# 20. Нелинейно оптимиране

Общата задача на нелинейното оптимиране се определя като задача за намирането на максимума (минимума) на целевата функция $f(x_1, x_2, \dots, x_n)$ върху множеството $D$, определено от ограниченията

$$D = \left\{ x \in \mathbb{R}^n \;\middle|\; \begin{array}{l} g_i(x_1, x_2, \dots, x_n) \le 0, \; i = 1, 2, \dots, r \\ g_i(x_1, x_2, \dots, x_n) = 0, \; i = r+1, \dots, m \end{array} \right\}$$

Очевидно е, че типът оптимизация не е определящ; затова занапред ще разглеждаме задачи за max. Както и в ЗЛО, векторът $x^* = (x_1^*, x_2^*, \dots, x_n^*) \in D$ ще наричаме допустим план; а в случай, че за $\forall x \in D$ е изпълнено $f(x^*) \ge f(x)$ – $x^*$ е оптимален план.

Задачата за нелинейното оптимиране е толкова обща, че за нея не съществува общ метод за решаване.

## 20.1. Решение на задачата за условна оптимизация по метода на Лагранж.

Ясно е, че с въвеждане на фиктивни променливи (както в ЗЛО) множеството $D$ може да бъде описано само чрез уравнения.

Методът на Лагранж (известен от курса по математически анализ) се състои в свеждането на задачата за условен екстремум

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &\to \max \\ g_1(x_1, x_2, \dots, x_n) &= 0 \\ g_2(x_1, x_2, \dots, x_n) &= 0 \\ &\dots \\ g_m(x_1, x_2, \dots, x_n) &= 0 \end{aligned}$$

към задачата за безусловна оптимизация на функцията на Лагранж

$$\mathcal{L}(x_1, x_2, \dots, x_n; \lambda_1, \lambda_2, \dots, \lambda_m) = f(x_1, x_2, \dots, x_n) + \sum_{i=1}^{m} \lambda_i g_i(x_1, \dots, x_n),$$

където $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathbb{R}^m$ е вектор, състоящ се от та–

ка наречените множители на Лагранж.
В случай на гладкост (диференцируемост) на функциите $f(x)$ и $g_i(x)$, $i=1,2,\dots,m$ е справедлива следната
Теорема 1. (Необходимо условие за условен екстремум). Ако $x^*=(x_1^*, x_2^*, \dots, x_n^*)$ е точка на условен екстремум и рангът на матрицата от първите частни производни

$$\left\| \frac{\partial g_i(x^*)}{\partial x_j} \right\|_{m\times n}$$

е равен на $m$, то съществуват такива $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*$, не равни едновременно на нули при които

$$\nabla \Phi(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) = 0$$

Забл.: С $\nabla \Phi$, $\nabla f$ сме означили градиентите на $\Phi$ и $f$ съответно, т.е. горното равенство ако го разпишем, добива вида:

$$\frac{\partial \Phi}{\partial x_1} = \frac{\partial f}{\partial x_1} + \sum_{i=1}^{m} \lambda_i^* \frac{\partial g_i}{\partial x_1} = 0$$

$$\frac{\partial \Phi}{\partial x_2} = \frac{\partial f}{\partial x_2} + \sum_{i=1}^{m} \lambda_i^* \frac{\partial g_i}{\partial x_2} = 0$$

$$\dots\dots\dots\dots\dots$$

$$\frac{\partial \Phi}{\partial x_n} = \frac{\partial f}{\partial x_n} + \sum_{i=1}^{m} \lambda_i^* \frac{\partial g_i}{\partial x_m} = 0$$

От Теорема 1 произлиза метод за търсене на условен екстремум (метод на Лагранж). Той се състои от следните стъпки:
1. Съставяне на функцията на Лагранж $\Phi(x,\lambda)$;
2. Намиране на частните производни от първи ред по всички променливи, т.е.

$\frac{\partial\phi(x,\lambda)}{\partial x_j}$ $(j=1,2,\dots,n)$ и $\frac{\partial\phi(x,\lambda)}{\partial\lambda_i}$ $(i=1,2,\dots,m)$

3. Решаване на системата от уравнения

$$\begin{cases} \frac{\partial\phi(x,\lambda)}{\partial x_j}=0, & (j=1,2,\dots,n) \\ \frac{\partial\phi(x,\lambda)}{\partial\lambda_i}=g_i(x)=0, & (i=1,2,\dots,m) \end{cases}$$

($n+m$ на брой уравнения за $n+m$ на брой неизвестни $x_1,x_2,\dots,x_n;\lambda_1,\lambda_2,\dots,\lambda_m$) и намиране на критичните стойности $x_1^*,x_2^*,\dots,x_n^*;\lambda_1^*,\lambda_2^*,\dots,\lambda_m^*$.

4. Изследване на тези критични стойности с помощта на някакво достатъчно условие за екстремум.

Заб.: Присъствието на последния етап се обяснява с това, че Теорема 1 е необходимо (но не и достатъчно) условие за наличието на условен екстремум. Такива достатъчни условия съществуват, но обхващат тесни класове функции.

20.2. Оптимизационни задачи за изпъкнали функции.

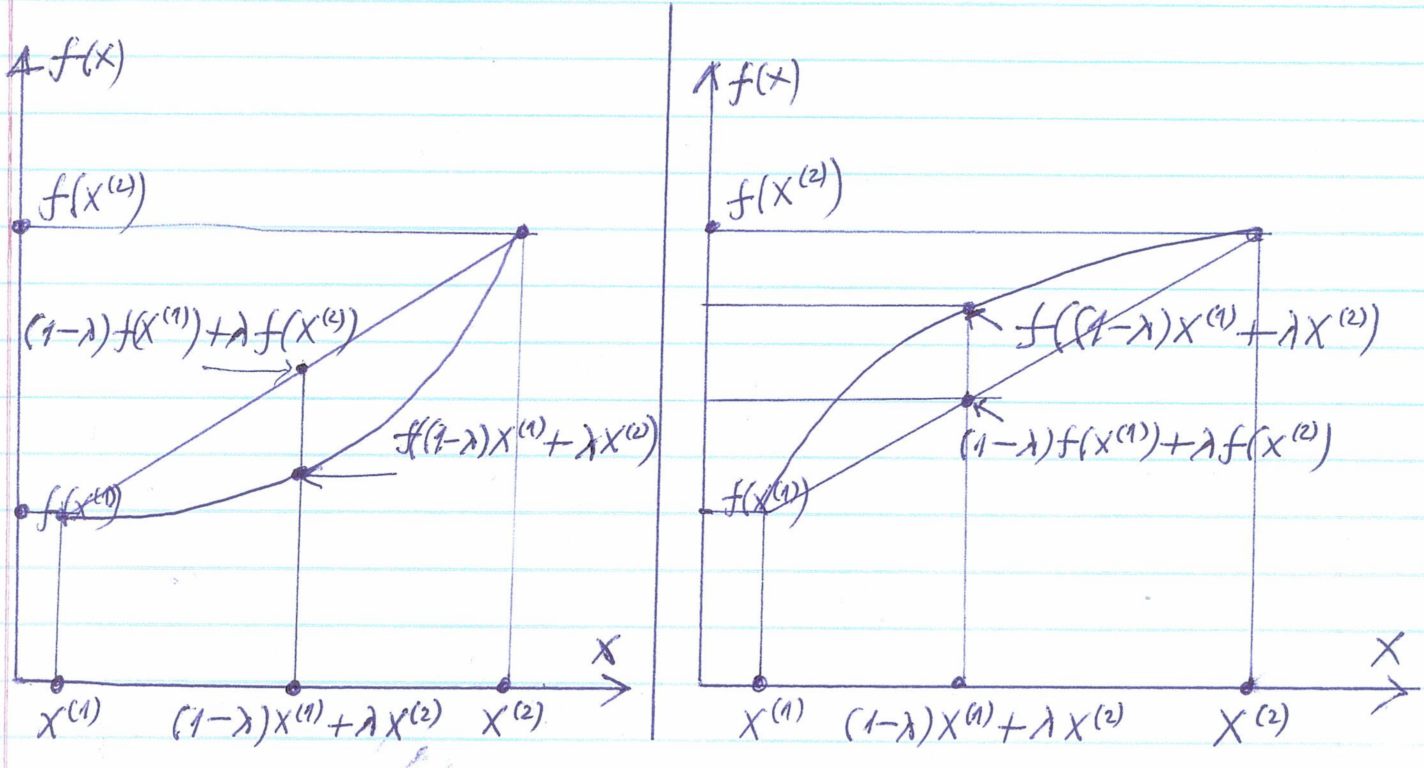
Определение: Функцията $f(x)=f(x_1,x_2,\dots,x_n)$ се нарича изпъкнала (вдлъбнатата) в областта $D$, ако за $\forall\lambda\in[0,1]$ е изпълнено неравенството

$$f((1-\lambda)x^{(1)}+\lambda x^{(2)})\le(1-\lambda)f(x^{(1)})+\lambda f(x^{(2)})$$

$$\left(f((1-\lambda)x^{(1)}+\lambda x^{(2)})\ge(1-\lambda)f(x^{(1)})+\lambda f(x^{(2)})\right),$$

където $x^{(1)},x^{(2)}\in D$ (произволни).

Геометричния смисъл на понятието изпъкналост (вдлъбнатост) може да се види на чертежа по-долу. Вижда се, че графиката на изпъкналата (вдлъбнатата) функция лежи под (над) правата линия, съединяваща точките $(x^{(1)},f(x^{(1)}))$ и $(x^{(2)},f(x^{(2)}))$.




Може да се докаже, че достатъчно условие за изпъкналост на функцията $f(x)=f(x_1,x_2,\dots,x_n)$ е положителната дефинитност на хесиана (матрицата на Хесе)

$$H(f)=\left\|\frac{\partial^2 f(x)}{\partial x_i\,\partial x_j}\right\|_{n\times n} \quad \text{за } \forall x\in D$$

Съответно, достатъчното условие за вдлъбнатост на $f(x)$ е отрицателната определеност на хесиана $H(f)$.

От друга страна, от курса по линейна алгебра е известен критерият на Силвестър (за положителна и отрицателна дефинитност на произволна квадратна матрица). Образуваме всички главни минори на матрицата, в случай на хесиан те са:

главен минор от ред 1 : $\dfrac{\partial^2 f}{\partial x_1^2}$

главен минор от ред 2 : $\begin{vmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1\partial x_2} \\ \dfrac{\partial f}{\partial x_2\partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} \end{vmatrix}$

главен минор
от ред 3:

$$\begin{vmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{vmatrix}$$

и т.н.
(С $|\cdot|$ е означена детерминанта на квадратна матрица)
Критерият на Силвестър гласи, че
1. Една матрица е положително дефинитна $\Leftrightarrow$ всички главни минори са положителни (в случай на хесиан – изпъкнала функция);
2. Една матрица е отрицателно дефинитна $\Leftrightarrow$ ~~ал~~ главните минори от нечетен ред са отрицателни, а тези от четен ред – положителни (в случай на хесиан – вдлъбната функция).
За изпъкналите (вдлъбнати) функции е справедливо следното достатъчно условие:
Теорема 2. Ако $f(x)$ е изпъкнала (вдлъбната) функция и $\nabla f(x^*)=0$, то $x^*$ е точка на глобален минимум (максимум).
Доказателство:
То ще се проведе в случай на вдлъбната функция, за противоположния случай то ще бъде аналогично с точност до знака.
Нека $x$ е произволна точка, различна от $x^*$. Тогава за $\forall \lambda \in (0;1]$ по силата на вдлъбнатостта ще бъде изпълнено неравенството

$$f((1-\lambda)x^*+\lambda x) \geq (1-\lambda)f(x^*)+\lambda f(x)$$

Изваждаме от двете страни на неравенството $f(x^*)$; тогава за лявата страна получаваме

$$f((1-\lambda)x^* + \lambda x) - f(x^*) = f(x^* - \lambda x^* + \lambda x) - f(x^*) =$$
$$= f(x^* + \lambda(x - x^*)) - f(x^*),$$

а за дясната

$$(1-\lambda)f(x^*) + \lambda f(x) - f(x^*) = \cancel{f(x^*)} - \lambda f(x^*) +$$
$$+ \lambda f(x) - \cancel{f(x^*)} = \lambda(f(x) - f(x^*))$$

Така последното неравенство добива вида

$$f(x^* + \lambda(x - x^*)) - f(x^*) \geq \lambda(f(x) - f(x^*))$$

Ако въведем вектор $e = x - x^*$ и означим $\Delta x = \lambda(x - x^*) = \lambda e$, то дължината на вектора $\Delta x$ ще бъде $\|\Delta x\| = \lambda\|e\|$.
Следователно,

$$f(x) - f(x^*) \leq \frac{f(x^* + \Delta x) - f(x^*)}{\lambda} = \frac{f(x^* + \Delta x) - f(x^*)}{\|\Delta x\|}\|e\|$$

Нека $\lambda \to 0$. Тогава (отчитайки, че векторите $\Delta x$ и $e$ са колинеарни) ще имаме

$$f(x) - f(x^*) \leq \frac{\partial f(x^*)}{\partial e}\|e\|$$

По условие $\nabla f(x^*) = 0$. Това означава, че за всеки вектор $e$ (а това ще рече и за всяка точка $x$) съгласно формулата, изразяваща производната по направление чрез градиента

$$\frac{\partial f(x^*)}{\partial e} = Pr_e \nabla f(x^*) = 0$$

(с думи: производната по направление на вектор $e$ е проекцията върху $e$ на градиента $\nabla f(x^*) =$

$= \left(\frac{\partial f}{\partial x_1}(x^*), \frac{\partial f}{\partial x_2}(x^*), \ldots, \frac{\partial f}{\partial x_n}(x^*)\right).$

Така ние получихме, че за произволна т. $x \neq x^*$ е в сила $f(x) - f(x^*) \leq 0 \Rightarrow f(x) \leq f(x^*)$, т.е. $x^*$ е точка на глобален максимум.

Тъй като изпъкналите функции имат много „полезни" оптимизационни качества, те заемат изключително важно място в математическото оптимиране. Съответният раздел е получил името изпъкнало оптимиране и се състои в намирането на максимум на вдлъбната (минимум на изпъкнала) функция върху изпъкнало множество.

20.3. Двойственост в нелинейното оптимиране

Разглеждаме задачата

$$f(x) \longrightarrow max$$

$$D = \{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, 2, \ldots, m\}$$

Съставяме функцията на Лагранж

$$\Phi(x;\lambda) = f(x_1, x_2, \ldots, x_n) - \sum_{i=1}^{m} \lambda_i g_i(x_1, x_2, \ldots, x_n)$$

Двойката вектори $(\bar{x}; \bar{\lambda})$ се наричат седлова точка за функцията $\Phi(x;\lambda)$ ~~в някоква~~, ако за $\forall x$ и $\forall \lambda$ е изпълнено

$$\Phi(x;\bar{\lambda}) \leq \Phi(\bar{x};\bar{\lambda}) \leq \Phi(\bar{x};\lambda)$$

В качеството на пример за седлова точка може да се разгледа точката $(0,0)$ за функцията $\Phi(x;\lambda) \stackrel{!}{=} -x^2 + \lambda^2$, определена върху множеството $\mathbb{R} \times \mathbb{R}$. И наистина $\Phi(0;0) = 0$; $\Phi(x,0) = -x^2$, а $\Phi(0,\lambda) = \lambda^2$.

Теорема 3 (Кун – Тъкър). Ако $(\bar{x}; \bar{\lambda})$ е седлова точка за функцията на Лагранж при $\lambda \geq 0$, то $\bar{x}$ е оптимален план на задачата и е изпълнено условието

$$\sum_{i=1}^{m} \bar{\lambda}_i g_i(\bar{x}) = 0$$

Тази централна в нелинейното оптимиране теорема представлява достатъчно условие за условен екстремум.
Тъй като за $\forall i$ $\lambda_i \geq 0$ и $g_i(x) \leq 0$, то ще имаме

$$\bar{\lambda}_i g_i(\bar{x}) \leq 0$$

Тогава горната сума от $m$ на брой събираеми е нула, когато всички те са нули, т.е.:

$$\bar{\lambda}_1 g_1(\bar{x}) = \bar{\lambda}_2 g_2(\bar{x}) = \ldots = \bar{\lambda}_m g_m(\bar{x}) = 0$$

Това означава следното:
1. Ако за някое $i$ $\bar{\lambda}_i > 0$, то тогава $g_i(\bar{x}) = 0$;
2. Ако за някое $j$ $g_j(\bar{x}) < 0$, то $\bar{\lambda}_j = 0$.

Доказателство:
По определение за седлова точка ще имаме

$$f(x) - \sum_{i=1}^{m} \bar{\lambda}_i g_i(x) \leq f(\bar{x}) - \sum_{i=1}^{m} \bar{\lambda}_i g_i(\bar{x}) \leq f(\bar{x}) - \sum_{i=1}^{m} \lambda_i g_i(\bar{x}) \quad (*)$$

за $\forall x$ и $\lambda \geq 0$.
От второто неравенство следва, че

$$\sum_{i=1}^{m} \lambda_i g_i(\bar{x}) \leq \sum_{i=1}^{m} \bar{\lambda}_i g_i(\bar{x}) \quad \text{за } \lambda_i \geq 0$$

Обаче такова неравенство може да е изпълнено само при $g_i(\bar{x}) \geq 0$, $i = 1, 2, \ldots, m$. Действително, ако съществува такъв номер $k$, че да е изпълнено $g_k(\bar{x}) > 0$, то ако положим $\lambda_i = 0$ за $i \neq k$

и изберем достатъчно голямо $u_k > 0$, то изразът от лявата страна на горното неравенство

$$\sum_{i=1}^{m} \lambda_i g_i(\bar{x}) = \lambda_k g_k(\bar{x})$$

може да стане по-голям от фиксираната дясна страна на неравенството

$$\sum_{i=1}^{m} \bar{\lambda}_i g_i(\bar{x})$$

От това, че за $\forall i = 1, 2, \dots, m$ имаме $g_i(\bar{x}) \leq 0$ следва, че $\bar{x}$ е допустим план на задачата.
Ако в лявата страна на въпросното неравенство поставим $\lambda_1 = \lambda_2 = \dots = \lambda_m = 0$, то получаваме, че

$$\sum_{i=1}^{m} \bar{\lambda}_i g_i(\bar{x}) \geq 0$$

Обаче, от $\bar{\lambda}_i \geq 0$ и $g_i(\bar{x}) \leq 0$ следва, че

$$\sum_{i=1}^{m} \bar{\lambda}_i g_i(\bar{x}) \leq 0$$

Така получаваме, че е изпълнено равенството

$$\sum_{i=1}^{m} \bar{\lambda}_i g_i(\bar{x}) = 0$$

Като заместим това равенство в първото неравенство от (*) получаваме

$$f(\bar{x}) \geq f(x) - \sum_{i=1}^{m} \bar{\lambda}_i g_i(x)$$

За всеки допустим план $x$ са изпълнени неравенствата $g_i(x) \leq 0$. В съчетание с $\bar{\lambda}_i > 0$ можем да твърдим, че

$$\sum_{i=1}^{m} \bar{\lambda}_i g_i(x) \leq 0,$$

следователно

$$f(x) - \sum_{i=1}^{m} \bar{\lambda}_i g_i(x) \geq f(x)$$

Окончателно получаваме, че за всеки допустим план е в сила неравенството $f(\bar{x}) \geq f(x)$, следователно $\bar{x}$ е оптимален план.
Твърдение, обратно на Теорема 2, т.е. достатъчна условия за екстремум на общата задача на нелинейното оптимиране се оказват възможни само ако се наложат допълнителни условия. Най-важното от тези условия е условието на Слейтър.
Определение. Ограничителната функция $g_i(x)$ на задачата на нелинейното оптимиране е регулярна по Слейтър, ако съществува такъв допустим план $\tilde{x}$, че

$$g_i(\tilde{x}) < 0,$$

т.е. $\tilde{x}$ е вътрешна точка относно ограничението $g_i(x)$.
Съществуват много различни варианти на необходимо условие на Кун-Тъкър. Ще приведем едно от тях:
Теорема 4. Ако $(D, f)$ е задача на изпъкналото оптимиране с решение $\bar{x}$; целевата функция $f(x)$ и ограничителните функции $g_i(x)$, $i = 1, 2, \ldots, m$ са диференцируеми, нелинейните ограничения са регулярни по Слейтър, то съществува такъв вектор $\bar{\lambda} \geq 0$, че точката $(\bar{x}, \bar{\lambda})$ е седлова точка на функцията на Лагранж.
Няма да привеждаме доказателство, защото то е доволно сложно.
Значението на теоремата на Кун-Тъкър се състои в това, че тя позволява да се свърже оптимизационната задача с търсенето на сед-

лови точки на функцията на Лагранж, което означава максимализиране на тази функция по $x$ и минимализиране по $\lambda$.
Определяме $F(x)$ като функция, поставяща в съответствие на $\forall x$ минималното значение на $\phi(x;\lambda)$ по $\lambda$:

$$F(x) = \min_{\lambda \geqslant 0} \phi(x;\lambda)$$

и по аналогия

$$G(\lambda) = \max_{x \in D} \phi(x;\lambda)$$

Да разгледаме задачата за максимализиране на $F(x)$

$$F(x) = \min_{\lambda \geqslant 0} \phi(x;\lambda) \to \max \; ; x \in D$$

и задачата за минимализиране на $G(\lambda)$

$$G(\lambda) = \max_{x \in D} \phi(x;\lambda) \to \min \; ; \lambda \geqslant 0$$

Очевидно е, че $F(x) = \min_{\lambda \geqslant 0} \phi(x;\lambda) = f(x)$, така че задачата за максимализиране на $F(x)$ е равносилна на задачата за максимализиране на $f(x)$ при $x \in D$. Аналогични изводи могат да се направят и за $G(\lambda)$. Горните две задачи представляват дуална двойка. Това е едно обобщение на теорията на двойствеността в линейното оптимиране. Съответно, при определени условия дуалната двойка има свойства, аналогични с тези от линейното оптимиране. В частност, при произволни $x$ и $\lambda$ е изпълнено неравенството

$$F(x) \leq G(\lambda)$$