Численные методы

УДК 519.853+519.632

# О СКОРОСТИ СХОДИМОСТИ РЕГУЛЯРИЗОВАННОГО МЕТОДА СГЛАЖИВАНИЯ ТОЧНЫХ ФУНКЦИЙ ШТРАФА

А.А.Каплан

Пусть $H$ – гильбертово пространство; $f, g^j : H \to R$ – выпуклые дифференцируемые по Фреше функции; множество $X^*$ решений задачи

$$f(x) \to min\,! \tag{1}$$

$$g^j(x) \le 0, \; j = \overline{1,m}, \tag{2}$$

непусто и ограничено и выполнено условие Слейтера.

Для (1)-(2) строится минимизирующая последовательность

$$z^k \approx \underset{x \in H}{argmin} \{ F_k(x) + \|x - z^{k-1}\|^2 \}, \tag{3}$$

где

$$F_k(x) = f(x) + \frac{1}{2}\sum_{j=1}^{m} a_j^k \left(g^j(x) + \sqrt{g^{j2}(x) + \tau_k}\,\right), \tag{4}$$

$\lim a^k = a > y$ ( $y$ – произвольный вектор Лагранжа), $\tau_k > 0$, $\lim \tau_k = 0$, $z^0$ – любое.

Итерационный процесс (3)-(4) естественно рассматривать [1,2] как регуляризованный вариант метода штрафов с функциями штрафа $V_k(x) = \frac{1}{2}\sum_{j=1}^{m} a_j^k \left(g^j(x) + \sqrt{g^{j2}(x) + \tau_k}\,\right)$, аппроксимирующими точную функцию штрафа $V(x) = \sum_{j=1}^{m} a_j \, max[0, g^j(x)]$.

С учетом того, что

$$\lim_{k \to \infty} F_k(x) = f(x) + V(x) \equiv F(x)$$

и $\operatorname{Arg}\min\limits_{x\in H} F(x) = X^*$, исследование процесса[ж)] (3)-(4) может быть осуществлено в рамках следующего формального подхода к решению задач безусловной минимизации выпуклой функции.

Пусть $\varphi, \varphi_k : H \to R$ – выпуклые непрерывные функции ( $k=1,2,\dots$); множество $X^\varphi = \{x : \varphi(x) = \inf\limits_{z\in H} \varphi(z) \equiv \varphi_{min}\}$ непусто и ограничено; $\{\delta_k\}$, $\{\varepsilon_k\}$ – заданные последовательности положительных чисел, $\lim \delta_k = \lim \varepsilon_k = 0$; $\Omega = \{x \in H : \varphi(x) \le c\}$.

Предполагая

$$\sup_{x\in\Omega} |\varphi_k(x) - \varphi(x)| \le \delta_k, \quad k=1,2,\dots, \tag{5}$$

определим итерационную последовательность $\{x^k\}$ :

$$\varphi_k(x^k) + \|x^k - x^{k-1}\|^2 \le \min_{x\in H} \{\varphi_k(x) + \|x - x^{k-1}\|^2\} + \varepsilon_k \tag{6}$$

( $x^0$ – любое).

В [2] для $H = R^n$ установлено, что если

$$\sum_{k=1}^{\infty} (\varepsilon_k + \delta_k) < \infty \quad \text{и} \quad c > \varphi(x^0) + \sum_{k=1}^{\infty} (2\delta_k + \varepsilon_k), \tag{7}$$

то $\lim\limits_{k\to\infty} \inf\limits_{x\in X^\varphi} \|x^k - x\| = 0$ (теорема 6 с учетом замечания 3), причем в случае $\sum\limits_{k=1}^{\infty} (\sqrt{\varepsilon_k} + \sqrt{\delta_k}) < \infty$ последовательность $\{x^k\}$ сходится к некоторой точке $\bar{x} \in X^\varphi$ (теорема 7). Отсюда, например, сразу вытекает, что при $a^k = a$, $k=1,2,\dots$,

$$F_k(z^k) + \|z^k - z^{k-1}\|^2 \le \min_{x\in R^n} \{F_k(x) + \|x - z^{k-1}\|^2\} + \varepsilon_k$$

и $\sum\limits_{k=1}^{\infty} (\tau_k^{1/4} + \varepsilon_k^{1/2}) < \infty$ последовательность $\{z^k\}$ сходится к некоторому решению задачи (1)-(2).

Для бесконечномерного пространства $H$ аналогичные результаты о сходимости $\{x^k\}$ не установлены. Более того, до сих пор не выяснено, сходится ли в сильном смысле последовательность

$$v^k = \arg\min_{x\in H} \{\varphi(x) + \|x - v^{k-1}\|^2\},$$

если $X^\varphi$ состоит из одной точки.

Однако оценки скорости сходимости к 0 последовательности

---

ж) Естественно, с уточнением условия выбора $z^k$.

$\{\gamma_k\}$, где $\gamma_k = \varphi(x^k) - \varphi_{min}$, полученные ниже для итерационного процесса (5)-(6) в бесконечномерном случае, являются новыми и при $H = R^n$. На их основе легко устанавливаются и оценки сходимости $f(z^k)$ к $\bar{f} = min\{f(x) : g^j(x) \le 0,\ j = \overline{1,m}\}$ в методе (3)-(4) - при соответствующем критерии определения $z^k$ и и выборе параметров $a^k$, $\tau_k$.

ЛЕММА. Пусть выполнено условие (7) и множество $\Omega$ ограниченное. Тогда последовательность $\{x^k\}$, определенная согласно (5)-(6), ограничена.

ДОКАЗАТЕЛЬСТВО. Предположим, что при фиксированном $k$ имеет место

$$\varphi(x^{k-1}) \le \varphi(x^0) + \sum_{i=1}^{k-1} (2\delta_i + \varepsilon_i).$$

Из неравенства (6) ввиду (5) следует

$$\varphi_k(x^k) \le \varphi_k(x^{k-1}) + \varepsilon_k \le \varphi(x^{k-1}) + \delta_k + \varepsilon_k \le$$
$$\le \varphi(x^0) + \sum_{i=1}^{k-1}(2\delta_i + \varepsilon_i) + \delta_k + \varepsilon_k.$$

Пусть $x$ такое, что

$$\varphi(x) = \varphi(x^0) + \sum_{i=1}^{k}(2\delta_i + \varepsilon_i) \equiv q_k.$$

Тогда $\varphi_k(x) \ge \varphi(x) - \delta_k \ge \varphi_k(x^k)$, и остается заметить, что в точке $\bar{x} \in X^\varphi$ справедливо $\varphi_k(\bar{x}) \le \varphi(\bar{x}) + \delta_k < q_k - \delta_k$. С учетом выпуклости $\varphi_k$ это означает, что $x^k \in \{x : \varphi(x) \le q_k\}$ и, таким образом, $\{x^k\} \subset \Omega$. ■

Обозначим $\bar{x}^k = arg\,min_{z \in X^\varphi} \|x^k - z\|$ и будем считать, что $\Omega$ ограничено, выбор $\{\delta_k\}$, $\{\varepsilon_k\}$ и $c$ подчинен условию (7), причем при всех $k$ выполнено неравенство

$$\tau_k \equiv \varepsilon_{k+1} + b\delta_{k+1} \le \frac{\alpha}{4}\left(\frac{\gamma_0}{1+\alpha(k+1)\gamma_0}\right)^2. \tag{8}$$

Здесь

$$b \ge 4q^{-1} \sup_k \gamma_k + 4, \quad q \ge 4 \sup_k \|x^k - \bar{x}^k\|^2,$$
$$0 < \alpha \le \frac{q^{-1}}{2}, \qquad \alpha \sup_k \gamma_k \le 1/4. \tag{9}$$

Легко видеть, что при $\gamma_0 > 0$ эти условия непротиворечивы (существование $q < \infty$ является следствием леммы).

ТЕОРЕМА I. Пусть выбор $\{\delta_k\}, \{\varepsilon_k\}, c$ и $\alpha$ подчинен условиям (7)-(9), $Q$ - ограниченное множество. Тогда для итерационного процесса (5)-(6) справедлива оценка

$$\gamma_k < \frac{\gamma_0}{1+\alpha k \gamma_0} . \tag{10}$$

ДОКАЗАТЕЛЬСТВО. На основании (6) и выпуклости $\varphi_{k+1}$ при $\lambda \in [0,1]$ имеем

$$\varphi_{k+1}(x^{k+1}) \le \varphi_{k+1}(x^k + \lambda(\bar{x}^k - x^k)) + \lambda^2 \|\bar{x}^k - x^k\|^2 + \varepsilon_{k+1} \le$$

$$\le \lambda \varphi_{k+1}(\bar{x}^k) + (1-\lambda)\varphi_{k+1}(x^k) + \lambda^2 \|\bar{x}^k - x^k\|^2 + \varepsilon_{k+1} \equiv \ell_k(\lambda) + \varepsilon_{k+1}.$$

Минимизируя $\ell_k$ на промежутке $[0,1]$, рассмотрим возможные варианты.

(а) Если $\lambda_k \equiv \arg\min_{0 \le \lambda \le 1} \ell_k(\lambda) = 0$, то $\ell_k(\lambda_k) = \varphi_{k+1}(x^k)$, причем $\lambda_k = 0$ возможно, лишь если

$$\varphi_{k+1}(x^k) \le \varphi_{k+1}(\bar{x}^k).$$

Тем самым

$$\varphi_{k+1}(x^{k+1}) \le \varphi_{k+1}(\bar{x}^k) + \varepsilon_{k+1},$$

откуда

$$\gamma_{k+1} \le 2\delta_{k+1} + \varepsilon_{k+1} . \tag{11}$$

(б) Если $\lambda_k = 1$, то $\ell_k(\lambda_k) = \varphi_{k+1}(\bar{x}^k) + \|\bar{x}^k - x^k\|^2$, причем в этом случае, как нетрудно видеть,

$$\frac{\varphi_{k+1}(x^k) - \varphi_{k+1}(\bar{x}^k)}{2\|\bar{x}^k - x^k\|^2} \ge 1,$$

так что

$$\ell_k(\lambda_k) \le \frac{\varphi_{k+1}(x^k) + \varphi_{k+1}(\bar{x}^k)}{2},$$

$$\varphi_{k+1}(x^{k+1}) \le \frac{\varphi_{k+1}(x^k) + \varphi_{k+1}(\bar{x}^k)}{2} + \varepsilon_{k+1}$$

и с учетом (5)

$$\gamma_{k+1} \le \frac{\gamma_k}{2} + 2\delta_{k+1} + \varepsilon_{k+1} . \tag{12}$$

(в) Если $\lambda_k \in (0,1)$, то $\varphi_{k+1}(x^k) > \varphi_{k+1}(\bar{x}^k)$,

$$\lambda_k = \frac{\varphi_{k+1}(x^k)-\varphi_{k+1}(\bar{x}^k)}{2\|\bar{x}^k-x^k\|^2} \tag{I3}$$

и

$$\varphi_{k+1}(x^{k+1}) \leq \varphi_{k+1}(x^k) - \frac{(\varphi_{k+1}(x^k)-\varphi_{k+1}(\bar{x}^k))^2}{4\|\bar{x}^k-x^k\|^2} + \varepsilon_{k+1}. \tag{I4}$$

В случае $\varphi(x^k)-\varphi(\bar{x}^k) > 2\delta_{k+1}$ ввиду (5) имеем

$$(\varphi_{k+1}(x^k)-\varphi_{k+1}(\bar{x}^k))^2 > (\varphi(x^k)-\varphi(\bar{x}^k))^2 - \\ -4\delta_{k+1}(\varphi(x^k)-\varphi(\bar{x}^k)) + 4\delta_{k+1}^2$$

и из (5), (I3), (I4) следует

$$\gamma_{k+1} < \gamma_k - \gamma_k^2 q^{-1} + 4\gamma_k \delta_{k+1} q^{-1} + \varepsilon_{k+1} + 2\delta_{k+1} < \\ < \gamma_k - 2\alpha\gamma_k^2 + \varepsilon_{k+1} + b\delta_{k+1}. \tag{I5}$$

Если же $\varphi(x^k)-\varphi(\bar{x}^k) \leq 2\delta_{k+1}$, то на основании (5),(6)

$$\gamma_{k+1} \leq \varphi_{k+1}(x^{k+1}) - \varphi(\bar{x}^k) + \delta_{k+1} \leq \varphi_{k+1}(x^k) - \varphi(\bar{x}^k) + \delta_{k+1} + \varepsilon_{k+1} \leq \\ \leq \gamma_k + 2\delta_{k+1} + \varepsilon_{k+1} \leq 4\delta_{k+1} + \varepsilon_{k+1}. \tag{I6}$$

Теперь с учетом выбора $\alpha$ и $b$ нетрудно проверить, что неравенство

$$\gamma_{k+1} \leq \gamma_k - 2\alpha\gamma_k^2 + \tau_k \tag{I7}$$

является следствием любого из неравенств (II), (I2), (I5), (I6) и тем самым выполняется при любом $k$.

Имея в виду $\gamma_k \geq 0$ и $\tau_k \leq \frac{\alpha}{4}\left(\frac{\gamma_0}{1+\alpha(k+1)\gamma_0}\right)^2$, покажем, что при любом $k$ справедлива оценка (I0).

При $k=0$ получаем

$$\gamma_1 \leq \gamma_0 - 2\alpha\gamma_0^2 + \frac{\alpha}{4}\left(\frac{\gamma_0}{1+\alpha\gamma_0}\right)^2 < \frac{\gamma_0}{1+\alpha\gamma_0}.$$

Пусть при $k=\ell$

$$\gamma_\ell \leq \frac{\gamma_0}{1+\alpha\ell\gamma_0}. \tag{I8}$$

Если $\tau_\ell \geq \alpha\gamma_\ell^2$, то из (8)

$$\gamma_\ell \leq \frac{\gamma_0}{2(1+\alpha(\ell+1)\gamma_0)},$$

и тогда

$$\gamma_{\ell+1} < \frac{\gamma_0}{2(1+\alpha(\ell+1)\gamma_0} + \frac{\alpha}{4}\left(\frac{\gamma_0}{1+\alpha(\ell+1)\gamma_0}\right)^2 < \frac{\gamma_0}{1+\alpha(\ell+1)\gamma_0}. \tag{I9}$$

Если же $\tau_\ell < \alpha\gamma_\ell^2$ , то ввиду (I7) $\gamma_{\ell+1} < \gamma_\ell - \alpha\gamma_\ell^2$ . Используя то обстоятельство, что функция $y - \alpha y^2$ при $0 \le \alpha y \le 1/2$ возрастающая и $\alpha\gamma_0 \le 1/4$ , на основании (I8) имеем

$$\gamma_{\ell+1} < \frac{\gamma_0}{1+\alpha\ell\gamma_0} - \alpha\left(\frac{\gamma_0}{1+\alpha\ell\gamma_0}\right)^2 <$$

$$< \frac{\gamma_0}{1+\alpha\ell\gamma_0}\left(1 - \frac{\alpha\gamma_0}{1+\alpha(\ell+1)\gamma_0}\right) = \frac{\gamma_0}{1+\alpha(\ell+1)\gamma_0}. \qquad (20)$$

Объединяя (I9), (20), завершаем индукцию. ■

Допустим, что функция $\varphi$ удовлетворяет на $\Omega$ условию Липшица с константой $L$ и справедливо неравенство

$$\inf_{x \in \Omega \setminus X^\varphi} \frac{\varphi(x) - \varphi_{min}}{\rho^2(x, X^\varphi)} \ge d > 0, \qquad (21)$$

где $\rho(x, X^\varphi) = \min_{z \in X^\varphi} \|x - z\|$.

Проанализируем в этом предположении более детально случай, когда $0 < \lambda_k < 1$ и $\varphi(x^k) - \varphi(\bar{x}^k) > 2\delta_{k+1}$ . Будем считать, что $d < 1/2$.

Из (5), (I3), (I4) имеем

$$\varphi(x^{k+1}) - \varphi(\bar{x}^k) < \varphi(x^k) - \varphi(\bar{x}^k) - \frac{(\varphi(x^k) - \varphi(\bar{x}^k))^2}{4\|\bar{x}^k - x^k\|^2} +$$

$$+ \frac{\varphi(x^k) - \varphi(\bar{x}^k)}{\|\bar{x}^k - x^k\|^2}\delta_{k+1} + \varepsilon_{k+1} + 2\delta_{k+1}.$$

Если $\|\bar{x}^k - x^k\| \le \delta_{k+1}^{1/3}$ , то $\varphi(x^k) - \varphi(\bar{x}^k) \le L\,\delta_{k+1}^{1/3}$ , откуда ввиду (5), (6)

$$\gamma_{k+1} \le L\,\delta_{k+1}^{1/3} + 2\delta_{k+1} + \varepsilon_{k+1}. \qquad (22)$$

Если же $\|\bar{x}^k - x^k\| > \delta_{k+1}^{1/3}$ , то

$$\frac{\varphi(x^k) - \varphi(\bar{x}^k)}{\|\bar{x}^k - x^k\|^2}\delta_{k+1} < (\varphi(x^k) - \varphi(\bar{x}^k))\delta_{k+1}^{1/3} < L_1\,\delta_{k+1}^{1/3},$$

где $L_1 = \sup_k \gamma_k$ . Тем самым

$$\gamma_{k+1} < \gamma_k - \frac{\gamma_k^2}{4\|\bar{x}^k - x^k\|^2} + L_1\,\delta_{k+1}^{1/3} + 2\delta_{k+1} + \varepsilon_{k+1},$$

и, используя (2I), получаем

$$\gamma_{k+1} < \gamma_k - \frac{d}{4}\gamma_k + L_1\delta_{k+1}^{1/3} + 2\delta_{k+1} + \varepsilon_{k+1}. \tag{23}$$

Полагая $\tau_k' = max\{\tau_k, L\delta_{k+1}^{1/3} + 2\delta_{k+1} + \varepsilon_{k+1}, L_1\,\delta_{k+1}^{1/3} + 2\delta_{k+1} + \varepsilon_{k+1}\}$, имеем при выполнении условий теоремы I, что неравенство

$$0 \le \gamma_{k+1} \le \left(1 - \frac{d}{4}\right)\gamma_k + \tau_k' \tag{24}$$

является следствием любого из неравенств (II), (I2), (I6), (22), (23), т.е. выполняется при любом $k$ .

Итак, справедлив следующий результат.

ТЕОРЕМА 2. Пусть выполнены предположения теоремы I; функция $\varphi$ удовлетворяет на $\Omega$ условию Липшица; справедливо неравенство (2I). Тогда если $\tau_k'$ достаточно быстро стремится к 0, то последовательность $\{\gamma_k\}$ сходится к 0 со скоростью геометрической прогрессии.

ЗАМЕЧАНИЕ I. Если функции $\varphi_k$ дифференцируемы по Фреше на $H$ , ввиду сильной выпуклости $\varphi_k(x) + \|x - x^{k-1}\|^2$ из неравенства

$$\|\nabla\varphi_k(x^k) + 2(x^k - x^{k-1})\| \le \varepsilon_k \tag{25}$$

следует, что

$$\varphi_k(x^k) + \|x^k - x^{k-1}\|^2 \le \min_{x\in H}\left\{\varphi_k(x) + \|x - x^{k-1}\|^2\right\} + \frac{\varepsilon_k^2}{2}.$$

Это позволяет очевидным образом переформулировать полученные результаты для случая, когда точки $x^k$ определяются по градиентному критерию (25); именно, в соотношениях (7),(8) и в выражении для $\tau_k'$ вместо $\varepsilon_k$ следует взять $\varepsilon_k^2/2$ .

Для функций $F$ и $F_k$ легко проверяется неравенство

$$|F(x) - F_k(x)| \le \frac{1}{2}\tau_k^{1/2}\sum_{j=1}^m a_j +$$

$$+\frac{1}{2}\sum_{j=1}^m |a_j^k - a_j|\left(g^j(x) + \sqrt{g^{j2}(x) + \tau_k}\right),$$

позволяющее распространить теорему I (а при подходящих условиях относительно $F$ - и теорему 2) на итерационный процесс (3), (4) решения исходной задачи (I), (2).

В частности, при $a^k = a$ ( $k=1,2,\ldots$ ) оценка (10) имеет место, если подчинить выбор $\{\tau_k\}$ условию $\tau_k^{1/2} \le \delta_k$ , а $z^k$ определять по критерию

$$F_k(z^k) + \|z^k - z^{k-1}\|^2 \le \min_{x \in H} \{F_k(x) + \|x - z^{k-1}\|^2\} + \varepsilon_k$$

с прежними предположениями относительно $\{\varepsilon_k\}$ , $\{\delta_k\}$ (годится $\Omega = H$ ).

В общем случае $\gamma_k = F(x^k) - F_{min}$ . Однако можно показать, что если точки $x^k$ определены по градиентному критерию (25) с $\varphi_k = F_k$ и $\lim \varepsilon_k = 0$ , то, начиная с некоторого номера, справедливо*) $g^j(x^k) \le 0$, $j=\overline{1,m}$ , и тем самым $F(x^k) = f(x^k)$ , $\gamma_k$ есть отклонение $f(x^k)$ от $\bar{f}$ .

ЗАМЕЧАНИЕ 2. Класс функций, удовлетворяющих условию (21), далеко не ограничивается сильно выпуклыми функциями. Так, например, (21) очевидным образом выполнено, если $H = R^n$ , $\varphi$ есть *supremum* конечного числа аффинных функций и $X^\varphi$ ограничено. Однако проверка условия (21) , вообще говоря, связана с серьезными трудностями. Не останавливаясь здесь на этом вопросе, укажем лишь, что теорема 2 оказывается полезной при исследовании регуляризованного метода штрафов в применении к эллиптическим вариационным неравенствам, если ядро соответствующего дифференциального оператора конечномерно.

## Литература


1. Каплан А.А. Об одном подходе к решению задач выпуклого программирования // Докл. АН СССР. - 1981. - Т.258, № 4. - С. 785-788.
2. Каплан А.А. Алгоритмы выпуклого программирования, использующие сглаживание точных функций штрафа // Сиб. мат. журн.- 1982. - Т.23, № 4. - С.53-64.



Поступила в ред.-изд. отдел
04.06.1987 г.


*) Для $H = R^n$ см. [2, теорема 3 с учетом замечания 5].