ПРЯМЫЕ УСЛОВИЯ ОПТИМАЛЬНОСТИ В ЗАДАЧАХ МАТЕМАТИЧЕСКОГО ПРОГРАММИРОВАНИЯ

По исторически сложившимся причинам *задачей математического программирования* принято называть задачу поиска в E^n максимальных элементов функционала $F(x)$, на множестве элементов x удовлетворяющих условиям вида:

$$f_i(x) \geq 0 \text{ , } i = [1, m] \text{ ,}$$

или, максимизировать $F(x)$ при условиях:

$$\begin{cases} f_1(x) \geq 0, \\ f_2(x) \geq 0, \\ \quad \ldots \\ f_m(x) \geq 0. \end{cases}$$

Соответственно в координатной форме эта задача записывается:

Найти максимум $F(\xi_1, \xi_2, \ldots, \xi_n)$ по $\{\xi_1, \xi_2, \ldots, \xi_n\}$,

при условиях: (3.5.1)

$$\begin{cases} f_1(\xi_1, \xi_2, \ldots, \xi_n) \geq 0, \\ f_2(\xi_1, \xi_2, \ldots, \xi_n) \geq 0, \\ \qquad \ldots \\ f_m(\xi_1, \xi_2, \ldots, \xi_n) \geq 0. \end{cases}$$

Необходимые условия разрешимости задачи математического программирования

Общая формулировка необходимых условий такова:

> *пересечение конуса допустимых вариаций и конуса улучшающих вариаций в* E^n *для рассматриваемой задачисостоит только из нулевого элемента.*

Напомним вначале

Определение	Ограничение $f_i(x) \geq 0$ называется *активным* на элементе x^*, если $f_i(x^*) \leq 0$.

Рассмотрим конкретный случай.

Пусть функционалы $\{ F(x), f_i(x), i = [1, m] \}$ непрерывно дифференцируемы на элементе $x^* \in E^n$, и $J(x^*)$ – множество индексов активных на элементе x^* ограничений, набор элементов $\{\operatorname{grad} f_i(x^*), i \in J(x^*)\}$ линейно независим и, наконец, $\delta x = x - x^*$, тогда оказывается справедливой

Теорема **Пусть функционал $F(x)$ на множестве R принимает максимальное значение на элементе x^*. Тогда необходимое условие оптимальности имеет вид**

$$(\operatorname{grad} F(x^*), \delta x) \le 0 \quad \text{и} \quad \forall \delta x \in M(x^*) :$$

$$(\operatorname{grad} f_i(x^*), \delta x) \ge 0, \ \ i \in J(x^*).$$

В отличие от учебных примеров, задача практического отыскания точек 'подозрительных на экстремум' оказывается существенно сложнее.

Поясним это следующим примером.

Пусть в некоторой акватории с помощью только глубинометра требуется найти наиболее глубокую точку дна.

Будем считать, что дно достаточно гладкое, глубинометр соответственно достаточно чувствительный. Каков должен быть алгоритм поиска?

Приведем пример использования схемы *градиентного подъема* с непрерывным изменением величины "шага по направлению" (*метод Коши*). Формально поиск локального максимального элемента для функционала $F(x)$ в этом случае может быть сведен к решению следующей задачи Коши: $\dfrac{dx}{d\tau} = \operatorname{grad} F(x);\ x(0) = x^0$.

Например, в E^2 для $F(\xi_1, \xi_2) = -2\xi_1^2 - \xi_2^2$ с $\| x^0 \| = \left\| \begin{matrix} 1 \\ 1 \end{matrix} \right\|$ имеем $\begin{cases} \dot{\xi}_1 = -4\xi_1, \\ \dot{\xi}_2 = -2\xi_2, \end{cases}$ что дает $\begin{cases} \xi_1(\tau) = e^{-4\tau}, \\ \xi_2(\tau) = e^{-2\tau} \end{cases}$ с фазовой траекторией, задаваемой в силу условий $\xi_1(0) = 1$; $\xi_2(0) = 1$ уравнением $\xi_1 - \xi_2^2 = 0$, которая при $\tau = 0$ выходит из начального элемента x_0 и при $\tau \to +\infty$ асимптотически стремится к максимальному элементу $\| x^* \| = \left\| \begin{matrix} 0 \\ 0 \end{matrix} \right\|$.

Общая схема поиска локального экстремума

Вполне очевидно, что метод поиска в E^n локальных экстремумов функционала $F(\xi_1, \xi_2, \dots, \xi_n)$, основанный на исследовании критических его элементов (или же решений уравнения $\operatorname{grad} F(x) = o$), практически пригоден лишь для крайне ограниченного числа случаев.

Реально реализуемой альтернативой является использование так называемых *итеративных численных методов*. Все эти методы также основаны на необходимых или достаточных условиях экстремума функционала в E^n и могут быть представлены в виде следующей схемы поиска, например, максимума:

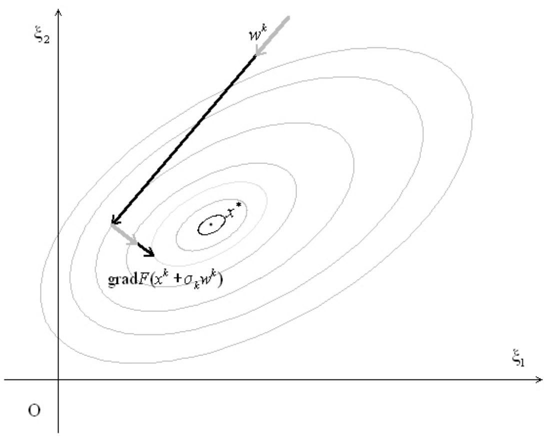
1°. Для некоторого начального элемента x^0 находятся ненулевое *улучшающее направление максимизации* $w^0 \in E^n$ и положительное число $\sigma_0 < +\infty$ – *величина шага* по данному направлению – такие, что на элементе $x^1 = x^0 + \sigma_0 \cdot w^0$ верно неравенство $F(x^1) > F(x^0)$.

2°. Если задача в пункте 1° решена успешно, то элемент x^0 заменяется на x^1 и процедура пункта 1° повторяется для некоторого нового $w^1 \in E^n$. Если же оказалось, что множество улучшающих направлений состоит только из нулевого элемента или же $\sigma_0 = 0$, то элемент x^0 принимается за x^* – максимальный, либо, при $\sigma_0 = +\infty$, констатируется факт отсутствия максимума у $F(x)$.

Таким образом, поиск экстремального элемента сводится к итерационной процедуре вида

$$x^{k+1} = x^k + \sigma_k \cdot w^k, \quad k = 0,1,2,\ldots ,$$

условия сходимости которой к искомому экстремальному элементу x^* (то есть, $\rho(x^k, x^*) \to 0$ при $k \to \infty$) зависят, как от способа выбора w^k и σ_k, так и от свойств максимизируемого функционала.



Методы поиска локального гладкого экстремума

Предположим, что функционал $F(x)$ непрерывно дифференцируем в E^n и выполнены следующие условия:

1°. Начальный элемент x^0 принадлежит окрестности максимального элемента x^*, в которой функционал $F(x)$ *строго выпуклый вверх*;

2°. Элемент *направления максимизации* w^k удовлетворяет ограничению

$$(\operatorname{grad} F(x^k), w^k) \geq \alpha \cdot \left| \operatorname{grad} F(x^k) \right|^2,$$

где α – некоторое фиксированное малое положительное число;

3°. Величина шага по выбранному направлению максимизации σ_k находится путем решения одномерной задачи

$$\sigma_k = \max_{\sigma \geq 0} F(x^k + \sigma \cdot w^k).$$

В этом случае последовательность элементов $\{x^k\}$ будет сходиться к x^*.

Заметим, что сформулированное в п. 3° правило выбора шага эффективно для *непрерывно дифференцируемых* функционалов в E^n. Оптимальная величина шага σ_k по направлению w^k в этом случае может быть найдена из уравнения

$$(\operatorname{grad} F(x^k + \sigma_k w^k), w^k) = 0.$$

Это правило гарантирует наибольшее локальное увеличение функционала по направлению w^k, что иллюстрирует рисунок 2.2.2.1.

В случаях, когда оптимизируемый функционал более гладкий (например, имеет непрерывные частные производные первого и второго порядка), для выбора w_k можно использовать другие, более эффективные схемы, такие как:

1) *метод наискорейшего подъема*, основанный на оценке скорости возрастания функционала $F(x)$ на элементе x^k по направлению w^k, которая в силу неравенства Коши–Буняковского максимальна при $w^k = \operatorname{grad} F(x^k)$, поскольку $\forall w$ с $|w| = 1$

$$\left| \frac{\partial F}{\partial w} \right| = \left| (\operatorname{grad} F, w) \right| \le \left| \operatorname{grad} F \right| \left| w \right| = \left| \operatorname{grad} F \right| .$$

Таким образом, направление "наискорейшего" *возрастания* функционала $F(x)$ на некотором фиксированном элементе x сонаправлено с элементом $\operatorname{grad} F(x)$, а направление "наискорейшего" *убывания* с элементом $-\operatorname{grad} F(x)$.

2) *метод квадратичной аппроксимации* (*метод Ньютона*), при котором улучшающее направление w_k удовлетворяет условию

$$\hat{\mathrm{Hess}}\, F(x_k) w^k = -\mathrm{grad}\, F(x_k)\,,$$

что в матричном и координатном представлениях есть система линейных уравнений

$$\left\| \hat{\mathrm{Hess}}\, F(x_k) \right\| \left\| w^k \right\| = -\left\| \mathrm{grad}\, F(x_k) \right\| \text{ или}$$

$$\sum_{j=1}^{n} \frac{\partial^2 F}{\partial \xi_i \partial \xi_j} \omega_j^k = -\frac{\partial F}{\partial \xi_i}\,,\ i = [1, n]\,,$$

где $\left\| w^k \right\| = \left\| \begin{matrix} \omega_1^k \\ \omega_2^k \\ \dots \\ \omega_n^k \end{matrix} \right\|$ – координатное представление элемента w^k.

Убедимся, что для квадратичного функционала $F(x)$ стационарный (то есть такой, что $\operatorname{grad} F(x) = o$) элемент находится по методу Ньютона за одну итерацию с $\sigma = 1$.

Действительно, в этом случае стационарный элемент существует и единственный. Разложение $F(x)$ по формуле Тейлора в окрестности любого x_0 при произвольном элементе вариации $dx = w$ (то есть отклонении от x_0) записывается *без остаточного члена*

$$F(x_0 + w) = F(x_0) + \sum_{j=1}^{n} \frac{\partial F}{\partial \xi_j} \omega_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{i=1}^{n} \frac{\partial^2 F}{\partial \xi_j \partial \xi_i} \omega_j \omega_i .$$

Тогда условие стационарности $F(x)$ будет иметь вид

$$\frac{\partial F}{\partial \omega_i} = \frac{\partial F}{\partial \xi_i} + \sum_{j=1}^{n} \frac{\partial^2 F}{\partial \xi_j \partial \xi_i} \omega_j = 0 \quad ; \quad i = [1, n],$$

что и доказывает проверяемое утверждение. Тип стационарного элемента (максимум, минимум или "седло") в этом случае зависит от знаковой определенности $\left\| \hat{\operatorname{Hess}} F(x) \right\|$.