Iterační algoritmy na volné lokální extrémy: gradientní a Newtonova metoda, nelineární nejmenší čtverce. Lokální extrémy vázané rovnostmi, metoda Lagrangeových multiplikátorů. Lineární programování. Konvexní množiny a funkce, konvexní optimalizační úlohy. (Optimalizace)

Jsou tady prakticky jen okopírované relevantní úseky skript, líp než ve skriptech to popsat prakticky nejde.

## Iterační algoritmy

Sloupcový vektor parciálních derivací funkce  $f: \mathbb{R}^n \to \mathbb{R}$  se nazývá její **gradient** a značí se

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} = f'(\mathbf{x})^T$$

( $\nabla$  čteme 'nabla'). Je to tedy transpozice Jacobiho matice  $f'(\mathbf{x})$ , což je řádkový vektor<sup>7</sup>. Gradient je sloupcový vektor parciálních derivací skalární funkce.

Označme jako

$$B_{\varepsilon}(\mathbf{x}) = \{ \mathbf{y} \in \mathbb{R}^n \mid ||\mathbf{x} - \mathbf{y}|| < \varepsilon \}$$
(8.21)

n-rozměrnou kouli bez hranice<sup>8</sup> se středem  $\mathbf{x} \in \mathbb{R}^n$  a poloměrem  $\varepsilon > 0$ . Všimněte si, že pro speciální případ n = 1 (tedy  $x \in \mathbb{R}$ ) je množina (8.21) interval  $B_{\varepsilon}(x) = (x - \varepsilon, x + \varepsilon)$ .

Mějme nyní množinu  $X \subseteq \mathbb{R}^n$ . Bod  $\mathbf{x} \in \mathbb{R}^n$  se nazývá její

- vnitřní bod, jestliže existuje  $\varepsilon > 0$  takové, že  $B_{\varepsilon}(\mathbf{x}) \subseteq X$ ,
- hraniční bod, jestliže pro každé  $\varepsilon > 0$  je  $B_{\varepsilon}(\mathbf{x}) \cap X \neq \emptyset$  a  $B_{\varepsilon}(\mathbf{x}) \cap (\mathbb{R}^n \setminus X) \neq \emptyset$ .

Extrém funkce f na množině X v bodě  $\mathbf{x}$  je volný, když je  $\mathbf{x}$  vnitřní bod množiny X. Vnitřní bod množiny X je takový, jehož nějaké okolí (koule bez hranice) leží v X.

Iterační metody na hledání lokálního minima spojité funkce  $f: \mathbb{R}^n \to \mathbb{R}$  mají tvar

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{v}_k,\tag{9.1}$$

kde vektor  $\mathbf{v}_k \in \mathbb{R}^n$  je směr hledání a skalár  $\alpha_k > 0$  je délka kroku. V sestupných metodách (descent methods) hodnota účelové funkce monotonně klesá<sup>2</sup>,  $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ .

V každém kroku se k naší současné aproximaci přičte nějaký bod, který by ji měl posunout směrem cíli.

Nechť je funkce f diferencovatelná. Směr  $\mathbf{v}_k$  se nazývá **sestupný** v bodě  $\mathbf{x}_k$ , jestliže

$$f'(\mathbf{x}_k)\,\mathbf{v}_k < 0,\tag{9.2}$$

tedy směrová derivace ve směru  $\mathbf{v}_k$  je záporná. Pokud v bodě  $\mathbf{x}_k$  existuje sestupný směr, existuje  $\alpha_k > 0$  tak, že  $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ . Pokud v bodě  $\mathbf{x}_k$  sestupný směr neexistuje, vektor  $f'(\mathbf{x}_k)$  je nutně nulový (proč?) a tedy  $\mathbf{x}_k$  je stacionární bod.

Máme-li sestupný směr, optimální délku kroku  $\alpha_k$  najdeme minimalizací funkce f na polopřímce z bodu  $\mathbf{x}_k$  ve směru  $\mathbf{v}_k$ . Tedy minimalizujeme funkci jedné proměnné

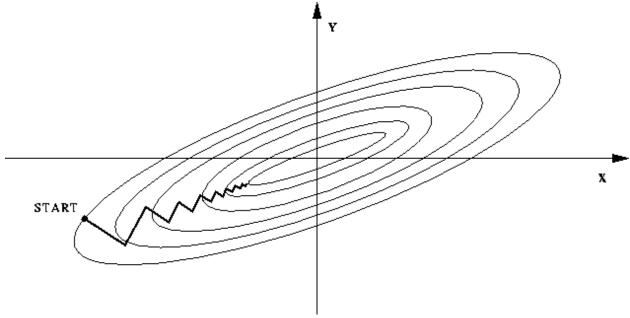
$$\varphi(\alpha_k) = f(\mathbf{x}_k + \alpha_k \mathbf{v}_k) \tag{9.3}$$

přes všechny  $\alpha_k \geq 0$ . Tato úloha je v kontextu vícerozměrné optimalizace nazývána jednoroz-měrné hledání (angl.  $line\ search$ ). Tuto úlohu stačí řešit přibližně, čímž se nebudeme zabývat

Sestupný směr je ten, pro který se posuneme z kopečka. Ten skalární součin v 9.2 znamená: derivace směřuje směrem růstu. Skalární součin je kladný, když vektory ukazují stejným směrem a záporný, když opačným. Když je skalární součin jakobiánu (derivace) a směru hledání záporný, je směr hledání sestupný (což chceme).

#### Gradientní metoda

Tato metoda volí jednoduchý směr hledání - záporný gradient. To znamená, že je vždy sestupný, protože jdeme vždy opačným směrem, než je směr největšího růstu. Nevýhoda je, že když máme třeba nějakou úzkou a dlouhou parabolu, tak může pomalu konvergovat (protože "skáče za strany na stranu"). To se stane když vlastní čísla hessiánu jsou velmi odlišná (ovlivňují natažení os elipsy).



#### Newtonova metoda

Taky metoda tečen, řeší soustavy nelineárních rovnic nebo minimalizaci funkce. Tj. hledá nulový bod nějakého zobrazení (které může představovat rovnici nebo soustavu). Když je to zobrazení derivace, tak vlastně hledáme lokální minimum, protože tam, kde je derivace nulová, je extrém.

### (1) Použití na soustavy - hledání kořenu

Nechť  $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^n$  je diferencovatelné zobrazení. Chceme řešit rovnici  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ , což je soustava n rovnic s n neznámými. Myšlenka Newtonovy metody je jednoduchá: místo hledání nulového bodu zobrazení  $\mathbf{g}$  (což je obecně velmi obtížné) opakujeme iteraci, která najde nulový bod afinní aproximace zobrazení  $\mathbf{g}$  v okolí aktuálního odhadu (což je snadné).

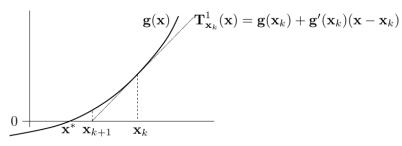
Afinní aproximace zobrazení  $\mathbf{g}$  v okolí bodu  $\mathbf{x}_k$  je zobrazení (viz (8.8) nebo (8.20))

$$\mathbf{T}_1(\mathbf{x}) = \mathbf{g}(\mathbf{x}_k) + \mathbf{g}'(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k). \tag{9.9}$$

Další odhad  $\mathbf{x}_{k+1}$  najdeme řešením nehomogenní lineární soustavy  $\mathbf{T}_1(\mathbf{x}_{k+1}) = \mathbf{0}$ . Pokud je Jacobiho matice  $\mathbf{g}'(\mathbf{x}_k)$  regulární, řešením je

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\mathbf{g}(\mathbf{x}_k). \tag{9.10}$$

Viz obrázek:



(2) Použití na hledání extrému - hledání stacionárního bodu

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T$$

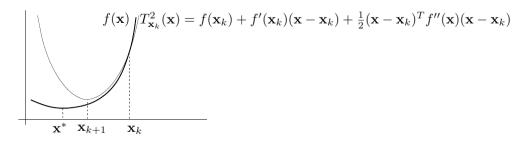
Hlavní výhodou Newtonovy metody je, že v blízkém okolí řešení obvykle konverguje velmi rychle (mnohem rychleji než gradientní metoda). Nevýhodou je, že je nutno začít s poměrně přesnou aproximací x0 skutečného řešení, jinak metoda snadno diverguje.

Newtonovu metodu lze použít pro hledání lokálního extrému dvakrát diferencovatelné funkce  $f: \mathbb{R}^n \to \mathbb{R}$  tak, že v metodě (9.10) položíme  $\mathbf{g}(\mathbf{x}) = f'(\mathbf{x})^T$ . Tím dostaneme iteraci

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T, \tag{9.11}$$

kde  $f''(\mathbf{x}_k)$  je Hessova matice funkce f v bodě  $\mathbf{x}_k$ . Neplet'te si tato dvě různá použití Newtonovy metody (tj. na hledání kořenů a na hledání lokálních extrémů)!

Iterace (9.10) se odvodila tak, že se zobrazení  $\mathbf{g}$  v okolí bodu  $\mathbf{x}_k$  aproximovalo afinním zobrazením  $\mathbf{T}_1$  a pak se našel nulový bod tohoto zobrazení. Lze ukázat (viz Cvičení 9.11), že iterace (9.11) lze odvodit také tak, že se funkce f aproximuje Taylorovým polynomem druhého stupně  $T_2$  (tedy kvadratickou funkcí) a pak se najde minimum této kvadratické funkce.



Iteraci (9.11) lze napsat v obecnějším tvaru (9.1), kde

$$\mathbf{v}_k = -f''(\mathbf{x}_k)^{-1} f'(\mathbf{x}_k)^T. \tag{9.12}$$

Výhodou tohoto zobecnění je možnost zvolit optimální (ne nutně jednotkovou) délku kroku  $\alpha_k$  pomocí jednorozměrné minimalizace (9.3). Metodě (9.11) s jednotkovou délkou kroku (tj.  $\alpha_k = 1$  pro každé k) se pak říká **čistá** Newtonova metoda.

Vektoru (9.12) říkáme **Newtonův směr**. Vidíme, že se od gradientního směru (9.4) liší násobením Hessovou maticí  $f''(\mathbf{x}_k)$ . Aby to byl sestupný směr, musí být

$$f'(\mathbf{x}_k)\mathbf{v}_k = -f'(\mathbf{x}_k)f''(\mathbf{x}_k)^{-1}f'(\mathbf{x}_k)^T < 0.$$

Toto platí, když  $f'(\mathbf{x}_k) \neq \mathbf{0}$  (tj.  $\mathbf{x}_k$  není stacionární bod) a matice  $f''(\mathbf{x}_k)$  je positivně definitní (neboť pak bude positivně definitní i její inverze, viz Cvičení 6.18).

V porovnání s gradientní metodou má Newtonova metoda (použitá na minimalizaci funkce) nevýhodu v tom, že musíme počítat Hessián  $f''(\mathbf{x}_k)$  a řešit soustavu  $f''(\mathbf{x}_k)\mathbf{v}_k = -f'(\mathbf{x}_k)^T$ , což pro velký počet proměnných je pomalé či nemožné. Všimněte si ale, že na rozdíl od §9.4.1 je zde matice  $\mathbf{g}'(\mathbf{x}_k) = f''(\mathbf{x}_k)$  symetrická, což může řešení soustavy ulehčit.

# 9.5 Nelineární metoda nejmenších čtverců

Mějme soustavu rovnic  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ , kde  $\mathbf{g}: \mathbb{R}^n \to \mathbb{R}^m$  (tedy je to soustava m rovnic s n neznámými). Soustavu nazveme přeurčenou, jestliže nemá žádné řešení. Chceme takovou přeurčenou soustavu řešit přibližně ve smyslu nejmenších čtverců. Tedy chceme minimalizovat funkci

$$f(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|^2 = \mathbf{g}(\mathbf{x})^T \mathbf{g}(\mathbf{x}) = \sum_{i=1}^m g_i(\mathbf{x})^2,$$
 (9.13)

kde  $g_i$  jsou složky zobrazení **g**. Speciálním případem je přibližné řešení lineární nehomogenní soustavy  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , kde  $\mathbf{g}(\mathbf{x}) = \mathbf{b} - \mathbf{A}\mathbf{x}$  (viz §5.1).

Zatímco v §9.3 a §9.4.2 bylo cílem minimalizovat *obecnou* funkci, zde chceme minimalizovat funkci ve speciálním tvaru (9.13). Nyní máme dvě možnosti. Buď můžeme nasadit na funkci (9.13) jednu z metod pro minimalizaci obecné funkce, k čemuž se vrátíme v §9.5.2. Nebo můžeme být chytřejší a využít speciálního tvaru funkce (9.13), což popíšeme v §9.5.1.

### 9.5.1 Gauss-Newtonova metoda

Aproximujme opět zobrazení  $\mathbf{g}$  v okolí bodu  $\mathbf{x}_k$  afinním zobrazením  $\mathbf{T}_1$  dle (9.9). Úloha (9.13) pak vyžaduje minimalizovat  $\|\mathbf{T}_1(\mathbf{x})\|^2$ . To je úloha lineárních nejmenších čtverců, kterou již známe z §5.1. Normální rovnice (5.3) má tvar

$$\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) = -\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}(\mathbf{x}_k). \tag{9.14}$$

Její řešení napišme pomocí pseudoinverze:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^+ \mathbf{g}(\mathbf{x}_k). \tag{9.15}$$

Metoda (9.15) je známa jako (čistá) **Gauss-Newtonova metoda**. Můžeme ji opět napsat obecněji ve tvaru (9.1) se směrem hledání (Gauss-Newtonův směr)

$$\mathbf{v}_k = -\mathbf{g}'(\mathbf{x}_k)^+ \mathbf{g}(\mathbf{x}_k). \tag{9.16}$$

Všmněte si, že pokud m = n a Jakobiho matice  $\mathbf{g}'(\mathbf{x}_k)$  je regulární, je  $\mathbf{g}'(\mathbf{x}_k)^+ = \mathbf{g}'(\mathbf{x}_k)^{-1}$ , tedy Gauss-Newtonova metoda (9.15) se redukuje na Newtonovu metodu (9.10).

Pokud má Jacobiho matice  $\mathbf{g}'(\mathbf{x}_k)$  lineárně nezávislé sloupce (viz §5.1), výraz (9.16) můžeme napsat jako

$$\mathbf{v}_k = -(\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k))^{-1} \mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}(\mathbf{x}_k) = -\frac{1}{2} (\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k))^{-1} f'(\mathbf{x}_k)^T, \tag{9.17}$$

kde ve výrazu na pravé straně jsme dosadili derivaci

$$f'(\mathbf{x}) = 2\mathbf{g}(\mathbf{x})^T \mathbf{g}'(\mathbf{x}) \tag{9.18}$$

účelové funkce (9.13) (odvoď<br/>te dle §8.4.2!). Vidíme, že Gauss-Newtonův směr (9.17) se liší od gradientního směru (9.4) pouze násobením matic<br/>í  $\frac{1}{2}(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}$ . Aby byl tento směr sestupný, musí být

$$f'(\mathbf{x}_k)\,\mathbf{v}_k = -\frac{1}{2}f'(\mathbf{x}_k)(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k))^{-1}f'(\mathbf{x}_k)^T < 0.$$

To platí, když  $f'(\mathbf{x}_k) \neq \mathbf{0}$  a matice  $\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k)$  je positivně definitní (viz Cvičení 6.18). Matice  $\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k)$  je positivně definitní, právě když  $\mathbf{g}'(\mathbf{x}_k)$  má lineárně nezávislé sloupce (viz Cvičení 6.19), což ovšem již předpokládáme kvůli existenci inverze. Tedy vidíme, že za přirozených podmínek je Gauss-Newtonův směr vždy sestupný.

Cistá Gauss-Newtonova metoda (tj. s jednotkovou délkou kroku) může divergovat, a to i když je počáteční odhad  $\mathbf{x}_0$  libovolně blízko lokálnímu minimu funkce (9.13). Protože ale Gauss-Newtonův směr je vždy sestupný, vhodnou volbou délky kroku  $\alpha_k$  lze vždy zajistit konvergenci.

Předpokládejme, že bychom optimalizovali naší účelovou funkci (9.13) přímo Newtonovou metodou z §9.4.2. Spočítejme (viz Cvičení 8.11.d) Hessián funkce (9.13):

$$f''(\mathbf{x}) = 2\mathbf{g}'(\mathbf{x})^T \mathbf{g}'(\mathbf{x}) + 2\sum_{i=1}^m g_i(\mathbf{x}) g_i''(\mathbf{x}). \tag{9.19}$$

Hessián je součtem členu obsahujícího derivace prvního rádu a členu obsahujícího derivace druhého řádu. Vidíme, že Gauss-Newtonův směr (9.17) se liší od Newtonova směru (9.12) zanedbáním členu druhého řádu v Hessiánu (9.19). Jinými slovy, Gauss-Newtonovu metodu je možno vnímat jako aproximaci Newtonovy metody na minimalizaci funkce (9.13) spočívající v tom, že zanedbáme členy druhého řádu, tedy skutečný Hessián (9.19) aproximujeme výrazem  $2\mathbf{g'}(\mathbf{x})^T\mathbf{g'}(\mathbf{x})$ .

To se projevuje tím, že Gauss-Newtonova metoda obvykle konverguje pomaleji než plná Newtonova metoda použitá na funkci (9.13). Ovšem vyhnuli jsme se počítání druhých derivací funkce  $\mathbf{g}$ , což je hlavní výhoda Gauss-Newtonovy metody.

### 9.5.3 Levenberg-Marquardtova metoda

Levenberg-Marquardtova metoda je široce používané vylepšení Gauss-Newtonovy metody, které její iteraci

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k))^{-1} \mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}(\mathbf{x}_k)$$
(9.20)

nahrazuje iterací

$$\mathbf{x}_{k+1} = \mathbf{x}_k - (\mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}'(\mathbf{x}_k) + \mu_k \mathbf{I})^{-1} \mathbf{g}'(\mathbf{x}_k)^T \mathbf{g}(\mathbf{x}_k)$$
(9.21)

kde  $\mu_k > 0$ . Přidání členu  $\mu_k \mathbf{I}$  je vlastně (Tichonovova) regularizace (viz §(5.17)). Potom:

- Pro malé  $\mu_k$  se (9.21) blíží Gauss-Newtonově iteraci.
- Pro velké  $\mu_k$  je  $(\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k) + \mu_k\mathbf{I})^{-1} \approx \mu_k^{-1}\mathbf{I}$ , tedy (9.21) je blízká iteraci  $\mathbf{x}_{k+1} = \mathbf{x}_k \frac{1}{2}\mu_k^{-1}f'(\mathbf{x}_k)^T$  (kde jsme dosadili (9.18)) gradientní metody s délkou kroku  $\mu_k^{-1}$ .

Tím jsou spojeny výhody Gauss-Newtonovy metody (typicky rychlá konvergence v okolí optima) a gradientní metody (spolehlivost i daleko od optima). Volbou parametru  $\mu_k$  spojitě přecházíme mezi oběma metodami.

Parametr  $\mu_k$  měníme se zvětšujícíc<br/>m se k pomocí jednoduché heuristiky. Začneme např. s<br/>  $\mu_0=10^3$  a pak v každé iteraci:

- Pokud iterace snížila účelovou funkci, iteraci přijmeme a  $\mu_k$  zmenšíme.
- Pokud iterace nesnížila účelovou funkci, iteraci odmítneme a  $\mu_k$  zvětšíme.

Zvětšování a zmenšování  $\mu_k$  děláme násobením a dělením konstantou, např. 10. Všimněte si, toto nahrazuje optimalizaci délky kroku  $\alpha_k$  (line search).

Motivaci pro přidání regularizace do Gauss-Newtonovy iterace (9.20) lze vidět i jinak. Matice  $\mathbf{g}'(\mathbf{x}_k)^T\mathbf{g}'(\mathbf{x}_k)$  může být singulární (to když sloupce  $\mathbf{g}'(\mathbf{x}_k)$  budou lineárně závislé) nebo blízká singulární. Pak její inverze neexistuje nebo je velmi citlivá na malé změny matice, což může neblaze ovlivnit konvergenci metody. Matice (9.21) je ale vždy positivně definitní (viz Cvičení 6.16), a tedy regulární.

## Extrémy vázané rovnostmi, Lagrangeovy multiplikátory

Hledejme minimum funkce  $f: \mathbb{R}^n \to \mathbb{R}$  na množině

$$X = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) = \mathbf{0} \}, \tag{10.1}$$

kde  $\mathbf{g} = (g_1, \dots, g_m) \colon \mathbb{R}^n \to \mathbb{R}^m$ . To odpovídá úloze (1.9) s omezeními typu rovnosti:

min 
$$f(x_1, ..., x_n)$$
  
za podmínek  $q_i(x_1, ..., x_n) = 0, \quad i = 1, ..., m.$  (10.2)

Mluvíme o minimu funkce f vázaném rovnostmi  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ .

Množina X je množina řešení soustavy m rovnic (obecně nelineárních) o n neznámých. Tato množina obvykle nemá žádné vnitřní body, proto nelze použít Větu 9.4. V některých případech ale lze vyjádřit všechna řešení soustavy  $\mathbf{g}(\mathbf{x}) = \mathbf{0}$  parametricky (neboli parametrizovat množinu X) a úlohu tak převést na úlohu bez omezení. To jsme použili v Příkladu 1.3, uveď me další příklady.

Věta 10.5. Nechť  $\mathbf{x} \in X$  je lokální extrém funkce f na množině X. Nechť f a  $\mathbf{g}$  jsou v bodě  $\mathbf{x}$  spojitě diferencovatelné. Nechť platí (10.12). Pak gradient  $\nabla f(\mathbf{x}) = f'(\mathbf{x})^T$  patří do ortogonálního prostoru množiny X v bodě  $\mathbf{x}$ , tj.

$$\nabla f(\mathbf{x}) \in \text{span}\{\nabla g_1(\mathbf{x}), \dots, \nabla g_m(\mathbf{x})\}.$$
 (10.14)

 $D\mathring{u}kaz$ . Protože  $\mathbf{x}$  je lokální extrém, dle Tvrzení 10.4 je gradient  $\nabla f(\mathbf{x})$  kolmý na každý vektor tečný k množině X v bodě  $\mathbf{x}$ . Protože platí (10.12), dle Věty 10.3 je tedy  $\nabla f(\mathbf{x})$  kolmý na tečný prostor null  $\mathbf{g}'(\mathbf{x})$ , neboli patří do ortogonálního prostoru (10.13).

Věta 10.5 se k řešení úlohy (10.2) obvykle používá následujícím způsobem. Podmínka (10.14) říká, že existují čísla  $(\lambda_1, \ldots, \lambda_m) = \lambda \in \mathbb{R}^m$  (tzv. Lagrangeovy multiplikátory) taková, že

$$f'(\mathbf{x}) + \lambda_1 g'_1(\mathbf{x}) + \dots + \lambda_m g'_m(\mathbf{x}) = f'(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}'(\mathbf{x}) = 0.$$
 (10.15)

Zaved'me Lagrangeovu funkci  $L: \mathbb{R}^{n+m} \to \mathbb{R}$  jako

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) = f(\mathbf{x}) + \lambda_1 g_1(\mathbf{x}) + \dots + \lambda_m g_m(\mathbf{x}).$$
(10.16)

Očividně je  $L(\mathbf{x}, \lambda) = f(\mathbf{x})$  pro všechna  $\mathbf{x} \in X$  a  $\lambda \in \mathbb{R}^m$ . Dále si všimneme, že:

- Podmínku (10.15) lze psát jako<sup>3</sup>  $\partial L(\mathbf{x}, \boldsymbol{\lambda})/\partial \mathbf{x} = \mathbf{0}$ .
- Rovnost  $\partial L(\mathbf{x}, \boldsymbol{\lambda})/\partial \boldsymbol{\lambda} = \mathbf{g}(\mathbf{x})^T = \mathbf{0}$  je ekvivalentní omezením.

Platí-li přepoklady Věty 10.5, existuje tedy vektor  $\lambda \in \mathbb{R}^m$  takový, že  $L'(\mathbf{x}, \lambda) = \mathbf{0}$ , tj. bod  $(\mathbf{x}, \lambda) \in \mathbb{R}^{m+n}$  je stacionární bod funkce L.

Věta 10.5 udává podmínky prvního řádu na extrémy vázané rovnostmi. Říká, že pokud  $(x, \lambda)$  je stacionární bod Lagrangeovy funkce, pak bod x je 'podezřelý' z lokálního extrému funkce f na množině X. Jak poznáme, zda tento bod je lokální extrém, případně jaký? Podmínky druhého řádu pro vázané extrémy uvádíme nepovinně v §10.4. Zde pouze zdůrazníme, že druh lokálního extrému nelze zjistit podle definitnosti Hessovy matice L"(x,  $\lambda$ ), tedy je chybou použít Větu 9.5 na funkci L.

## Lineární programování

Úloha **lineárního programování** (LP, také zvané lineární optimalizace) je minimalizace nebo maximalizace lineární funkce za omezujících podmínek ve tvaru lineárních rovnic a nerovnic. Zde **lineární rovnicí** rozumíme relaci

$$a_1x_1 + \cdots + a_nx_n = b$$
,

neboli krátce  $\mathbf{a}^T \mathbf{x} = b$ . Lineární nerovnicí rozumíme jednu z relací

$$a_1x_1 + \dots + a_nx_n \le b$$
,  $a_1x_1 + \dots + a_nx_n \ge b$ ,

neboli  $\mathbf{a}^T \mathbf{x} \leq b$  či  $\mathbf{a}^T \mathbf{x} \geq b$ . Lineární program je tedy úloha (1.9) ve které je funkce f lineární (tj. tvaru (3.6)) a funkce  $g_i, h_i$  jsou afinní (tj. tvaru (3.23)).

min 
$$f(x_1, \ldots, x_n)$$
  
za podmínek  $g_i(x_1, \ldots, x_n) \leq 0, \quad i = 1, \ldots, m$   
 $h_i(x_1, \ldots, x_n) = 0, \quad i = 1, \ldots, l$   
 $x_1, \ldots, x_n \in \mathbb{R}$  (1.9)

# Konvexní množiny a funkce

Funkce  $f: \mathbb{R}^n \to \mathbb{R}$  je **konvexní** na konvexní množině  $X \subseteq \mathbb{R}^n$ , jestliže

$$\mathbf{x} \in X, \ \mathbf{y} \in X, \ 0 \le \alpha \le 1 \implies f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) \le (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}).$$
 (15.1)

Funkce f je **konkávní** na množině X, jestliže je funkce -f konvexní na množině X. Rozlišujte pojmy konvexní množina a konvexní funkce, jde o různé věci. Všimněte si, že množina X musí být konvexní (pojem konvexní nebo konkávní funkce na nekonvexní množině není definován). Pokud X je celý definiční obor funkce f, odkaz na X můžeme vynechat a říkáme jen, že funkce f je konvexní.

Podmínku (15.1) lze zobecnit pro více než dva body: funkce f je konvexní, právě když

$$\begin{vmatrix}
\mathbf{x}_1, \dots, \mathbf{x}_k \in X \\
\alpha_1, \dots, \alpha_k \ge 0 \\
\alpha_1 + \dots + \alpha_k = 1
\end{vmatrix}
\implies f(\alpha_1 \mathbf{x}_1 + \dots + \alpha_k \mathbf{x}_k) \le \alpha_1 f(\mathbf{x}_1) + \dots + \alpha_k f(\mathbf{x}_k).$$
(15.2)