ЛЕКЦИЯ 13
Глава 4. МЕТОДЫ ДЛЯ ЗАДАЧ УСЛОВНОЙ ОПТИМИЗАЦИИ И КОМПЛЕМЕНТАРНЫХ ЗАДАЧ
(продолжение)

Содержание лекции

1. Штрафы и модифицированные функции Лагранжа (завершение)
   - Модифицированные функции Лагранжа
   - Точные штрафы

Задача оптимизации со смешанными ограничениями

Задача

$$f(x) \to \min, \quad x \in D, \tag{1}$$

$$D = \{x \in \mathbb{R}^n \mid F(x) = 0,\ G(x) \leqslant 0\}, \tag{2}$$

где

- $f : \mathbb{R}^n \to \mathbb{R}$ — заданная гладкая функция;
- $F : \mathbb{R}^n \to \mathbb{R}^l$ и $G : \mathbb{R}^n \to \mathbb{R}^m$ — заданные гладкие отображения.

Сведение к ограничениям-равенствам через «слэки»

$$f(x) \to \min, \quad (x,\, \sigma) \in \tilde{D}, \tag{3}$$

$$\tilde{D} = \{(x,\, \sigma) \in \mathbb{R}^n \times \mathbb{R}^m \mid F(x) = 0,\ g_i(x) + \sigma_i^2 = 0,\ i = 1,\, \ldots,\, m\}. \tag{4}$$

Модифицированные функции Лагранжа

Смысл введения модифицированных функции Лагранжа для задач с ограничениями-равенствами — обеспечение положительной определенности их матриц Гессе в естественных предположениях (при сохранении ключевых свойств функции Лагранжа).

Функция Лагранжа задачи (3), (4) (с ограничениями-равенствами)

$$\tilde{L} : (\mathbb{R}^n \times \mathbb{R}^m) \times (\mathbb{R}^l \times \mathbb{R}^m) \to \mathbb{R}, \quad \tilde{L}((x, \sigma), (\lambda, \mu)) =$$

$$= f(x) + \langle \lambda, F(x) \rangle + \sum_{i=1}^{m} \mu_i (g_i(x) + \sigma_i^2) = L(x, \lambda, \mu) + \sum_{i=1}^{m} \mu_i \sigma_i^2 .$$

Модифицированные функции Лагранжа

Семейство модифицированных функций Лагранжа задачи (3), (4)

$$\tilde{L}_c : (\mathbb{R}^n \times \mathbb{R}^m) \times (\mathbb{R}^l \times \mathbb{R}^m) \to \mathbb{R}, \quad \tilde{L}_c((x, \sigma), (\lambda, \mu)) =$$

$$= \tilde{L}((x, \sigma), (\lambda, \mu)) + \frac{c}{2}\left(\|F(x)\|^2 + \sum_{i=1}^{m} (g_i(x) + \sigma_i^2)^2 \right).$$

Подзадача метода модифицированных функций Лагранжа для задачи (3), (4)

$$\tilde{L}_c((x, \sigma), (\lambda, \mu)) \to \min, \quad (x, \sigma) \in \mathbb{R}^n \times \mathbb{R}^m \qquad (12)$$

(для заданных $\lambda \in \mathbb{R}^l$ и $\mu \in \mathbb{R}^m$).

Модифицированные функции Лагранжа

Глобальный минимум по $\sigma \in \mathbb{R}^m$ в (12) достигается в (единственной с точностью до знаков компонент) точке $\sigma(x, \mu, c) \in \mathbb{R}^m$,

$$\sigma_i(x, \mu, c) = \begin{cases} 0, \text{ если } g_i(x) + \mu_i/c > 0, \\ \sqrt{-(g_i(x) + \mu_i/c)}, \text{ иначе}, \end{cases} \qquad i = 1, \ldots, m.$$

Поэтому

$$g_i(x) + (\sigma_i(x, \mu, c))^2 = g_i(x) + \max\{0, -g_i(x) - \mu_i/c\} =$$

$$= \max\{g_i(x), -\mu_i/c\} \quad \forall\, i = 1, \ldots, m. \qquad (13)$$

Модифицированные функции Лагранжа

Это приводит к следующему понятию:

Модифицированная функция Лагранжа задачи (1), (2)

$$L_c : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m \to \mathbb{R}, \quad L_c(x, \lambda, \mu) = \min_{\sigma \in \mathbb{R}^m} \tilde{L}_c((x, \sigma), (\lambda, \mu)) =$$

$$= \tilde{L}_c((x, \sigma(x, \mu, c)), (\lambda, \mu)) = f(x) + \langle \lambda, F(x) \rangle + \frac{c}{2}\|F(x)\|^2 +$$

$$+ \frac{1}{2c}\sum_{i=1}^{m}((\max\{0, cg_i(x) + \mu_i\})^2 - \mu_i^2). \tag{14}$$

Эта функция дифференцируема, но, вообще говоря, только один раз.

Метод модифицированных функций Лагранжа

Метод модифицированных функций Лагранжа

Состоит в последовательном решении задач безусловной оптимизации

$$L_c(x, \lambda, \mu) \to \min, \quad x \in \mathbb{R}^n, \tag{15}$$

сопровождающемся пересчетом значения λ и μ.

Метод модифицированных функций Лагранжа

Алгоритм 2

Выбираем монотонно неубывающую последовательность $\{c_k\} \subset \mathbb{R}_+ \setminus \{0\}$ и точки $\lambda^0 \in \mathbb{R}^l$ и $\mu^0 \in \mathbb{R}^m$, и полагаем $k = 0$.

1. Вычисляем $x^{k+1} \in \mathbb{R}^n$ как стационарную точку задачи (15) с целевой функцией, задаваемой формулой (14) при $c = c_k$, $\lambda = \lambda^k$ и $\mu = \mu^k$.
2. Полагаем
$$\lambda^{k+1} = \lambda^k + c_k F(x^{k+1}), \tag{16}$$
$$\mu^{k+1} = \max\{0,\, \mu^k + c_k G(x^{k+1})\}. \tag{17}$$
3. Увеличиваем номер шага k на 1 и переходим к п. 1.

Метод модифицированных функций Лагранжа

Происхождение формулы (17) для пересчета множителей: для задачи (3), (4) (с ограничениями-равенствами) должно быть

$$\mu_i^{k+1} = \mu_i^k + c_k(g_i(x^{k+1}) + (\sigma_i(x^{k+1},\, \mu^k,\, c_k))^2), \quad i = 1,\, \ldots,\, m,$$

а это и есть (17) (в силу (13)).

Как и в случае ограничений-равенств, при этом

$$\frac{\partial L}{\partial x}(x^{k+1},\, \lambda^{k+1},\, \mu^{k+1}) = f'(x^{k+1}) + (F'(x^{k+1}))^\top(\lambda^k + c_k F(x^{k+1})) +$$

$$+ (G'(x^{k+1}))^\top(\max\{0,\, \mu^k + c_k G(x^{k+1})\}) = \frac{\partial L_{c_k}}{\partial x}(x^{k+1},\, \lambda^k,\, \mu^k) = 0. \tag{18}$$

Кроме того,

$$\mu^{k+1} \geqslant 0. \tag{19}$$

Глобальная сходимость

Теорема 4

Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ и отображения $F : \mathbb{R}^n \to \mathbb{R}^l$ и $G : \mathbb{R}^n \to \mathbb{R}^m$ непрерывно дифференцируемы на $\mathbb{R}^n$.
Тогда если последовательность $\{(x^k, \lambda^k, \mu^k)\}$ сгенерирована алгоритмом 2, последовательность $\{x^k\}$ сходится к точке $\bar{x} \in \mathbb{R}^n$, которая допустима в задаче (1), (2), а последовательность $\{(\lambda^k, \mu^k)\}$ имеет предельную точку $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^l \times \mathbb{R}^m$, то $\bar{x}$ — стационарная точка задачи (1), (2), а $\bar{\lambda}$ и $\bar{\mu}$ — отвечающие ей множители Лагранжа.

Для версий алгоритма 2 с более конкретными правилами управления параметром штрафа можно получить более тонкие теоремы о глобальной сходимости: характеризующие предельные точки $\{x^k\}$ (без предположений о ее сходимости), в том числе и недопустимые.

Глобальная сходимость

Доказательство

Из выполнения для любого k (18) и (19) следует, что

$$\frac{\partial L}{\partial x}(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0, \quad \bar{\mu} \geqslant 0,$$

и остается доказать выполнение условия дополняющей нежесткости.

Для любого $i \in \{1, \ldots, m\} \setminus A(\bar{x})$, в силу положительности и монотонного неубывания значений c_k найдется $\delta > 0$ такое, что для любого достаточно большого k выполняется $c_k g_i(x^k) < -\delta$. Поэтому, согласно (17), либо $\mu_i^{k+1} = 0$, либо $\mu_i^{k+1} \leqslant \mu_i^k - \delta$, откуда следует, что $\mu_i^k = 0$ для любого достаточно большого k (идентификация неактивных ограничений). Значит, $\bar{\mu}_i = 0$, что и требовалось.

Глобальная сходимость

Полностью избежать возможности существования недопустимых предельных точек нельзя (даже при отсутствии ограничений-неравенств).

Пример

При $n = l = 1$, $m = 0$, $f \equiv 0$, $F(x) = x^2 + 1$ задача (1), (2) не имеет допустимых точек.
При этом

$$L_c(x, \lambda) = \lambda(x^2 + 1) + \frac{c}{2}(x^2 + 1)^2,$$

$$\frac{\partial L_c}{\partial x}(x, \lambda) = 2x(\lambda + c(x^2 + 1)),$$

и при любых λ и c точка $\bar{x} = 0$ является стационарной в задаче (16), (17).

Локальная сходимость и скорость сходимости

Результат о локальной сходимости и скорости сходимости можно получить немедленно из теоремы 4.2.5, используя сведение к задаче с ограничениями-равенствами (3), (4), но при этом приходится требовать выполнение условия строгой дополнительности (ср. с теоремой 3 для метода квадратичного штрафа).
Непосредственный анализ для исходной задачи (1), (2) позволяет получить следующий более тонкий результат (ср. с теоремой 4.3.2 о локальной сверхлинейной сходимости метода последовательного квадратичного программирования).

Локальная сходимость и скорость сходимости

Теорема 5

Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ и отображения $F : \mathbb{R}^n \to \mathbb{R}^l$ и $G : \mathbb{R}^n \to \mathbb{R}^m$ дважды дифференцируемы в некоторой окрестности точки $\bar{x} \in \mathbb{R}^n$, и их вторые производные непрерывны в этой точке. Пусть $\bar{x}$ — стационарная точка задачи (1), (2), а $\bar{\lambda} \in \mathbb{R}^l$ и $\bar{\mu} \in \mathbb{R}^m$ — отвечающие ей множители Лагранжа, причем в точке $\bar{x}$ выполнены строгое условие Мангасариана–Фромовица и достаточное условие второго порядка оптимальности (11).

Локальная сходимость и скорость сходимости

Теорема 5 (завершение)

Тогда найдутся $\bar{c} > 0$ и $\delta > 0$ такие, что для любого $c_0 \geqslant \bar{c}$ и любого начального приближения $(x^0, \lambda^0, \mu^0) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$, достаточно близкого к $(\bar{x}, \bar{\lambda}, \bar{\mu})$, существует последовательность $\{(x^k, \lambda^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^m$ такая, что для всякого $k = 0, 1, \ldots$ точка x^{k+1} является стационарной в задаче (14) при $c = c_k$, $\lambda = \lambda^k$ и $\mu = \mu^k$, и выполнены соотношения (16), (17) и

$$\|(x^{k+1} - x^k, \lambda^{k+1} - \lambda^k, \mu^{k+1} - \mu^k)\| \leqslant \delta; \tag{20}$$

любая такая последовательность сходится к $(\bar{x}, \bar{\lambda}, \bar{\mu})$ с линейной скоростью; скорость сходимости сверхлинейная, если $c_k \to +\infty$.

Штрафы

Семейство штрафных функций

$$\varphi_c : \mathbb{R}^n \to \mathbb{R}, \quad \varphi_c(x) = f(x) + c\psi(x). \tag{6}$$

Степенной штраф

$$\psi(x) = \|\Psi(x)\|^p, \quad x \in \mathbb{R}^n, \tag{8}$$

где

$$\Psi : \mathbb{R}^n \to \mathbb{R}^l \times \mathbb{R}^m, \quad \Psi(x) = (F(x), \max\{0, G(x)\}).$$

Точные штрафы (т.е. не предполагающие бесконечного увеличения параметра штрафа) — одно из важнейших средств современной численной оптимизации.

Скорость сходимости по функции

Теорема 6

Пусть функция $f : \mathbb{R}^n \to \mathbb{R}$ липшицева в окрестности точки $\bar{x} \in \mathbb{R}^n$, а отображения $F : \mathbb{R}^n \to \mathbb{R}^l$ и $G : \mathbb{R}^n \to \mathbb{R}^m$ дифференцируемы в этой окрестности, причем их производные непрерывны в точке $\bar{x}$. Пусть $\bar{x}$ — локальное решение задачи (1), (2), причем в точке $\bar{x}$ выполнено условие регулярности Мангасариана–Фромовица.

Скорость сходимости по функции

Теорема 6 (продолжение)

Тогда найдутся числа $M_1 > 0$ и $M_2 > 0$ такие, что если функция φ_{c_k} задана формулами (6), (8) при некотором $p > 0$, и для последовательностей $\{c_k\} \subset \mathbb{R}_+$ и $\{x^k\} \subset \mathbb{R}^n$ имеют место соотношения

$$\varphi_{c_k}(x^k) \leqslant \varphi_{c_k}(\bar{x}) = f(\bar{x}) \quad \forall\, k, \tag{21}$$

$$\{c_k\} \to \infty, \quad \{x^k\} \to \bar{x} \quad (k \to \infty), \tag{22}$$

то для любого достаточно большого k справедливо следующее:

Скорость сходимости по функции

Теорема 6 (продолжение)

- если $p > 1$, то
$$0 \leqslant f(\bar{x}) - f(x^k) \leqslant \frac{M_1}{c_k^{1/(p-1)}}, \quad \operatorname{dist}(x^k, D) \leqslant \frac{M_2}{c_k^{1/(p-1)}};$$
- если $p \in (0, 1]$, то
$$f(x^k) = f(\bar{x}), \quad x^k \in D;$$
  в частности, если $\bar{x}$ является строгим локальным решением задачи (1), (2), то $x^k = \bar{x}$.

Скорость сходимости по функции

Доказательство

Ключевой ингредиент — оценка расстояния до допустимого множества

$$\mathrm{dist}(x,\, D) = O(\|\Psi(x)\|)$$

при $x \in \mathbb{R}^n$ стремящемся к $\bar{x}$, справедливая при выполнении условия регулярности Мангасариана–Фромовица.

Свойство (21) имеет место, если x^k является глобальным минимумом функции φ_{c_k} на некотором множестве, содержащем $\bar{x}$. Типично для методов штрафа.
Второе соотношение в (22) — сходимость по аргументу.
Оба эти свойства выполняются, если $\bar{x}$ — строгое локальное решение задачи (1), а $x^k = x(c_k)$ определяется как в теореме 1 о «локальной» сходимости.

Точные штрафы

Следствие 1

Пусть в условиях теоремы 6 точка $\bar{x}$ — строгое локальное решение задачи (1), (2), и пусть $p \in (0, 1]$.
Тогда существует число $\bar{c} \geqslant 0$ такое, что для каждого $c > \bar{c}$ точка $\bar{x}$ является строгим локальным решением задачи

$$\varphi_c(x) \to \min, \quad x \in \mathbb{R}^n, \tag{7}$$

с целевой функцией, задаваемой формулой (6).

Уменьшение p приводит к увеличению скорости сходимости.
При достаточно малом p (а именно, при $p \in (0, 1]$) степенной штраф является *точным*: искомое (строгое) локальное решение $\bar{x}$ задачи (1), (2) является локальным решением подзадачи безусловной оптимизации (7) при любом $c > \bar{c}$ (т.е. нет необходимости бесконечно увеличивать c).

Точные штрафы

Уменьшение p негативно влияет на гладкость φ_c: при $p \in (0, 1]$ вводимый в (8) степенной штраф не является гладким.
Другая трудность: $\bar{c}$ неизвестно.

Негладкие точные степенные штрафы обычно используются не непосредственно, а в сочетании с другими методами, в целях глобализации сходимости последних.