ТЕОРЕТИЧНИ ОСНОВИ НА НЕЛИНЕЙНОТО ПОГРАМИРАНЕ

9.1. БЕЗУСЛОВНА ОПТИМИЗАЦИЯ

Класическата теория на оптимизацията се основава на използу-ването на диференциалното смятане за определяне на екстремумите на функции при *отсьствие* или *наличие* на ограничения, т.е. при реша-ването на *безусловни* или *условни* екстремални задачи. Точката $X^* = [x^*_1, x^*_2 \; ... \; , x^*_n]^T$ е *точка на максимум* на функцията f(X), ако стойно-стите на f(X) във всяка точка от достатъчно близка околност на X* не превишават f(X*), т.е. ако условието

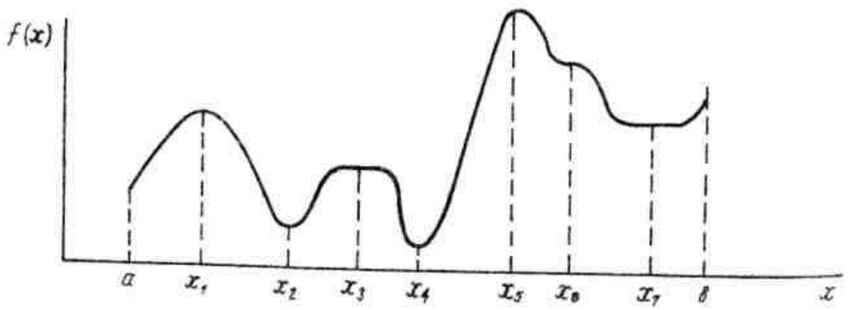
$$\mathbf{f(X^* + \Delta X) \leq f(X^*)}$$

се изпълнява за всички $\Delta X = [\Delta x_1, \Delta x_2, ..., \Delta x_n]^T$ такива, че $| x_i|$ са достатъчно малки, $i=\overline{1,n}$.

Подобно, X* е *точка на минимум* на f(X), ако

$$\mathbf{f(X^* + \Delta X) \geq f(X^*).}$$

На фиг.9.1 е показана функция от една променлива *f(x).* Тя е определена върху интервала $a \leq x \leq b$, който не се разглежда като ограничение на *f(x).* Точките x_1, x_3, x_5 са точки на *локални максимуми.* Тъй като



Фиг. 9.1

$f(x_5) = max\{f(x_1), f(x_3), f(x_5)\}$, точката x_5 е точка на *глобален мак-имум за* областта *[a,b].* Аналогично, x_2, x_4 *и* x_7 са точки на *локални минимуми,* а точката x_4, в която $f(x_4) = min\{f(x_2), f(x_4), f(x_7)\}$, *е* точка на *глобален минимум.*

В околността на x_3 неравенството е несгрого, т.е. има точки з които $f(x_3 + \Delta x) = f(x_3)$. Казва се, че x_3 е точка на *нестрог (слаб) максимум.* Подобно, x_7 е точка на *нестрог (слаб) минимум.* Останали-те точки, в конто се изпълняват строги неравенства, с изключение на x_6, са точки на *строги (силни) екстремуми.*

Известно е, че производната на *f(x)* има нулева стойност в точ-ките на екстремум и в *инфлексните* и *седлови* точки. На фиг.9.1. x_6 е инфлексна точка.

9.1.1. НЕОБХОДИМИ И ДОСТАТЪЧНИ УСЛОВИЯ ЗА СЪЩЕСТВУВАНЕ НА ЕКСТРЕМУМИ

Разглежда се функция от *n* променливи f(X). Предполага се, че първите и вторите частни производни на f(X) са непрекъснати във всяка точка X. Векторът на *градиента* на f(X) в точката X = X' ще означим с $\nabla f(X')$, т.е.

В сила са следните теореми:

$$\nabla f(\mathbf{X}') = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots \frac{\partial f}{\partial x_n} \right] \bigg|_{\mathbf{X}=\mathbf{X}'}$$

а матрицата на вторите частни производни, т.нар. *хесианова* матрица

$$- \text{с } \mathbf{H}, \mathbf{H} = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right), i, j = \overline{1, n}.$$

Теорема 9.1. Необходимо условие X* да е точка на екстремум на функцията /(X) е

$$\nabla f(\mathbf{X}^*) = \mathbf{0},$$

Доказателство. В околността на X* f(X) може да се представи чрез ред на Тейлор

$$f(\mathbf{X}^* + \Delta\mathbf{X}) = f(\mathbf{X}^*) + \nabla f(\mathbf{X}^*)\Delta\mathbf{X} + \frac{1}{2}(\Delta\mathbf{X})^T\mathbf{H}\Delta\mathbf{X}\Big|_{\mathbf{X}^*+\theta.\Delta\mathbf{X}},$$

където $0 < \theta < 1$. Ако $[\Delta x_i]$ са малки, остатъчният член $1/2(\Delta X)^T H \Delta X$ е безкрайно малка величина от втори ред $0(\Delta x_i^2)$, поради което

$$f(X^* + \Delta X) - f(X^*) = \nabla f(X^*), \Delta X + 0(\Delta x^2_i) \approx \nabla f(X^*) \Delta X$$

Нека X* е точка на максимум. Да допуснем обратното, че в точката X* $\nabla f(X^*) \neq 0$, т.е. че за някое j, $\frac{\partial f(\mathbf{X}^*)}{\partial x_j} > 0$ или $\frac{\partial f(\mathbf{X}^*)}{\partial x_j} < 0$. Тогава може да се избере такъв знак на Δx_j, че $\Delta x_j \frac{\partial f(\mathbf{X}^*)}{\partial x_j} > 0$. При това, ако останалите Δx_i се положат равни на нула, от разложението в ред на Тейлор следва, че

$$f(X^* + \Delta X_j) > f(X^*),$$

което противоречи на предположението, че X* е точка на максимум Следователю) градиентът $\nabla f(X^*) = 0$.

Условието $\nabla f(X^*) = 0$ се удовлетворява също и в инфлексните и в седловите точки. Поради това то е *необходимо, но недостатъчно* условие на определяне на екстремума. Всички точки, конто го удовлет-воряват, се наричат *стационарни* точки.

Теорема 9.2. За да бъде стационарната точка X* точка на максимум (минимум), е *достатьчно* хесиановата матрица H в точката X* да е отрицателно (положително) определена.

Доказателство. От теоремата на Тейлор

$$f(\mathbf{X}^* + \Delta\mathbf{X}) = f(\mathbf{X}^*) + \nabla f(\mathbf{X}^*)\Delta\mathbf{X} + \frac{1}{2}(\Delta\mathbf{X})^T\mathbf{H}\Delta\mathbf{X}\Big|_{\mathbf{X}^*+\Theta\Delta\mathbf{X}}$$

където $0 < \theta < 1$. Тъй като X* е стационарна точка, $\nabla f(X^*) = 0$ и

Нека X* да е точка на максимум. Тогава $f(X^* + \Delta X) < f(X^*)$ при $\Delta\mathbf{X} \neq 0$, което означава, че $\frac{1}{2}\ (\Delta\mathbf{X})^T\mathbf{H}\Delta\mathbf{X}\big|_{\mathbf{X}^*+\Theta\Delta\mathbf{X}} < 0$. Поради непрекъснатостта на вторите производни знакът на $1/2(\Delta X)^T H\Delta X$ е един и същ в точките X* и X* $+\theta\Delta X$. А квадратичната форма $(\Delta X)^T H\Delta X|_{x^*}$ е отрицателна тогава и само тогава, когато $H|_{x^*}$ е отрицателно определена матрица. Това е и достатъчното условие за съществуване на максимум в точката X*.

По подобен начин се доказва и твърдението за съществуване на минимум.

Известно е, че необходимите и достатъчни условия една квадратична форма $Q(X) = X^T AX$ да е положително или отрицателно определена (полуопределена) са следните:

-Q(X) е *положително определена,* $Q(X) > 0$, ако ъгловите (воде-щите) главни минори на нейната матрица А са положителни. Тогава А е *положително определена* матрица, $A > 0$.

-Q(X) е *положително полуопределена,* $Q(X) \geq 0$, ако матрицата А е изродена и всички нейни главни минори са неотрицателни. В този случай А е *положително полуопределена* матрица, $A \geq 0$;

— Q(X) е *отрицателно определена,* $Q(X) < 0$, ако стойността на k-тия ъглов главен минор на матрицата А има знака на $(-1)^k$, $k = 1,n$. Тогава А е *отрицателно определена матрица,* $A < 0$;

— Q(X) е *отрицателно полуопределена,* $Q(X) \leq 0$, ако k-тият ъглов главен минор на матрицата А е или равен на нула, или има знак $(-1)^k$, $k = 1,n$. Тогава А е *отрицателно полуопределена* матрица, $A \leq 0$.

Ако H(X*) е *неопреде лена* матрица, X* е *седлова* точка. Ако H(X*) е *полуопределена,* X* може да е точка на екстремум. *Пример 9.1.* Дадена е функцията

$$f(x_1,x_2,x_3) = x_1 + 3x_3 + 2x_2x_3 - 2x_1^2 - 2x_2^2 - x_3^2.$$

От необходимото условие $\nabla f(X^*) = 0$ се получава системата уравнения,

$$\frac{\partial f}{\partial x_1} = 1 - 4x_1 = 0$$
$$\frac{\partial f}{\partial x_2} = 2x_3 - 4x_2 = 0$$
$$\frac{\partial f}{\partial x_3} = 3 + 2x_2 - 2x_3 = 0,$$

решението на която е $X^* = \left[\frac{1}{4}, \frac{3}{2}, 3\right]$

За да се провери достатъчното условие, се определи хесиановата матрица

$$\mathbf{H}_{\mathbf{X}^*} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 \partial x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{bmatrix}_{\mathbf{X}^*} = \begin{bmatrix} -4 & 0 & 0 \\ 0 & -4 & 2 \\ 0 & 2 & -2 \end{bmatrix}$$

Водещите главни минори на H_{X^*} са равни на —4,16,-16. Следователно H_X е отрицателно определена и $\mathbf{X}^* = \left[\frac{1}{4}, \frac{3}{2}, 3\right]$ е точка на максимум.

В случай на функция от една променлива матрицата H съдържа един елемент и достатъчните условия са

$f''(x^*) < 0$ - за максимум в точката x^*,

$f''(x^*) > 0$ - за минимум в точката x^*.

Ако $f''(x^*) = 0$, необходимо е да се анализират производните от по-висок ред. В сила е следната теорема:

Теорема 9.3. Ако в стационарната точка x^* първите (n — 1) производни на $f(x)$ са равни на нула, а $f^{(n)}(x^*) \neq 0$, точката x^* е

а) инфлексна точка, ако *n* е *нечетно.*

б) точка на екстремум, ако *n* е *четно: на максимум* при $f^{(n)}(x^*) < 0$ или на *минимум* при $f^{(n)}(x^*) > 0$.

9.1.2. МЕТОД НА НЮТОН-РАФСЪН

Системата нелинейни алгебрични уравнения, която възниква от необходимото условие $\nabla/(X) = 0$, обикновено се решава с използува не на числени методи. Процесът на решение е итерационен и методът на Нютон-Рафсън е широко разпространена итерационна процедура за решаване на нелинейни системи. Всъщност той е *градиентен* метод на безусловна оптимизация.

Разглежда се нелинейната система

$$f_i(X) = 0, i = \overline{1,m}.$$

Ако X_k е зададена точка, от теоремата на Тейлор следва, че

$$f_i(X) \approx f_i(X_k) + \nabla f_i(X_k)(X - X_k),\ i = \overline{1,m}$$ и началната

система се представя приблизително във вида

$$f_i(X_k) + \nabla f_i(X_k)(X - X_k) = 0,\ i = \overline{1,m}$$

или в матрична форма

$$\mathbf{F}_k + \frac{\partial \mathbf{F}_k}{\partial \mathbf{x}}(\mathbf{X} - \mathbf{X}_k) = \mathbf{0},$$

където

$$\mathbf{F}_k = [f_1(\mathbf{X}_k), f_2(\mathbf{X}_k), \ldots f_m(\mathbf{X}_k)]^T,\ \frac{\partial \mathbf{F}_k}{\partial \mathbf{X}} = \begin{bmatrix} \nabla f_1(\mathbf{X}_k) \\ \nabla f_2(\mathbf{X}_k) \\ \vdots \\ \nabla f_m(\mathbf{X}_k) \end{bmatrix}.$$

При предположение, че $f_i(X)$ са независими, матрицата $\frac{\partial \mathbf{F}_k}{\partial \mathbf{X}}$ е неизро-

$$\mathbf{X} = \mathbf{X}_k - \left[\frac{\partial \mathbf{F}_k}{\partial \mathbf{X}}\right]^{-1} \mathbf{F}_k.$$

Процесът започва от някаква начална точка X_0. Като се използува последното съотношение, се определи X и се полага $X_1 = X$. На след-вашата итерация се пресмята X_2 при зададеното X_1 и т.н., докато се постигне $X_r \approx X_{r-1}$ като X_r е приблизителното решение на системата. Иолученото рекурентно съотношение често е неудобно за практически пресмятания поради необходимостта да се определя обратна матрица. Поради това то се преобразува така, че вместо "обръщане на матрица" да се използува решаване на система линейни уравнения (например по метода на Гаус с избор на водещ елемент). За целта съотношението се представя във вида

$$\mathbf{X} - \mathbf{X}_k = -\left[\frac{\partial \mathbf{F}_k}{\partial \mathbf{X}}\right]^{-1} \mathbf{F}_k.$$

След умножаване на двете страни отляво по $\frac{\partial \mathbf{F}_k}{\partial \mathbf{X}}$ и въвеждане на означението лX$_к$ = X — X_k, се получава

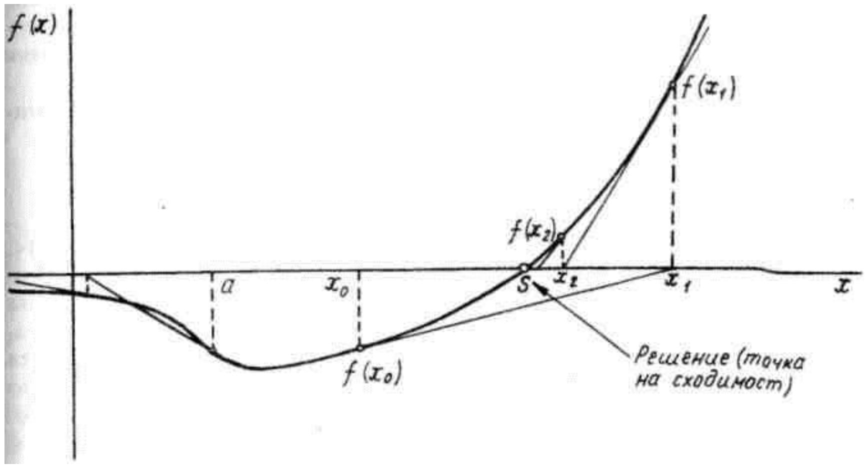
$$\frac{\partial \mathbf{F}_k}{\partial \mathbf{X}} \Delta \mathbf{X}_k = -\mathbf{F}_k.$$

Тази линейна система се решава спрямо неизвестните ΔX_k, след което се намира новата точка $X = X_k + \Delta X_k$.

В случая на функция от една променлива рекурентното съотноше-ние е твърде просто

$$x_{k+1} = x_к - f(x_k)/f'(x_k).$$

Геометричната интерпретация е показана на фиг.9.2.



Фиг. 9.2

Сходимостта на метода зависи от началното приближение X_0. Ако след няколко итерации (например 7-8) процесът не завърши, избира се ново приближение X_0, откъдето процесът започва отначало. От фиг.9.2 се вижда, че ако X_0 е в точка а, процесът е разходим.

9.2. УСЛОВНА ОПТИМИЗАЦИЯ

В условните екстремални задачи ограниченията могат да бъдат във вид на равенства или неравенства.

9.2.1. ОГРАНИЧЕНИЯ ВЪВ ВИДА НА РАВЕНСТВА

Ще бъдат разгледани *методът на Якоби,* който е обобщение на симплекс-метода на ЛП и *методът на Лагранж,* който може да се раз-глежда като развитие на метода на Якоби.

9.2.1.1. МЕТОД НА ЯКОБИ (МЕТОД НА УСЛОВИИ Я ГРАДИЕНТ)

Разглежда се задачата

Да се минимизира $W = f(X)$ при ограниченията

$$g(X) = 0,$$

където $X = [x_1, x_2, \ldots, x_n]^T, g = [g_1, \ldots, g_m]^T$. Предполага се, че функци-ите f(X) и $g_i(X)$ са два пъти непрекъснато диференцируеми.

Основната идея при метода на условния градиент е да се намери аналитичен израз за първите частни производни на функцията f(X) във всички точки, които удовлетворяват ограниченията g(X) = 0. При това стационарни ще са точките, в които тези частни производни стават равни на нула. Частните производни и градиентът, построени в точките, конто удовлетворяват ограничението g(X) = 0, се наричат *условни* частни производни и *условен* градиент.

Идеята е илюстрирана на фиг.9.3. Функцията $f(x_1, x_2)$ се минимизира при ограничението

$$g_1(x_1, x_2) = x_2 - a = 0,$$

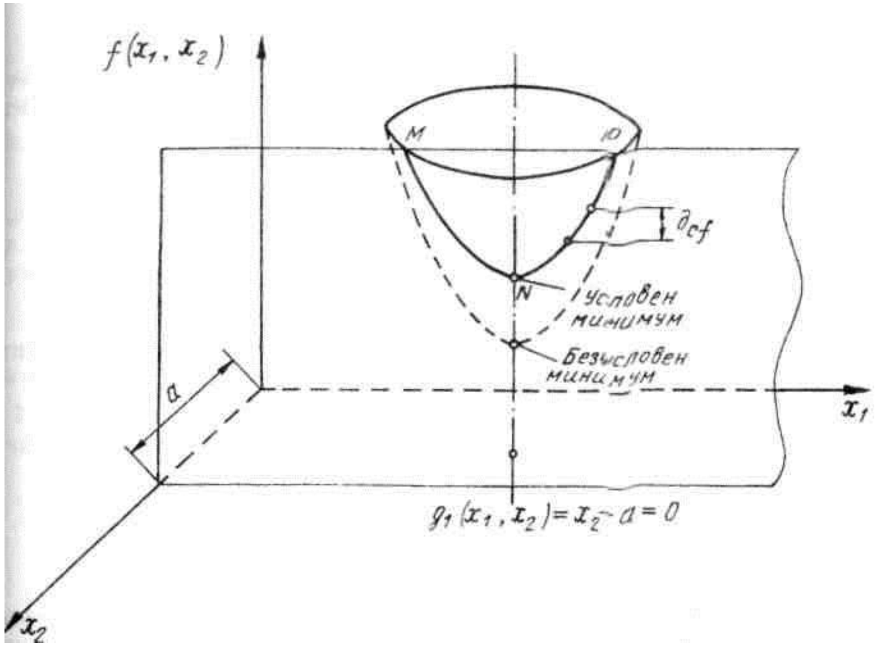
където *a* е константа. Вижда се, че стойностите на $f(x_1, x_2)$ в точките, които удовлетворяват ограничението, се представят от кривата, коя-то преминава през точките *M, N* и *P.* Според метода компонентите на условния градиент се оценяват във всяка точка от тази крива и точката, в която условните производни стават равни на нула, е стационарната точка на задачата с ограничение - точка *N.* Показано е също *условното* нарастване $d_c f$ на функцията $f(x_1,x_2)$.

От теоремата на Тейлор следват изразите

$$f(\mathbf{X} + \Delta\mathbf{X}) = f(\mathbf{X}) + \nabla f(\mathbf{X})\Delta\mathbf{X} + 0(\Delta x_j^2),$$

$$\mathbf{g}(\mathbf{X} + \Delta\mathbf{X}) = \mathbf{g}(\mathbf{X}) + \nabla\mathbf{g}(\mathbf{X})\Delta\mathbf{X} + 0(\Delta\mathbf{X}_j^2).$$

След прехвърляне на f(X) и g(X) от ляво на съответните равенства и полагане $\Delta x_j \to 0, j = \overline{1,n}$ двете уравнения се представят във вида



Фиг. 9.3.

$$\partial f(\mathbf{X}) = \nabla f(\mathbf{X})\partial\mathbf{X},$$

$$\partial g(\mathbf{X}) = \nabla \mathbf{g}(\mathbf{X})\partial\mathbf{X}.$$

Тъй като в допустимата област g(X) = 0, трябва и $\partial\mathbf{g}(\mathbf{X}) = \mathbf{0}$ и може да се запише

$$\partial f(\mathbf{X}) - \nabla f(\mathbf{X})\partial\mathbf{x} = 0$$

$$\nabla g(\mathbf{X})\partial\mathbf{X} = \mathbf{0}.$$

Това е система от $(m + 1)$ уравнения с $(n + 1)$ неизвестни -$\partial f(\mathbf{X})$ и $\partial\mathbf{X}$. Неизвестното $\partial f(X)$ може да се определи, ако са извсестни $\partial\mathbf{X}$, така че системата има *m* уравнения с *n* неизвестни.

Ако $m > n$, поне $m - n$ уравнения са излишни и могат да бъдат Отстранени, след което броят на независимите уравнения става $m \leq n$. Согато $m = n$, решението е тривиално: $\partial X = 0$. В този случай точката X няма допустима околност и пространството на решението се състои от една точка. Очевидно, интерес представлява случаят, когато $m < n$.

Нека $X = [Y^T Z^T]^T$, където

$Y = [y_1, y_1, \ldots, y_m]^T$ *и* $Z = [z_1, z_2, \ldots, z_{n-m}]^T$

са съответно *зависимите* и *независимите* променливи, конто съответствуват на X. С тези означения градиентите на f и g се записват във формата

$$\nabla f \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = [\nabla_y f, \nabla_z f],$$

$$\nabla \mathbf{g} \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = [\nabla_y \mathbf{g}, \nabla_z \mathbf{g}].$$

Вижда се, че $\nabla_y g$ и $\nabla_z g$ са матрици, за конто ще въведем означенията J и C

$$\mathbf{J} = \nabla_y \mathbf{g} = \begin{bmatrix} \nabla_y g_1 \\ \vdots \\ \nabla_y g_m \end{bmatrix}_{m \times m}, \quad \mathbf{C} = \nabla_z \mathbf{g} = \begin{bmatrix} \nabla_z g_1 \\ \vdots \\ \nabla_z g_m \end{bmatrix}_{m \times (n-m)}$$

и които се наричат съответно *матрица на Якоби (якобиян* J) и *матрица на управлението.* Предполага се, че якобиянът J е неизродена матрица, защото $g_i(X)$ са *m* независими уравнения по определение, а компонен-тите на Y може да се изберат сред тези на X така, че J да е неизродена. Като се използуват въведените вектори и матрици, системата уравнения с неизвестните $\partial f(X)$ и ∂X може да се запише във вида

$$\partial f \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = \nabla_y f \partial \mathbf{Y} + \nabla_z f \partial \mathbf{Z}$$

$$\mathbf{J} \partial \mathbf{Y} = -\mathbf{C} \partial \mathbf{Z},$$

откъдето

$$\partial \mathbf{Y} = -\mathbf{J}^{-1} \mathbf{C} \partial \mathbf{Z}.$$

Последните уравнения определят зависимите променливи $\partial \mathbf{Y}$ като функция на независимите $\partial \mathbf{Z}$. След заместване на $\partial \mathbf{Y}$ в израза на $\partial f \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}$

$$\partial f \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix} = [\nabla_z f - \nabla_y f \mathbf{J}^{-1} \mathbf{C}] \partial \mathbf{Z}.$$

От това уравнение се определят условните производни и условният градиент $\nabla_c f$ на функцията f спрямо вектора Z

Условният гради

$$\nabla_c f = \frac{\partial_c f \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}}{\partial_c \mathbf{Z}} = \nabla_z f - \nabla_y f \mathbf{J}^{-1} \mathbf{C}.$$

е равен на нула в стационарните точки.

Достатъчните условия са аналогични на тези от раздела 9.1.1. Елементите на хесиана съответствуват на елементите на вектора Z и се наричат *условии* втори производни. Те се определят, като се изпол-зува съотношението

$$\nabla_c f = \nabla_z f - \mathbf{VC}.$$

Векторът $\frac{\partial \nabla_c f}{\partial z_i}$ е i-тият ред на условния хесиан. Тъй като $\partial Y = —J^{-1} C \partial Z$,V е функция на Y, а Y - функция на Z. Поради това час-тната производна на $\nabla_c f$ по z_i се определя по правилото

$$\frac{\partial V_j}{\partial z_i} = \frac{\partial V_j}{\partial y_j} \frac{\partial y_j}{\partial z_i}$$

Пример 9.2. Зададени са:

$$f(X) = 4x_1^2 + 3x_2^2 + 7x_1x_3^2$$

$$g_1(X) = 2x_1x_3 + 3x_2 + x_2^2 - 26 = 0$$

$$g_2(X) = 4x_1^2 + 3x_1x_2 + x_3^2 - 29 = 0.$$

и допустимата точка $X_0 = [1,3,4]$. Необходимо е да се оцени условното изменение $\partial_c f$ на /(X) в допустимата околност на точката X_0.

Нека $Y = [x_1, x_3]$ и $Z = x_2$. Тогава

$$\nabla_y f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_3}\right] = [8x_1 + 7x_3^2, 14x_1x_3],$$

$$\nabla_z f = \frac{\partial f}{\partial x_2} = 6x_2,$$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_3} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 2x_3 & 2x_1 \\ 8x_1 + 3x_2 & 2x_3 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 3+ & 2x_2 \\ 3x_1 & \end{bmatrix}.$$

Нека е зададено малко изменение $\partial x_2 = 0{,}02$ на независимата промен-лива x_2. За да се оцени $\partial_c f$ в допустимата околност на $X_0 = [1,3,4]$, се пресмята

$$\mathbf{J}^{-1}\mathbf{C} = \begin{bmatrix} 8 & 2 \\ 17 & 8 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 3 \end{bmatrix} \approx \begin{bmatrix} 8/30 & -2/30 \\ -17/30 & 8/30 \end{bmatrix} \begin{bmatrix} 9 \\ 3 \end{bmatrix} = \begin{bmatrix} 2,20 \\ -4,30 \end{bmatrix}$$

Оттук

$$\partial_c f = [\nabla_z f - \nabla_y f \mathbf{J}^{-1}\mathbf{C}]\partial \mathbf{Z} = [6 \times 3 - [120, 56] \begin{bmatrix} 2,20 \\ -4,30 \end{bmatrix}]\partial x_2 \approx$$

$$\approx -5,2\partial x_2 = -0,104.$$

Измененията ∂x_1 и ∂x_3 на зависимите променливи x_1 и x_3 се определят от

$$\partial \mathbf{Y} = -\mathbf{J}^{-1}\mathbf{C}\partial \mathbf{Z}$$

При $\partial x_2 = 0,02$

$$\begin{bmatrix} \partial x_1 \\ \partial x_3 \end{bmatrix} = -\begin{bmatrix} 2,20 \\ -4,30 \end{bmatrix} 0,02 = \begin{bmatrix} -0,044 \\ 0,086 \end{bmatrix}$$

За да се провери точността на получената оценка $\Delta_c f$, може да се пре-сметнат стойностите на f(X) в точките $X = X_0$ и $X = X_0 + \partial X$, където

$$X_0 + \partial X = [1-0,044, 3 + 0,02, 4 + 0,086] = [0,956, 3,02, 4,086].$$

Получават се $f(X_0) = 143$ и $f(X_0 + \partial X) = 142,743$, откъдето $\partial_c f = f(X_0 + \partial X) - f(X_0) = -0,257$. Тази стойност се различава от стойността на $\partial_c f$, определена по формулите по-горе. Разликата се дължи на линей-ната апроксимация в околността на точката X_0. Формулите определят достатъчно точно $\partial_c f$ само при много малки отклонения от X_0.

Пример 9.3. Да се минимизира функцията

$$f(X) = x_1^2 + 2x_2^2 + x_3^2$$

при ограниченията

$$g_1(X) = x_1 + 2x_2 + 3x_3 - 3 = 0$$

$$g_2(X) = 7x_1 + 3x_2 + x_3 - 7 = 0$$

Полагат се $Y = [x_1, x_2]^T$, $z = x_3$ и се определят:

$$\nabla_y f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right] = [2x_1, 4x_2], \nabla_z f = \frac{\partial f}{\partial x_3} = 2x_3,$$

$$\mathbf{J} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 7 & 3 \end{bmatrix}, \mathbf{J}^{-1} = \begin{bmatrix} -3/11 & +2/11 \\ +7/11 & -1/11 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \frac{\partial g_1}{\partial x_3} \\ \frac{\partial g_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Оттук

$$\nabla_c f = \frac{\partial_c f}{\partial_c x_3} = \nabla_z f - \nabla_y f \mathbf{J}^{-1}\mathbf{C} = 2x_3 -$$

$$- [2x_1, 4x_2] \begin{bmatrix} -3/11 & 2/11 \\ 7/11 & -1/11 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} =$$

$$= \frac{14}{11}x_1 - \frac{80}{11}x_2 + 2x_3.$$

Необходимото условие за стационарна точка $\nabla_c f = 0$ и двете огра-ничения $g_1(X) = 0$ и $g_2(X) = 0$ образуват системата уравнения

$$\begin{bmatrix} \frac{14}{11} & -\frac{80}{11} & 2 \\ 1 & 2 & 3 \\ 7 & 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 7 \end{bmatrix}$$

чието решение е $X^* = [0, 799, 0,289, 0, 541]^T$.

За да се провери достатъчното условие, се определи условният хе-сиан. Тъй като независимата променлива е x_3, от израза за $\nabla_c f$ се получава

$$\frac{\partial_c^2 f}{\partial_c x_3^2} = \frac{14}{11}\frac{dx_1}{dx_3} - \frac{80}{11}\frac{dx_2}{dx_3} + 2 = \left[\frac{14}{11}, -\frac{80}{11}\right] \begin{bmatrix} dx_1/dx_3 \\ dx_2/dx_3 \end{bmatrix} + 2.$$

От метода на Якоби е известно, че

$$\partial Y = -J^{-1}C\partial Z,$$

т.е.

$$\begin{bmatrix} dx_1/dx_3 \\ dx_2/dx_3 \end{bmatrix} = -\mathbf{J}^{-1}\mathbf{C} = \begin{bmatrix} 7/11 \\ -20/11 \end{bmatrix}.$$

След заместване в израза за условната втора производна

$$\frac{\partial_c^2 f}{\partial_c x_3^2} = \left[\frac{14}{11}, -\frac{80}{11}\right] \begin{bmatrix} 7/11 \\ -20/11 \end{bmatrix} = \frac{1698}{121} > 0.$$

Следователно X* е точка на минимум.

При използването на метода на Якоби могат да възникнат затруднения, свързани с получаването на J^{-1} при голям брой на ограниче-нията. В такива случаи може да се използва правилото на Крамер за определяне на ∂f чрез независимите променливи ∂Z, след като се получават и условните производни $\frac{\partial_c f}{\partial_c z_j}$ Анализ на чувствителността Методът на Якоби може да се използва за анализ на чувствителността на оптималната стойност на f към малки изменения на десните части на ограниченията. За разлика от линейния случай обаче, ре-зултатите от анализа са валидни в малка околност около точката на екстремума.

Нека дясната част на ограниченията g(X) = 0 е изменена на ∂g Тогава вместо условието за допустимост ∇ g(X)∂X = 0, използвано при извеждането на съотношенията по-горе, ще имаме∇g(X)∂X = ∂g, което след въвеждането на векторите Y и Z може да се представи във вида

$$\left[\nabla_y \mathbf{g}\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix}, \nabla_z \mathbf{g}\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix}\right] \begin{bmatrix}\partial \mathbf{Y}\\ \partial \mathbf{Z}\end{bmatrix} = \partial \mathbf{g}$$

или

$$J\partial Y + C\partial Z = \partial g.$$

По такъв начин системата уравнения с неизвестни ∂f(X) и ∂X се за-писва във формата

$$\partial f(\mathbf{Y}, \mathbf{Z}) = \nabla_y f \partial \mathbf{Y} + \nabla_z f \partial \mathbf{Z}$$
$$\partial \mathbf{g} = \mathbf{J}\partial \mathbf{Y} + \mathbf{C}\partial \mathbf{Z}.$$

Тъй като$\partial \mathbf{g} \neq \mathbf{0}$, от второто уравнение се получава

$$\partial \mathbf{Y} = \mathbf{J}^{-1}\partial \mathbf{g} - \mathbf{J}^{-1}\mathbf{C}\partial \mathbf{Z}.$$

След заместване в уравненията за ∂f

$$\partial f\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix} = \nabla_y f \mathbf{J}^{-1}\partial \mathbf{g} + \nabla_c f \partial \mathbf{Z},$$

където, както беше положено по-рано

$$\nabla_c f = \nabla_z f - \nabla_y f \mathbf{J}^{-1}\mathbf{C}.$$

Полученият израз за $\partial f\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix}$ може да се използва за анализ на изменениято на f в допустима околност на допустимата точка X_0, породени от малки изменения dg и ∂Z.

Нека X_0 е точка на екстремум, т.е. X_0 = X*. В тази точка условният градиент$\nabla_c f$ =0и изразът за $\partial f\begin{pmatrix}\mathbf{Y}\\ \mathbf{Z}\end{pmatrix}$ се записва във вида

$$\partial f\begin{pmatrix}\mathbf{Y}^*\\ \mathbf{Z}^*\end{pmatrix} = \nabla_{y^*} f \mathbf{J}^{-1}\partial \mathbf{g}\begin{pmatrix}\mathbf{Y}^*\\ \mathbf{Z}^*\end{pmatrix}$$

или

$$\frac{\partial f}{\partial \mathbf{g}} = \nabla_y f \mathbf{J}^{-1}.$$

Величините $\frac{\partial f}{\partial \mathbf{g}}$ се наричат *коефициенти на чувствителност.* Те описват "скоростта" на изменение на f спрямо изменения на g. В точката X* тези коефициенти *не зависият* от избора на компонентите на Y, което следва от факта, че изразът, който ги определя, не съдържа Z (т.е. те имат постоянни стойности).

Пример 9.4. В предишния пример беше получено, че $X^* = [x^*_1, x^*_2, x^*_3]^T = [0{,}799, 0{,}289, 0{,}541]^T$. Тъй като $Y^* = [x^*_1, x^*_2]^T$

$$\nabla_{y^*} f = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right]_* = [2x_1^*, 4x_2^*] = [1,598, 1,156].$$

Оттук коефициентите на чувствителност са

$$\left[\frac{\partial f}{\partial g_1}, \frac{\partial f}{\partial g_2}\right] = \nabla_{y^*} f \mathbf{J}^{-1} = [1,598, 1,156]\begin{bmatrix} -3/11 & 2/11 \\ 7/11 & -1/11 \end{bmatrix} =$$
$$= [0,2998, 0,1854]$$

т.е. при ∂g_1 = 1,f* нараства приблизително с 0,2998, а при ∂g_2 = 1,f* нараства с 0,1854.

Методът на Якоби може да бъде използван за решаване на задачи на ЛП. Показано е, че най-важните идеи на симплекс-метода могат да бъдат интерпретирани и изведени чрез метода на Якоби.

9.2.1.2. МЕТОД НА МНОЖИТЕЛИТЕ НА ЛАГРАНЖ

Беше показано, че коефициентите на чувствителност $\partial f/\partial g$ могат да се използват за оценка на влиянието на малки изменения в десните части на ограниченията върху оптималната стойност на f и че тези коефициенти са константи в точката на оптимума X*.

Нека

$$\boldsymbol{\lambda} = \nabla_{y^*} \mathbf{J}^{-1} = \frac{\partial f}{\partial \mathbf{g}}.$$

Оттук

$$\partial f - \boldsymbol{\lambda} \partial \mathbf{g} = 0.$$

Това уравнение отразява необходимите условия за стационарни точки тъй като изразът за $\partial f / \partial g$ е получен при предположение, че $\nabla_c f = 0$. Тези условия се представят в по-удобна форма, като се вземат частните

производни по всички x_j

$$\frac{\partial}{\partial x_j}(f - \boldsymbol{\lambda}\mathbf{g}) = 0, j = \overline{1,n}.$$

Получените по този начин уравнения и ограниченията g(X) = 0 дават възможност да се определят допустимите вектори X и $\boldsymbol{\lambda}$, конто удовлетворяват *необходимите* условия за стационарност.

Описаната процедура е основа на метода на множителите на Ла-гранж. Съставя се т.нар. *функция на Лагранж L(X,* $\boldsymbol{\lambda}$)

$$L(\mathbf{X}, \boldsymbol{\lambda}) = f(\mathbf{X}) - \boldsymbol{\lambda}\mathbf{g}(\mathbf{X}),$$

където $\boldsymbol{\lambda}$ е вектор - ред на *множителите на Лагранж,* които имат сми-съла на въведените по-горе коефициенти на чувствителност. Уравне-нията

$$\frac{\partial L}{\partial \mathbf{X}} = \mathbf{0} \quad \text{и} \quad \frac{\partial L}{\partial \boldsymbol{\lambda}} = \mathbf{0}$$

определят същите необходими условия, които бяха дадени по-горе. Следователно функцията на Лагранж може да се използува направо за генериране на необходимите условия. Това означава, че определянето на екстремум на f(X) при ограниченията g(X) = 0 е еквивалентно на задачата за безусловна оптимизация на функцията на Лагранж L(X, $\boldsymbol{\lambda}$). Достатъчни условия при метода на Лагранж могат да бъдат фор-мулирани, като се използува т.нар. *оградена* хесианова матрица

$$\mathbf{H}^B = \begin{bmatrix} \mathbf{0} & \vdots & \mathbf{P} \\ \cdots & \cdots & \cdots \\ \mathbf{P}^T & \vdots & \mathbf{Q} \end{bmatrix}_{(m+n)\times(m+n)},$$

където

$$\mathbf{P} = \begin{bmatrix} \nabla g_1(\mathbf{X}) \\ \vdots \\ \nabla g_m(\mathbf{X}) \end{bmatrix}_{m \times n}, \quad \mathbf{Q} = \left[\frac{\partial^2 L(\mathbf{X}, \boldsymbol{\lambda})}{\partial x_i, \partial x_j}\right]_{n \times n} \text{ за всички } i \text{ и } j.$$

Ако стационарната точка на $L(\mathbf{X}, \boldsymbol{\lambda})$ е $(\mathbf{X}^*, \boldsymbol{\lambda}^*)$,матрицата H^B се фор-мира от стойностите на елементите й в тази точка. Точката X* е:

а) точка на максимум, ако започвайки с водещия главен минор от ред (2 m+1) *последните (n-m)* водещи главни минори от $H^в$ образуват знакопроменлив ред, първият член на който има знак $(-1)^{m+1}$.

б) точка на минимум, ако започвайки с водещия главен минор от ред (2m+1), *последните (n-m)* водещи главни минори от $H^в$ имат знак $(-1)^m$

Тези условия са *достатьчни,* но не и *необходимы,* т.е. ако те не се удовлетворяват в стационарната точка, тя пак би могла да бъде точка на екстремум. Съществуват други условия, конто са необходими и до-статъчни, но практическото приложение на конто често е свързано с непреодолими изчислителни трудности.

Пример 9.5. Разглежда се задачата от примера 9.3. Функцията на Лагранж е

$$L(\mathbf{X}, \boldsymbol{\lambda}) = x_1^2 + 2x_2^2 + x_3^2 - \lambda_1(x_1 + 2x_2 + 3x_3 - 3) - \\ - \lambda_2(7x_1 + 3x_2 + x_3 - 7).$$

Необходимите условия са:

$$\frac{\partial L}{\partial x_1} = 2x_1 - \lambda_1 - 7\lambda_2 = 0$$
$$\frac{\partial L}{\partial x_2} = 4x_2 - 2\lambda_1 - 3\lambda_2 = 0$$
$$\frac{\partial L}{\partial x_3} = 2x_3 - 3\lambda_1 - \lambda_2 = 0$$
$$\frac{\partial L}{\partial \lambda_1} = -(x_1 + 2x_2 + 3x_3 - 3) = 0$$
$$\frac{\partial L}{\partial \lambda_2} = -(7x_1 + 3x_2 + x_3 - 7) = 0.$$

Решението на тази система е X* = $[x^*_1, x^*_2, x^*_3]^T$ = [0, 799,0,289,0,541]^T, $\boldsymbol{\lambda}^* = [\lambda_1^*, \lambda_2^*]$ = [0,2998,0,1854]. Вижда се, че решението обхваща реше-нията на задачите от примерите 9.3 и 9.4.

За проверка на достатъчните условия се съставя матрицата H^B

$$\mathbf{H}^B = \begin{bmatrix} 0 & 0 & \vdots & 1 & 2 & 3 \\ 0 & 0 & \vdots & 7 & 3 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 7 & \vdots & 2 & 0 & 0 \\ 2 & 3 & \vdots & 0 & 4 & 0 \\ 3 & 1 & \vdots & 0 & 0 & 2 \end{bmatrix}$$

В случая n = 3,m = 2,n — m = 1. Необходимо е да се провери само детерминантата на H^B, която трябва да има знак $(-1)^2$, за да бъде X* точка на минимум. Получава се, че $|H^B| = 1940 > 0$, т.е. X* е точка на минимум.

В общия случай за решаването на нелинейната алгебрична система, възникнала от необходимите условия, може да бъде използван ме-тодът на Нютон-Рафсън.

9.2.2. ОГРАНИЧЕНИЯ ВЪВ ВИД НА НЕРАВЕНСТВА

9.2.2.1 ОБОБЩЕН МЕТОД НА МНОЖИТЕЛИТЕ НА ЛАГРАНЖ

Разглежда се задачата

да се максимизира $J = f(X)$ при ограничения

$$g_i(X) \leq 0, i = 1, m$$

Предполага се, че ограниченията $X \geq 0$, ако съществуват, са включени в горните ограничения.

Основната идея е следната. Ако точката на *безусловная* оптимум на f(X) не удовлетворява всички ограничения, условният оптимум трябва да се достига в *гранична точка* на областта на допустимите решения. Това означава, че едно или няколко ограничения ще се изпълняват като *равенства.*

Алгоритъмът може да се представи обобщено в следния вид:

Стъпка 1. Решава се задачата *без* ограничения

$$\text{Да се максимизира} \quad J = f(X)$$

Ако всички ограничения се удовлетворяват от оптималното решение X*, изчисленията се прекратяват. Иначе се полага $\kappa = 1$ и се преминава към стъпка 2.

Стъпка 2. Произволно избрани *к* ограничения се правят *активны* (превръщат се в равенства) и се определя максимумът на f(X) при наличие на *к* ограничения във вид на равенства по метода на Лагранж-Ако решението е допустимо спрямо останалите ограничения, изчисленията се прекратяват. Определен е *локален* оптимум - един от всички възможни оптимуми, конто биха се получили в резултат на максими-зирането на /(X) при *всички* възможни комбинации от *к* ограничения, k = 1,m. В обратния случай се активизират други *к* ограничения и стъпката се повтаря. Ако *всички* подмножества от *к* активни ограничения не довеждат до допустимо решение, преминава се към стъпка 3.

Стъпка 3. Ако *к — т,* изчисленията се прекратяват. Допустими решения не съществуват. Иначе се полага $\kappa = \kappa + 1$ и се преминава към стъпка 2.

Процедурата *не гарантира* получаването на глобален оптимум. Необходимо е също да се има предвид и следното. Ако r < *s*, опти-малното решение f_r* при r ограничения може да не е по-добро от това при *s* ограничения (f_s*). Стойността f_r* би била по-добра от f_s* *само* ако r-те ограничения са подмножество от s-те ограничения.

Процедурата може да се модифицира така, че да се определя глобален оптимум, ако f(X) е унимодална функция (с един екстремум). За целта на стъпка 2 по-горе е необходимо да се определят и сравняват условните оптимуми, съответстващи на *всички възможни комбинации от к* ограничения, *к = 1,т* и от тях да се избере най-добрият. Това обаче може да доведе до много голям брой на решаваните задачи и да направи практически неизпълнима такава процедура, ако броят на ограниченията е голям.

9.2.2.2. УСЛОВИЯ НА КАРУШ- КУН - ТАКЕР

Условията на Каруш-Кун-Такер са теоретична основа на нелиней-ното програмиране. Това са *необходими,* а при определени ограничения и *достатъчни* условия за идентифициране на стационарни точки при ограничения във вид на *неравенства.*

Разглежда се задачата

да се максимизира J = f(X) при ограничения

$$g(X) \leq 0.$$

Ограниченията от вид неравенства се преобразуват във вид на равенства чрез въвеждане на *неотрицателни* допълнителни променливи s^2. Нека $S = [s_1, s_2, ..., s_m]^T$ и $S^2 = [s_1^2, s_2^2, ..., s_m^2]^T$, където го е броят на ограниченията неравенства. Тъй като i-то ограничение е от вида $g_i(X) + s_i^2 = 0$, функцията на Лагранж може да бъде записана във формата

$$L(\mathbf{X}, \boldsymbol{\lambda}, \mathbf{S}) = f(\mathbf{X}) - \boldsymbol{\lambda}[\mathbf{g}(\mathbf{X}) + \mathbf{S}^2].$$

При зададени ограничения от вида

$$g(X) \leq 0$$

едно от *необходимите* условия за оптималност е *неотрицателността (неположителността)* на $\boldsymbol{\lambda}$ в задачата за *максимизация (минимизация).*

Това може да се покаже по следния начин. Нека задачата е за максимизация. Известно е, че множителите на Лагранж $\boldsymbol{\lambda}$ изразяват скоростта на изменение на f спрямо $\mathbf{g}$, $\boldsymbol{\lambda} = \partial f / \partial \mathbf{g}$. Ако дясната част на ограничението нараства над нулата, пространството на решенията се разширява и f не може да намалява. Следователно $\boldsymbol{\lambda} \geq \mathbf{0}$. Подобно, ако в задачата за минимизация допустимата облает се разширява, стойността на f не може да се влошава (да нараства), т.е. $\boldsymbol{\lambda} \leq \mathbf{0}$. Ако ограниченията са равенства g(X) = 0, не се налагат ограничения върху знака на $\boldsymbol{\lambda}$

Останалите необходими условия се извеждат от условията за екстремум на L

$$\frac{\partial L}{\partial \mathbf{X}} = \nabla f(\mathbf{X}) - \boldsymbol{\lambda} \nabla \mathbf{g}(\mathbf{X}) = \mathbf{0}$$

$$\frac{\partial L}{\partial s_i} = -2\lambda_i s_i = 0, \quad i = \overline{1,m}$$

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = -(\mathbf{g}(\mathbf{X}) + \mathbf{S}^2) = \mathbf{0}$$

От втората група уравнения следват следните изводи:

а) ако $\lambda_i > 0$, то $s_i^2 = 0$. Следователно ресурсът, който съответствува на това ограничение, е дефицитен, т.е. напълно се изчерпва и ограничението става равенство.

б) ако $s_i^2 > 0$, то $\lambda_i = 0$. Следователно г'-тият ресурс не е дефицитен и неговото изменение не влияе на f, $\lambda_i = \partial f / \partial g_i = 0$.

От втората и третата група уравнения следва, че

$$\lambda_i g_i(\mathbf{X}) = 0, \quad i = \overline{1,m}.$$

Наистина, ако $\lambda_i > 0$, то $s_i = 0$ и 0,(X) = 0, т.е. $\lambda_i g_i(X) = 0$. Обратно, ако $g_i(X) < 0$, то $s_i^2 > 0$ и $\lambda_i = 0$, т.е. $\lambda_i g_i(\mathbf{X}) = 0$.

И така, необходимите условия на Каруш-Кун-Такер $[X \quad ,\boldsymbol{\lambda}]^T$ да е стационарна точка в задача за максимизация са следните:

$$\boldsymbol{\lambda} \geq \mathbf{0}$$

$$\nabla f(\mathbf{X}) - \boldsymbol{\lambda} \nabla \mathbf{g}(\mathbf{X}) = \mathbf{0}$$

$$\lambda_i g_i(\mathbf{X}) = \mathbf{0}, \quad i = \overline{1,m}$$

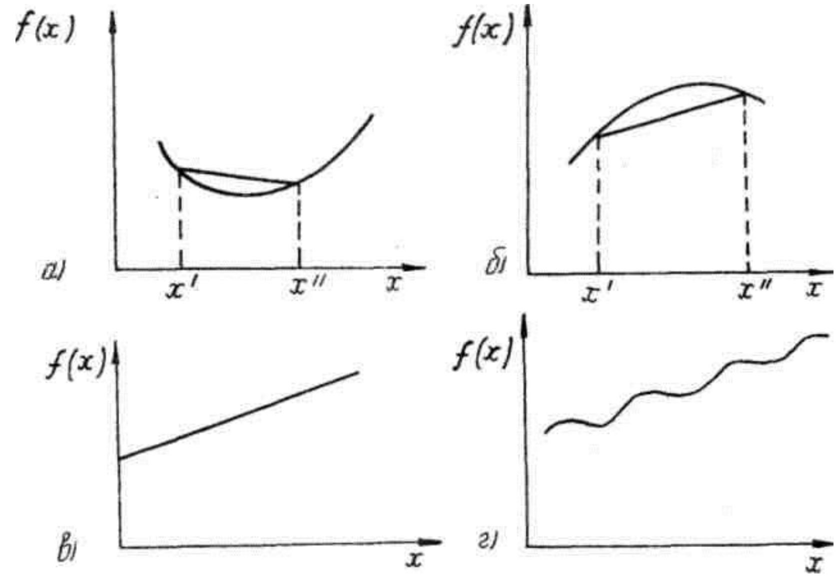
$$g(\mathbf{X}) \leq \mathbf{0}$$

В задача за минимизация първата група условия се заменя с $\boldsymbol{\lambda} \leq$ O, останалите условия не се променят. Не се налагат ограничения върху знака на множителите на Лагранж, конто съответствуват на ограничения равенства.

Достатъчност на условията на Каруш- Кун- Такер Необходимите условия на Каруш-Кун-Такер са и *достатьчни,* ако целевата функция f е *вдлъбната* - в задача за максимизация, или *изпъ-кнала* - в задача за минимизация, и ако пространството на решенията е *изпъкнало множество.* На фиг.9.4 а,б,в и г са показани съответно изпъ-кнала функция, вдлъбната функция, линейна функция, която се приема едновременно за изпъкнала и вдлъбната, и функция, която не е нито из-пъкнала, нито вдлъбната. По определение функцията от една промен-лива f(x) е изпъкнала, ако за всяка двойка стойности на *x,* например x' и x''

$$f[rx'' + (1 - r)x'| < rf(x'') + (1 - r)f(x')$$

за всяко r от интервала $0 \leq r \leq 1$. Тя е *строго изпъкнала,* ако неравен- ството $\leq$ се замени с $<$. Тя е *вдлъбната (строго вдлъбната)* функция, ако неравенството $\leq$ се замени с $\geq$ (или с $>$). Тези определения имат ясна геометрична интерпретация. Ако *[x',f(x')}* и *[x",f(x")]* са две точки от графичното изображение на f(x), [rx"+(l - r)x',rf(x")+(l — *r)f(x')]* описва точките от линейния сегмент между тези две точки, когато $0 \leq r \leq 1$. Според дефиницията по-горе f(x) е изпъкнала, ако линейният сегмент, свързващ кои и да е две точки от графиката на *f(x),* лежи изцяло над или върху тази графика. Необходимо и достатъчно условие за изпъкналост е $d^2f / dx^2 \geq 0$ за всяко x от дефиниционната облает на



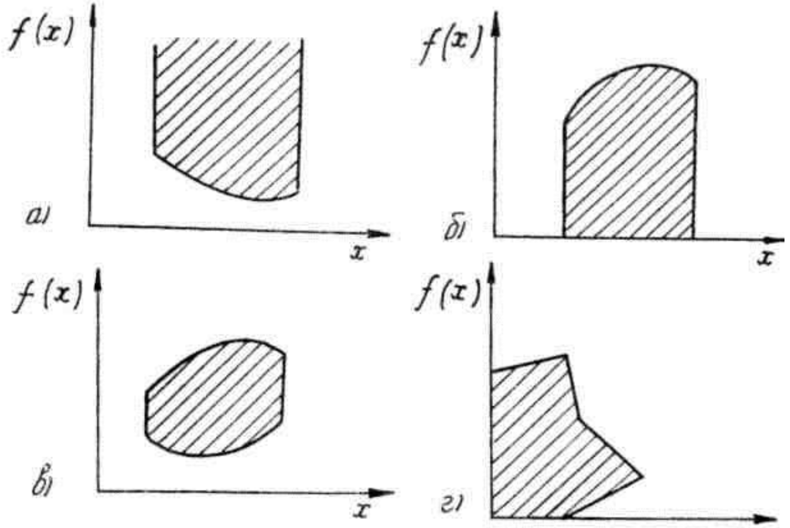
Фиг. 9.4 *а,б,в,г*

f(x).

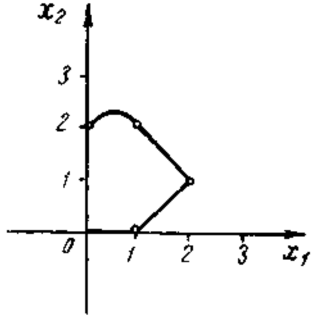
Дадените определения се запазват и за функция от n промени-ви след обобщаване на понятието линеен сегмент. Линейният сегмент свързващ две точки $X' = [x'_1, x'_2, \ldots, x'_n]$ и $X'' = [x''_1, x''_2, \ldots, x''_n]$ е мно-жеството от точки $X = [x_1, x_2, \ldots, x_n] = [rx''_1 + (1 - r)x'_1, rx''_2 + (1 - r)x'_2, \ldots, rx''_n + (1 - r)x'_n]$ такива, че $0 \leq r \leq 1$.

Изпъкналите функции притежават две важни свойства: сумата от изпъкнали функции е изпъкнала функция; ако f(X) е изпъкнала, w(X) = -f(X) е вдлъбната и обратно.

Върху понятията изпъкнала и вдлъбната функция се изгражда понятието *изпъкнало множество.* Ако f(X) е изпъкнала функция, ан-самбълът от точки над или върху графиката на f(X) е изпъкнало множество. Ако f(X) е вдлъбната, ансамбълът от точки под или върху графиката на f(X) е изпъкнало множество. На фиг.9.5 а,б и в са показани изпъкнали множества, определени съответно от изпъкнала, вдлъбната или изпъкнала и вдлъбната функция. По определение едно множество от точки е изпъкнало, ако линейният сегмент, който свързва две про-изволни негови точки, е изцяло в това множество. *Екстремна точка* на изпъкнало множество е точка от множеството, която не лежи върху нито едни линеен сегмент, свързващ две *други* точки от множеството. Например екстремните точки на изпъкналото множество на фиг.9.6 са (0,0), (1,0), (2,1), (1,2) (0,2) и безкрайният брой точки от кривата между (0,2) и (1,2).



Фиг. 9.5 а,б,в,г



Фиг. 9.6

На фиг.9.5-г е показано неизпъкнало множество.

Тъй като е по-лесно да се провери изпъкналостта на функция, отколкото да се докаже изпъкналостта на множество, по-нататък изпъ-кналостта на пространството на решенията се установява чрез проверка дали функциите $q_i(X)$ $i = \overline{1,m}$ удовлетворяват определени условия (изисквания). Разглежда се обобщената нелинейна задача

Да се максимизира (или минимизира) $J = f(X)$

при ограниченията

$$g_i(X) \leq 0, \quad i = \overline{1,k}$$

$$g_i(\mathbf{X}) \geq 0, \quad i = \overline{k+1, l},$$

$$g_i(\mathbf{X}) = 0, \quad i = \overline{l+1, m},$$

при което

$$L(\mathbf{X}, \boldsymbol{\lambda}, \mathbf{S}) = f(\mathbf{X}) - \sum_{i=1}^{k} \lambda_i [g_i(\mathbf{X}) + s_i^2] - \sum_{i=k+1}^{l} \lambda_i [g_i(\mathbf{X}) - s_i^2] - \sum_{i=l+1}^{m} \lambda_i g_i(\mathbf{X}).$$

Като се отчитат разгледаните свойства на изпъкналите функции, лесно могат да се формулират изисквания към $g_i(X)$, при конто $L(X, S, \boldsymbol{\lambda})$ е вдлъбната или изпъкнала функция, съответно за случайте на максимизация или минимизация на f(X). Тези изисквания са представени в табл.9.1. Когато те се удовлетворяват, необходимите условия на Ка-руш-Кун-Такер са и достатъчни.

При съставянето на таблицата е отчетено, че ако например $g_i(X)$ е вдлъбната, $\lambda_i g_i(\mathbf{X})$ е вдлъбната, когато $> \lambda_i$ 0 и изпъкнала, ако $\backslash \lambda_i < 0$.

Взето е също предвид и това, че $[g_i(X) + s_i*]$ е изпъкнала, когато $g_i(X)$ е изпъкнала, тъй като s^ е изпъкнала, както и това, че $[g_i(X) — s_i^2]$ е

Таблица 9

Оптими-зация	Изисквания			
	$f(\mathbf{X})$	$g_i(\mathbf{X})$	λ_i	
Максимизация	Вдлъбната	изпъкнала	≥ 0	$1 \leq i \leq k$
		вдлъбната	≤ 0	$k+1 \leq i \leq l$
		линейна	без огран.	$l+1 \leq i \leq m$
Минимизация	Изпъкнала	изпъкнала	≤ 0	$1 \leq i \leq k$
		вдлъбната	≥ 0	$k+1 \leq i \leq l$
		линейна	без огран.	$l+1 \leq i \leq m$

вдлъбната, когато $g_i(X)$ е вдлъбната, тъй като — s_i* е вдлъбната. Лесно е да се види, че в табл.9.1 са показани само част от условията за доста-тъчност на необходимите условия на Каруш-Кун-Такер. Допустимата област може да е изпъкнала и в случаи, когато не се удовлетворяват изискванията към $g_i(X)$ от горната таблица. Например [g_i (X) + s_i2*]* може да е изпъкнала и при "слабо" вдлъбната $g_i(X)$. *Пример 9.6.* Да се минимизира функцията

$$f(X) = 2x_1^2 + x_2^2 + 2x_3^2$$

при ограниченията

$g_1(X) = x_1 + 2x_2 - 6 < 0$

$g_2(X) = 2x_1 + x_3 - 5 < 0$

$g_3(X) = 2 - x_1 < 0$ $g_4(X) = 3 -$

$2x_2 < 0$

$g_5(X) = -x_3 < 0.$

Функцията f(X) е изпъкнала, пространството на решенията g(X) ≤ O е също изпъкнало. Следователно функцията на Лагранж L(X, $\boldsymbol{\lambda}$, S) при $\boldsymbol{\lambda} \leq \mathbf{0}$ е изпъкнала и нейната стационарна точка, определена от услови-ята на Каруш-Кун-Такер, е точка на условен глобален минимум.

Необходимите условия се записват във вида

$$[\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] \leq \mathbf{0}$$

$$[4x_1, 2x_2, 4x_3] - [\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5] \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

$$\lambda_1 g_1 = \lambda_2 g_2 = \lambda_3 g_3 = \lambda_4 g_4 = \lambda_5 g_5$$

$$g_i(\mathbf{X}) \leq 0, \quad i = \overline{1,5}.$$

След извършване на операциите се получава:

$$\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5 \leq 0 \quad (1)$$
$$4x_1 - \lambda_1 - 2\lambda_2 + \lambda_3 = 0 \quad (2)$$
$$2x_2 - 2\lambda_1 + 2\lambda_4 = 0 \quad (3)$$
$$4x_3 - \lambda_2 + \lambda_5 = 0 \quad (4)$$
$$\lambda_1(x_1 + 2x_2 - 6) = 0 \quad (5)$$
$$\lambda_2(2x_1 + x_3 - 5) = 0 \quad (6)$$
$$\lambda_3(2 - x_1) = 0 \quad (7)$$
$$\lambda_4(3 - 2x_2) = 0 \quad (8)$$
$$\lambda_5 x_3 = 0 \quad (9)$$
$$x_1 + 2x_2 - 6 \leq 0 \quad (10)$$
$$2x_1 + x_3 - 5 \leq 0 \quad (11)$$
$$x_1 \geq 2 \quad (12)$$
$$x_2 > 3/2 \quad (13)$$
$$x3 > 0 \quad (14)$$

Вижда се, че (7) и (12), (8 и 13) и (9) и (14) се удовлетворяват съответно от $x^*_1 = 2$, $x^*_2 = 3/2$ и $x^*_3 = 0$. Тези стойности удовлетворяват условия-та (10), (11), а за да се удовлетворят и условията (5) и (6), трябва да се изберат $\lambda_1^* = \lambda_2^* = 0$. От условие (4) следва, че $\lambda_5^* = 0$._ Накрая, за да се изпълнят (2) и (3), се избират λ_3^*_ = -8 и λ_4^* = -3/2.

И така, съществуват числа $\boldsymbol{\lambda}^* \leq \mathbf{0}, \boldsymbol{\lambda}^*$ = [0,0, -8, -3/2,0] такива, че X* = $[x^*_1, x^*_2, x^*_3]^T$ = $[2,3/2,0]^T$ и $\boldsymbol{\lambda}^*$ удовлетворяват всички необходими условия. Следователно X* е точката на условен глобален минимум на началната задача.

Определянето на оптималното решение непосредствено от условията на Каруш-Кун-Такер, когато задачите са по-сложни, е трудно, дори невъзможно. В такива случаи тези условия се използват за да се прове-ри дали някое предлагано решение е оптимално. Особено важни обаче са техните *непреки* приложения, свързани с разработването на алгори-тми на НЛП, на *теорията на дуалността* в НЛП и др.

9.3. ЗАКЛЮЧЕНИЕ

В главата са представени теоретичните основи на определянето на точки на екстремум при решаването на нелинейни безусловни и условни задачи. Когато ограниченията отстъстват или са във вид на равенства, достатъчните условия позволяват да се определи видът на екстремума - максимум или минимум. Когато тези условия се удовлет-воряват в разглежданата облает на определение на целевата функция и ограниченията, задачата е изпъкнала и локалният екстремум е и глобален екстремум.

Когато ограниченията са във вида на неравенства, необходимите условия на Каруш-Кун-Такер идентифицират и вида на екстремума - максимум или минимум. Не съществуват обаче достатъчни условия, удовлетворяването на които да показва, че определеният екстремум е глобален. Единствено предварителната проверка на изискванията от табл.9.1 може да потвърди в част от случаите глобалността на оптиму-м**а.**

Независимо от това, че някои от методите и условията не винами са удобни за непосредствени изчисления, разгледаните теоретични резултати са основа за разработване на ефективни числени методи и алгоритми.