при постоянном $\gamma \ge 0$), μ-алгоритмом, а на увеличении коэффициента штрафа γ (при постоянном $\mu \ge 0$) – γ-алгоритмом.

В работе [*] описан μ-алгоритм, а также свойства функции $F_{\mu\gamma}(z)$.

В частности, показано (лемма 2), что для μ-алгоритма при постоянном $\gamma \geq 0$ существует «пороговое» $\underline{\mu} \geq 0$ такое, что для любых $\mu \geq \underline{\mu}$ во всех точках ограниченной последовательности $\{z_k\} \in R^n \times \Lambda^+$ выполняется неравенство

$$b_k \leq (1-\varepsilon_1)(A_k p_k, p_k) - \varepsilon_1 \mu \left\| g^+(x_k) \right\|_1, \tag{7}$$

где $p_k = \bar{x}_k - x_k$, $\varepsilon_1 \in (0;1)$ – любая константа.

Это неравенство гарантирует релаксацию функции $F_{\mu\gamma}(z)$ в точке z_k вдоль направления $\Delta z_k = \{p_k, \bar{\lambda}_k - \lambda_k\}$:

$$\partial F_{\mu\gamma}(z_k) \Big/ \partial \Delta z_k \leq -\varepsilon_1 \left[\langle A_k p_k, p_k \rangle + \mu \left\| g^+(x_k) \right\|_1 \right] + \langle \bar{\lambda}_k, g^-(x_k) \rangle, \tag{8}$$

где $g^-(x_k)$ – вектор из R^m с компонентами $g_i^-(x) = \min\{0, g_i(x)\}$.

Поскольку правая часть неравенства (8) в точке $x_k \neq x_*$ отрицательна, производная функции $F_{\mu\gamma}(z)$ по направлению Δz_k убывает.

Заметим, что в отличие от μ (лемма 2 в [*]) коэффициент γ, при котором начинает выполняться неравенство (7) и, следовательно, (8) может расти неограниченно в точках $z_k \in R^n \times \Lambda^+$, $k \to \infty$, т. е. «порогового» $\underline{\gamma} \geq 0$, вообще говоря, не существует.

Напомним, что $z_k = \{x_k, \lambda_k\} \in R^n \times \Lambda^+$ – точки, которые будут генерироваться γ-алгоритмом при решении задачи (1), $\{\bar{x}_k, \bar{\lambda}_k\}$ – пара Куна – Таккера задачи (3), $\Delta\hat{\lambda}_k = \bar{\lambda}_k - \hat{\lambda}_k$ – направление спуска по λ, зависящее от выбора $\hat{\lambda}_k$.

Если $\hat{\lambda}_k = \{\hat{\lambda}_k^i\}$, $\hat{\lambda}_k^i = \min\{\lambda_k^i, \bar{\lambda}_k^i\}$, $i = 1,\dots,m$, то $\Delta\hat{\lambda}_k = \{\bar{\lambda}_k - \lambda_k\}^+$; если $\hat{\lambda}_k = \bar{\lambda}_k$, то $\Delta\hat{\lambda}_k = 0$.

При замене λ_k, $z_k = \{x_k, \lambda_k\}$ на $\hat{\lambda}_k$, $\hat{z}_k = \{x_k, \hat{\lambda}_k\}$ выражение (7) (см. (13) из [*]) принимает вид:

$$2\langle \Delta\hat{\lambda}_k, g^+(x_k) \rangle \leq (1-\varepsilon_1)\left[\langle A_k p_k, p_k \rangle + \mu \left\| g^+(x_k) \right\|_1 \right] + \gamma \left\| g^+(x_k) \right\|^2. \tag{9}$$

Приведем основное свойство γ-алгоритмов, вытекающее из неограниченного роста γ в точках x_k.

Лемма 1. Пусть неравенство (9) нарушено в точке $\hat{z}_k \in R^n \times \Lambda^+$ при $\gamma = \gamma(x_k)$ и выполняется при любых $\gamma > \gamma(x_k)$; последовательности $\{x_k\}$, $\{\bar{\lambda}_k\}$ ограничены, $\mu > 0$ фиксировано.

Если учесть $\gamma(x_k) \to \infty$ при $k \to \infty$, то $p_k \to 0$, $x_k \to X_*$, $\bar{\lambda}_k \to \Lambda_*$, $\left\| g^+(x_k) \right\| \le C\gamma_k^{-1} \to 0$.

Доказательство. Нарушение (9) при $\gamma = \gamma(x_k)$ вследствие (4) приводит к неравенству

$$(1-\varepsilon_1)\left[\langle A_k p_k, p_k \rangle + \mu \left\| g^+(x_k) \right\|_1\right] + \gamma_k \left\| g^+(x_k) \right\|^2 \le 2\left\langle \Delta\hat{\lambda}_k, g^+(x_k) \right\rangle, \qquad (10)$$

которое возможно только при $\left\langle \Delta\hat{\lambda}_k, g^+(x_k) \right\rangle > 0$. Из ограниченности $\{\bar{\lambda}_k\}$ и определения $\hat{\lambda}_k$ вытекает ограниченность последовательности $\{\hat{\lambda}_k\}$, $\{\Delta\hat{\lambda}_k\}$. С учетом ограниченности $\{x_k\}$, $\{\Delta\hat{\lambda}_k\}$ из (10) и (4) получим, что $\left\| g^+(x_k) \right\| \le C\gamma_k^{-1}$, где C – константа. Поэтому $g^+(x_k) \to 0$. Тогда на основании (10) и (4) выполняется $p_k \to 0$ – необходимое и достаточное условие того, что $x_k \to X_*$, $\bar{\lambda}_k \to \Lambda_*$ и $\left\| g^+(x_k) \right\| \le C\gamma_k^{-1} \to 0$. Лемма доказана.

Формулировка метода. Примем ε и ε_1 любыми константами из (0;1); $\gamma_0 = 1$, $x_0 \in R^n$ произвольно, $\mu > 0$ – любое фиксированное число, его не будем указывать в обозначениях и вместо $F_{\mu\gamma}(z)$ будем писать $F(z, \gamma_k)$.

Каждая k-ая итерация в точке z_k ($x_k \ne x_*$, $p_k \ne 0$), $k = 0,1,...$ состоит в следующем.

Шаг 1 (определение p_k и $\bar{\lambda}_k$). Задать матрицу $A_k = A_k^T$, удовлетворяющую (4) и, решая задачу (3), найти $\bar{x}_k$, $\bar{\lambda}_k$, p_k.

Шаг 2 (спуск по λ и сохранение штрафа $\gamma_k = \gamma_{k-1}$). Вычислить $\hat{\lambda}_k^i = \min\{\lambda_k^i; \bar{\lambda}_k^i\}$, $i = 1,...,m$ и найти $\hat{z}_k = \{x_k; \hat{\lambda}_k\}$, $\Delta\hat{\lambda}_k = \left(\bar{\lambda}_k - \hat{\lambda}_k\right)^+$, полагая $\hat{\lambda}_k = \bar{\lambda}_0$ при $k = 0$. В случае $\Delta\hat{\lambda}_k = 0$ принять $\gamma_k = \gamma_{k-1}$ и перейти к шагу 4; при нарушении к шагу 3.

Шаг 3 (вычисление нового γ_k и переопределение $\hat{z}_k$, $\Delta\hat{z}_k$).

Найти γ_k, как наименьшее из значений $\gamma = \gamma_{k-1}^{2^s}$, $s = 1,2,...$, для которого впервые выполняется (9). Взять в качестве нового $\hat{\lambda}_k$ вектор $\bar{\lambda}_k$ и, соответственно, новые $\hat{z}_k = \{x_k, \bar{\lambda}_k\}$, $\Delta\hat{\lambda}_k = 0$.

Шаг 4 (определение α_k). Найти α_k как наименьшее из значений $\alpha = 2^{-s}$, $s = 0,1,...$, для которого впервые выполнится неравенство

$$F(\hat{z}_k, +\alpha_k \Delta\hat{z}_k, \gamma_k) \le F(\hat{z}_k, \gamma_k) + \varepsilon\alpha\hat{\varphi}_k, \quad 0 < \varepsilon < 1, \qquad (11)$$

где

$$\tilde{\psi}_k = -\frac{\varepsilon_1}{2}\left[\left\langle A_k p_k, p_k\right\rangle + \mu\left\|g^+(x_k)\right\|_1\right] + \left\langle\tilde{\lambda}_k, g^-(x_k)\right\rangle,\ 0 < \varepsilon_1 < 1. \tag{12}$$

Шаг 5. Вычислить $z_{k+1} = \tilde{z}_k + \alpha_k \Delta\tilde{z}_k$, $k := k+1$.

Конец итерации.

Отметим, что при поиске γ_k в неравенстве (9) используются вычисленные на шаге 2 значения $\tilde{\lambda}_k$, $\tilde{z}_k$, $\Delta\tilde{\lambda}_k$, $\Delta\tilde{z}_k$. При этом в (11) в качестве $\tilde{\lambda}_k$ выступает вектор с компонентами ($i = 1,...,m$)

$$\tilde{\lambda}_k^i = \begin{cases} \min\left\{\lambda_k^i; \bar{\lambda}_k^i\right\}, & \text{если} \quad \gamma_k = \gamma_{k-1}, \\ \bar{\lambda}_k^i, & \text{если} \quad \gamma_k > \gamma_{k-1}. \end{cases} \tag{13}$$

Соответсвенно (13) определяется и $\Delta\tilde{\lambda}_k = \bar{\lambda}_k - \tilde{\lambda}_k$ в (11): $\Delta\tilde{\lambda}_k = \left(\bar{\lambda}_k - \lambda_k\right)^+$ при $\gamma_k = \gamma_{k-1}$ и $\Delta\tilde{\lambda}_k = 0$ при $\gamma_k > \gamma_{k-1}$.

Вследствие ограниченности объема публикации, сформулируем свойства γ -алгоритма в виде двух теорем без доказательств.

Теорема 1. Пусть градиенты $f(x)$, $g_i(x)$, $i = 1,...,m$ удовлетворяют условию Липшица, для вспомогательной задачи (3) выполняется предположение А, $x_0 \in R^n$ – произвольно, x_{k_t} – точки увеличения γ.

Тогда.

1. Если $\gamma_k \to \infty$, то при $k \to \infty$,

$$x_{k_t} \to X_*,\ x_k \to \Omega,\ F(z_k, \gamma_k) \to F_*, \tag{14}$$

где $F_* = f(x_*)$ – значение функции $f(x)$ в одной из стационарных точек $x_* \to X_*$. При дополнительном условии выпуклости задачи (1) будет

$$x_k \to X_*,\ F_* = f_*,\ f(x_k) \to f_*. \tag{15}$$

2. Если $\gamma_k = \gamma > 0$ – константа при $k \ge k_0$, то справедливы утверждения из [*] для μ-алгоритма о нелокальной сходимости $x_k \to X_*$ (теорема 1) и локальных оценках скорости сходимости (теорема 2).

Оценки локальной скорости сходимости устанавливались в зависимости от выполнения условий:

А1. Функции $f(x)$, $g_i(x)$, $i \in I(x_*)$ выпуклы в окрестности множества Ω, содержащего внутреннюю точку.

А2. Функции $f(x)$, $g_i(x)$, $i \in I(x_*)$ дважды непрерывно дифференцируемы в окрестности x_*, градиенты $g_i'(x)$, $i \in I(x_*)$ линейно-независимы и для всех x таких, что $\left\langle g_i'(x_*), x\right\rangle = 0$, $\forall i \in I(x_*) = \left\{i \in I(x_*) \middle| \lambda_*^i > 0\right\}$ имеет место $\left\langle L_{xx}(z_*)x, x\right\rangle > 0$.

А3. Градиенты $g_i'(x_*)$, $i \in I(x_*) = \{i = 1,...,m\}$ – линейно независимы и $\lambda_*^i > 0$ для всех $i \in I(x_*)$.

Ограниченность генерируемых алгоритмом значений γ_k исследуется в теореме 2.

Теорема 2. Пусть выполнены условия теоремы 1 и А1, $x_k \to x_*$ при $k \to \infty$. Тогда при дополнительном условии А2 или А3 будет $\gamma_k = \gamma > 0$ – константа для $k \ge k_0$, при этом $\lambda_k \to \lambda_*$, $\tilde{\lambda}_k \to \lambda_*$, $\Delta\tilde{\lambda}_k \to 0$ и справедливы оценки локальной скорости сходимости

$$\|x_k - x_*\| \le C_1 q^k, \ \|x_{k+1} - x_*\| \le C_2 \|x_k - x_*\|^2,$$

где $q \in (0;1)$, C_1, C_2. – константы.

Заключение. В отличие от μ-алгоритма [*], нелокальная сходимость (14) в γ-алгоритме гарантируется, вообще говоря, при $\gamma \to \infty$. Однако при их комбинации в предложенной схеме присутствие постоянного $\mu > 0$ положительно влияет на характер генерируемых γ_k. Так, если $\mu \ge \left\|\bar{\lambda}_k - \lambda_k\right\|^+$, $k \ge k_0$, то неравенство (9) выполнится, гарантируя ограниченность $\gamma_k \le \gamma$ как для выпуклой, так и невыпуклой задачи (1). Если же $g_i'(x_*)$, $i \in I(x_*)$ линейно-независимы в предельной точке x_*, то при любом постоянном $\mu > 0$ генерируется $\gamma_k \le \gamma$. Таким образом, сам факт присутствия коэффициента $\mu > 0$ обуславливает ограниченность генерируемых γ_k.

Л.О. Соболенко, С.Г. Ненахова, І.А. Шубенкова

МЕТОДИ ГЛАДКОГО ШТРАФУ НА ОСНОВІ КОМБІНОВАНОЇ ШТРАФНОЇ ФУНКЦІЇ

За допомогою розробленої загальної схеми апроксимуючих методів послідовного квадратичного програмування, в основі якої лежить релаксація штрафної функції, що містить гладкі та негладкі штрафи, описано методи гладких штрафів для розв'язування задач умовної оптимізації з нелінійними обмеженнями у вигляді нерівностей.

L.A. Sobolenko, S.G. Nenakhova, I.A. Shubenkova

SMOOTH PENALTY METHODS ON THE BASIS OF COMBINED PENALTY FUNCTION

By means of the developed general scheme of approximating methods of sequential square programming which is based on the relaxation of the penal function containing smooth and unsmooth penalties, smooth penalty methods for the solution of problems of the conditional optimization with nonlinear restrictions in the form of inequalities are described.

Получено 20.03.2014