

ZÁPADOČESKÁ UNIVERZITA V PLZNI  
FAKULTA APLIKOVANÝCH VĚD  
KATEDRA MATEMATIKY



BAKALÁŘSKÁ PRÁCE  
POUŽITÍ GRADIENTNÍCH METOD V ÚLOHÁCH NA NELINEÁRNÍ  
NEJMENŠÍ ČTVERCE

## Čestné prohlášení

Tímto předkládám k posouzení a obhajobě svoji bakalářskou práci zpracovanou během třetího ročníku studia na Fakultě aplikovaných věd Západočeské univerzity v Plzni. Prohlašuji, že jsem bakalářskou práci vypracoval samostatně a výhradně s použitím odborné literatury a pramenů uvedených v seznamu na konci práce.

V Plzni, dne 20. května 2016.

.....  
vlastnoruční podpis

## Poděkování

Chtěl bych poděkovat vedoucí své práce Ing. Haně Kopincové, PhD. za pomoc při tvorbě práce, cenné rady, výbornou komunikaci a čas, který mi věnovala při konzultacích. Dále bych chtěl poděkovat Ing. Janu Pospíšilovi, PhD. za cenné rady a pomoc při experimentech a v neposlední řadě také rodině za podporu.

## Abstrakt

Tato bakalářská práce se zaměřuje na nepodmíněnou optimalizaci. V první části jsou popsány klasické gradientní metody a poté metody s proměnnou metrikou. Dále se práce soustředí na problematiku nelineárních nejmenších čtverců a metody, které se těmito úlohami zabývají. Ve třetí části jsou jednotlivé metody testované na kvadratické funkci, populační úloze a také na funkci z finančního prostředí, která se týká oceňování opcí, a následně porovnány.

**Klíčová slova:** nepodmíněná optimalizace, gradientní metody, metody s proměnnou metrikou, nelineární nejmenší čtverce

## Abstract

This bachelor thesis is focused on unconstrained optimization. The gradient methods and the variable metric methods are described in the first part. In the second part we study the problems of non-linear least squares. Then we perform some numerical experiments, where we can compare and evaluate several methods. We minimize the quadratic function, the population problem and the function regarding option pricing.

**Key words:** unconstrained optimization, gradient methods, variable metric methods, non-linear least squares

# Obsah

<b>1</b>	<b>Úvod</b>	<b>5</b>
<b>2</b>	<b>Základní teorie</b>	<b>6</b>
2.1	Optimalizace . . . . .	6
2.2	Teoretické pojmy . . . . .	6
<b>3</b>	<b>Klasické gradientní metody</b>	<b>9</b>
3.1	Metoda největšího spádu . . . . .	9
3.2	Metoda sdružených gradientů . . . . .	10
3.3	Line-Search metody . . . . .	11
3.4	Rychlost konvergence . . . . .	12
<b>4</b>	<b>Metody s proměnnou metrikou</b>	<b>14</b>
4.1	Obecný úvod . . . . .	14
4.2	Metoda DFP . . . . .	14
4.3	Metoda BFGS . . . . .	15
4.4	Broydenovy metody . . . . .	16
<b>5</b>	<b>Nelineární nejmenší čtverce</b>	<b>17</b>
5.1	Základní myšlenka . . . . .	17
5.2	Gauss-Newtonova metoda . . . . .	17
5.3	Vliv metod s proměnnou metrikou . . . . .	18
<b>6</b>	<b>Numerické experimenty</b>	<b>20</b>
6.1	Testování 1 . . . . .	20
6.2	Testování 2 . . . . .	22
6.3	Příklad z praxe . . . . .	25
<b>7</b>	<b>Závěr</b>	<b>28</b>
<b>A</b>	<b>Příloha</b>	<b>32</b>

# 1 Úvod

V práci se budeme zabývat metodami nepodmíněné optimalizace a speciálně se zaměříme na úlohu nejmenších čtverců. Optimalizační metody mají vysoké praktické využití, což si ukážeme na testovacích úlohách. Předmětem experimentování bude i funkce z finančního prostředí, kde jsou optimalizační metody také ve velkém měřítku používány.

Kapitola 2 bude obsahovat úvod do optimalizace a seznámení čtenáře s touto problematikou. Zdefinujeme si pojmy jako gradient, Hessova matice nebo Jacobiho matice, které budou v průběhu práce často zmiňovány a s nimiž budeme pracovat v dalších kapitolách.

Hlavním předmětem kapitoly 3 budou klasické gradientní metody. Zahrnujeme sem i nejjednodušší třídu optimalizačních metod, a to metodu největšího spádu. V této metodě je největší spád ve směru gradientu, proto může být zmiňována v této kapitole společně s dalšími metodami, jako je například metoda sdružených gradientů. Tato metoda slouží k odstranění nevýhod metody největšího spádu, protože je založena na principu konjugovanosti, což má za následek urychlení konvergence oproti metodě největšího spádu [9]. První nelineární metoda sdružených gradientů byla popsána od autorů Fletchera a Reevese v práci [14]. V této kapitole si popíšeme i modifikace metod sdružených gradientů, a to Fletcher-Reevesovu [14] a Polak-Ribierovu [15]. V závěru kapitoly se zmíníme o rychlosti konvergence iteračních metod.

V následující kapitole 4 se zaměříme na metody s proměnnou metrikou, patřící mezi nejeftivnější metody pro řešení menších úloh. Principem těchto metod je aproximace Hessovy matice nebo aproximace její inverze. Po úvodu do těchto metod popíšeme metodu DFP, kterou William C. Davidon popsal už v roce 1959, ale objevila se až v roce 1991 v odborném časopise *SIAM Journal on Optimization* [16]. Další metodou s proměnnou metrikou je metoda BFGS, pojmenována po autorech Broyden [17], Fletcher [18], Goldfarb [19] a Shanno [20].

V kapitole 5 charakterizujeme úlohu nejmenších čtverců. Nejmenší čtverce znamenají, že výsledné řešení má minimalizovat součet čtverců odchylek vůči každé rovnici. [21] Hlavní metodou na problematiku nejmenších čtverců je Gauss-Newtonova metoda. Po popisu algoritmu se zaměříme na kombinace s metodou s proměnnou metrikou, konkrétně kumulativní korekci s metodou BFGS [22].

V následující kapitole 6 jsou provedeny na vybraných metodách z předchozích kapitol numerické experimenty. Prvním experimentem je porovnání několika metod na jednoduché kvadratické funkci dvou proměnných. Mezi zkoumaná kritéria patří souřadnice minima, funkční hodnota v minimu, celkový počet provedených iterací a čas výpočtu. V druhém experimentu je testována úloha na nelineární nejmenší čtverce, konkrétně úloha týkající se živočišné populace. Výsledky jsou opět porovnány a vše je doplněno ilustrativními obrázky. Třetí experiment směřuje do finančního prostředí, jelikož se týká oceňování opcí a hledání neznámých parametrů aproximující funkce.

V závěru je práce shrnuta, jsou okomentovány výsledky a zmíněny náměty pro případné rozšíření práce.

## 2 Základní teorie

### 2.1 Optimalizace

Optimalizace patří mezi matematické disciplíny a zabývá se hledáním extrému (minimalizací či maximalizací) funkcí více proměnných za případných omezujících podmínek. Jak už toto vysvětlení napovídá, nachází optimalizace široké využití v praktickém životě, zejména pak ve finančním sektoru, kde je při snaze minimalizovat náklady či maximalizovat zisk vysoce využívána.

Označme

$x \in \mathbb{R}^n$  - vektor proměnných

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ , je cílová (účelová) funkce neboli funkce proměnné  $x$ , kterou chceme optimalizovat.

**Definice 1** *Nechť je funkce  $f$ , dvakrát spojitě diferencovatelná, definovaná na prostoru*

$$f : \mathbb{R}^n \rightarrow \mathbb{R}.$$

*Úlohu, která řeší problém*

$$\min_{\forall x \in \mathbb{R}^n} f(x),$$

*nazýváme nepodmíněnou optimalizací.*

Více o nepodmíněné optimalizaci k nalezení zde [1].

**Poznámka 1** *Funkce  $f$  je dvakrát spojitě diferencovatelná. Tato vlastnost znamená, že je spojitá první i druhá derivace funkce  $f$ .*

### 2.2 Teoretické pojmy

V této části si zadefinujeme některé teoretické pojmy z oblasti numerické matematiky, které budou dále používány a zmiňovány.

**Definice 2** *Nechť  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  je funkce dvou proměnných. Řekneme, že funkce má v bodě  $(x_0, y_0)$  **parciální derivaci** podle proměnné  $x$  rovnou číslu  $f_x(x_0, y_0)$ , jestliže existuje konečná limita*

$$f_x(x_0, y_0) = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x, y) - f(x_0, y)}{\Delta x}.$$

[2]

**Poznámka 2** *Ekvivalentním značením parciální derivace  $f'_{x_i}$  je zápis  $\frac{\partial f}{\partial x}$ .*

**Definice 3** *Nechť  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $a \in Df$ . Nechť existuje  $\frac{\partial f(a)}{\partial x_i}$  pro  $\forall i = 1, \dots, n$ .*

Vektor

$$\text{grad}f(a) = \left( \frac{\partial f(a)}{\partial x_1}, \dots, \frac{\partial f(a)}{\partial x_n} \right) \quad (1)$$

se nazývá *gradient funkce*  $f$  v bodě  $a$ .

Vektor

$$\text{grad}f = \left( \frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n} \right) \quad (2)$$

se nazývá **gradient** funkce  $f$  neboli vektor parciálních derivací, který udává směr a velikost největšího růstu funkce.

Podrobnější popis gradientu a jeho vlastností dostupný zde [1, 9].

**Poznámka 3** Gradient značíme  $\text{grad} f$  nebo také  $\nabla f$ , symbol  $\nabla$  se nazývá nabla operátor.

### Vlastnosti gradientu:

Zavedeme si  $F, G$  jako vektorová pole,  $f, g$  jako funkce a  $a, b$  libovolná reálná čísla. Potom má gradient následující vlastnosti:[4]

- je lineární vůči reálným číslům  
 $\nabla(af + bg) = a \nabla f + b \nabla g$ ,
- splňuje Leibnizovo pravidlo pro funkce aneb chování operátorů vůči součinu  
 $\nabla(fg) = (\nabla f)g + f \nabla g$ ,
- gradient skalárního součinu vektorů splňuje vlastnost  
 $\nabla(F \cdot G) = \nabla F \cdot G + \nabla G \cdot F$ .

Dále se v práci budeme zabývat operacemi, pro které je důležité znát maticový počet. Matice budou tvořené parciálními derivacemi, prvního a druhého řádu. [5]

**Definice 4** Necht' je  $h = (h_1, h_2, \dots, h_n)$  diferencovatelné zobrazení z  $\mathbb{R}^n \rightarrow \mathbb{R}^m$  na otevřené množině  $\Omega \in \mathbb{R}^n$ . Matici parciálních derivací  $J_f(x)$  pro  $x \in \Omega$  ve tvaru

$$J_f(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \dots & \frac{\partial f_m(x)}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1(x)}{\partial x_n} & \dots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

nazýváme **Jacobiho maticí** zobrazení  $f$ .

**Poznámka 4** Termínem *Jakobián* se nejčastěji označuje determinant Jacobiho matice.

**Definice 5** Necht' je funkce  $f(x)$  dvakrát diferencovatelná funkce z  $\mathbb{R}^n \rightarrow \mathbb{R}$ , potom matici parciálních derivací druhého řádu

$$H(x) \equiv \frac{\partial^2 f}{\partial x_i \partial x_j}(x) \equiv \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n \partial x_n}(x) \end{bmatrix}$$



nazýváme **Hessovou maticí**.

Jestliže jsou parciální derivace druhého řádu v Hessově matici v bodě  $x_0$  spojité, jsou smíšené derivace záměnné, což znamená, že je Hessova matice symetrická, řádu  $n$ . (Podle  $n$  proměnných).

Hlavním úkolem optimalizačních metod je nalezení extrémní, ať už minimální, či maximální hodnoty. Následující definice se týkají především matic a jejich vlastností.

**Definice 6** *Nechť existuje  $\varepsilon > 0$  tak, že pro všechna  $x \in \{x \in \mathbb{R}^n; 0 \leq \|x - a\| \leq \varepsilon\}$  je  $f(x) < f(a)$ . Potom říkáme, že funkce nabývá v bodě  $a$  svého ostrého lokálního maxima. Pokud nahradíme znaménko mezi funkčními hodnotami za opačné, dostaneme předpis pro ostré lokální minimum.*

**Definice 7** *Bod, ve kterém je funkce diferencovatelná, a který splňuje nutnou podmínku existence lokálního extrému, což znamená, že jeho gradient je rovní nule, nazýváme stacionárním bodem.*

**Definice 8** *O symetrické matici  $n \times n$  můžeme říci, že je **pozitivně definitní**, právě tehdy, když pro každý nenulový vektor  $x$  dimenze  $n$  je jeho kvadratická funkce  $x^T Ax$  kladná a nulová nastane pouze v případě nulového vektoru. **Negativně definitní** nazveme matici se zápornou kvadratickou funkcí a nulovou jen pro nulový vektor. Pokud je funkce nezáporná, resp. nekladná, a existuje nulová v případě nulového vektoru, mluvíme o **pozitivně**, resp. **negativně semidefinitní** matici. V případě existence nenulových vektorů  $x_1, x_2$  tak, že:*

$$x_1^T Ax_1 > 0 \text{ a zároveň } x_2^T Ax_2 < 0,$$

*hovoříme o **indefinitní** matici.*

V naší práci se zaměříme na iterační metody nepodmíněné optimalizace. Na začátku máme počáteční bod  $x_1 \in \mathbb{R}^n$ , generující se posloupnost bodů  $x_i \in \mathbb{R}^n$ , kde  $i \in N$ . Proces iterace vysvětluje rovnice

$$x_{i+1} = x_i + \alpha_i d_i, \tag{3}$$

kde

$d_i \in \mathbb{R}^n$  je směrový vektor

a  $\alpha_i \geq 0$  je délka kroku. ( $N$  je množina přirozených čísel)

Nejefektivnější optimalizační metody řadíme do třech skupin:

- Metody sdružených gradientů
- Metody s proměnnou metrikou
- Modifikované Newtonovy metody

### 3 Klasické gradientní metody

V této části se seznámíme s iteračními algoritmy řešící klasický optimalizační problém

$$\min_{x \in \mathbb{R}^n} f(x),$$

kde je v každém kroku využíván směr poklesu hodnoty účelové funkce  $f$  k nalezení další iterace s ještě menší hodnotou  $f$ . Klasické gradientní metody využívají jako směr poklesu záporný gradient. Spádové směry jsou poměrně jednoduché na implementaci, mají nízkou cenu iterace a dají se aplikovat na řešení mnoha úloh. Jejich nevýhodou je velký počet iterací, jelikož nevyužívají explicitně informaci z předchozích iterací. Příkladem gradientní metody je metoda největšího spádu.

#### 3.1 Metoda největšího spádu

**Definice 9** Směr  $d_k$  nazveme spádovým, pokud  $g_k^T d_k < 0$ , tj. pokud je derivace funkce  $f$  v bodě  $x_k$  ve směru  $d_k$  záporná neboli klesá. Spádové směry jsou všechny směry  $d_k$ , které s vektorem  $-g_k$  svírají úhel menší než 90 stupňů. [1]

Tyto metody byly primárně odvozeny pro kvadratický problém

$$\min_{x \in \mathbb{R}^n} f(x), f(x) = \frac{1}{2} x^T A x - x^T b, \quad (4)$$

kde  $A$  je symetrická, pozitivně definitní matice (viz definice 8),  $f$  je ostře konvexní funkce a jednoznačné globální minimum je ve stacionárním bodě  $x^*$ , který splňuje rovnici

$$Ax^* = b.$$

Metodu největšího spádu lze použít i na nekvadratický případ, kde musíme řešit dva hlavní problémy:

- problém, jak vypočítat gradient,
- problém minimalizace v daném směru.

Detailnější popis této metody k nalezení v [1, 3].

#### Algoritmus metody největšího spádu

1. Je dána diferencovatelná funkce  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $x_0 \in \mathbb{R}^n$ , dále gradient  $g_0 = \nabla f(x_0)$ , koeficient  $k = 0$  a zastavovací podmínka  $\epsilon > 0$ .
2. **while**  $\|g_k\| > \epsilon$ 
  - (a) Vyber  $\alpha_k > 0$ , kde  $\alpha_k$  je vhodná délka kroku
  - (b)  $x_{k+1} = x_k - \alpha_k g_k$
  - (c)  $g_{k+1} = \nabla f(x_{k+1})$
  - (d)  $k = k + 1$
  - (e) **end**
3. Přibližné řešení  $x^* = x_k$ .

### 3.2 Metoda sdružených gradientů

Tato metoda byla vyvinuta za účelem odstranění nevýhod u metody největšího spádu (3.1) a Newtonovy metody. Byla sestavena pro čistě kvadratický problém,

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} x^T A x - x^T b, \quad (5)$$

kde  $A$  je symetrická pozitivně definitní matice (viz definice 8).

Globální minimum bude ve stacionárním bodě  $\mathbf{x}^*$ , který vyhovuje rovnici

$$A x^* = b.$$

**Definice 10** *Nechť  $A \in \mathbb{R}^{n \times n}$  je symetrická pozitivně definitní matice, nenulové vektory  $\mathbf{x}, \mathbf{y}$  nazveme sdružené vzhledem k  $A$  či  $A$ -ortogonální, pokud*

$$\mathbf{y}^T A \mathbf{x} = 0.$$

Je-li  $A$  symetrická, pozitivně definitní matice a nenulové vektory  $d_0, d_1, \dots, d_k$   $A$ -ortogonální (viz definice 10), pak jsou vektory  $d_0, d_1, \dots, d_k$  lineárně nezávislé.

Tyto metody využívají pouze  $n$ -dimenzionální vektory. Směrové vektory  $d_i \in \mathbb{R}^n, i \in N$  jsou generovány tak, že  $d_1 = -g_1$  a

$$d_{i+1} = -g_{i+1} + \beta_i d_i, \quad (6)$$

kde  $g_{i+1} = g(x_{i+1})$  je gradient funkce  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  v bodě  $x_{i+1}$  a  $\beta_i$  je vhodně definovaný skalární parametr.

Metody sdružených gradientů potřebují pouze  $O(n)$  operací na jednu iteraci, ale oproti metodám s proměnnou metrikou potřebují iterací více. Samozřejmě jsou ale tyto iterace levnější. Metody sdružených gradientů jsou vyvinuty pro řešení velkých úloh.

#### Algoritmus metody sdružených gradientů

Na počátku máme symetrickou, pozitivně definitní matici  $A \in \mathbb{R}^{n \times n}$  a  $b \in \mathbb{R}^n$ .

1. Zvolíme  $\epsilon > 0$  (zastavovací podmínka), počáteční  $x_0 \in \mathbb{R}^n$  a označíme  $g_0 = Ax_0 - b, p_1 = g_0, k = 1$ .
2. Základní cyklus pro sdružené gradienty:
  - (a) **while**  $\|g_{k-1}\| > \epsilon$
  - (b)  $\alpha_k = \|g_{k-1}\|^2 / p_k^T A p_k$
  - (c)  $x_k = x_{k-1} - \alpha_k p_k$
  - (d)  $g_k = g_{k-1} - \alpha_k A p_k$
  - (e)  $\beta_k = \|g_k\|^2 / \|g_{k-1}\|^2 = \frac{(g_k)^T (g_k - (g_{k-1}))}{\|g_{k-1}\|^2}$
  - (f)  $p_{k+1} = g_k + \beta_k p_k$
  - (g)  $k = k + 1$

### 3. Řešení $x^* = x_k$ .

Podrobnější popis gradientu a jeho vlastností k nalezení zde [3, 11].

Metoda sdružených gradientů se dá použít i pro obecnější funkce než kvadratické. Stačí aby funkce byla diferencovatelná a už budeme schopni gradient buďto najít nebo alespoň aproximovat.

Před řešením nekvadratických úloh je nutné udělat nějaké úpravy v algoritmu pro případy kvadratické. Například určení délky kroku  $\alpha_k$  vypočteme pomocí jednorozměrné minimalizace funkce  $f$  v bodě  $x_{k-1}$  ve směru  $-p_k$ , tj.

$$\alpha_k = \arg \min_{\alpha > 0} f(x_{k-1} - \alpha p_k) \quad (7)$$

Toto minimalizační pravidlo je podrobněji popsáno v [3].

Další úpravou je nahrazení rezidua  $g_{k-1} - \alpha_k A p_k$ , tj. gradientu kvadratické funkce

$$f(x) = \frac{1}{2} x^T A x - x^T b$$

gradientem nelineární funkce  $f$ , tj.

$$g_k = \nabla f(x_k). \quad (8)$$

U nekvadratických příkladů je matice  $A = H(x_k)$ , kde  $H(x_k)$  je Hessova matice druhých partiálních derivací. A je velmi nepraktické počítat Hessovu matici nově v každém kroku. Proto byly vyvinuty dvě metody, **Fletcher-Reevesova** a **Polak-Ribierova**, kde parametr  $\alpha_k$  je odhadnut jednorozměrnou minimalizací, například pomocí Line-Search metody (viz podkapitola 3.3), zatímco parametr  $\beta_k$  je odhadnut každou metodou jinak.

V případě kvadratických funkcí se vztahy pro výpočet  $\beta_k$  rovnají, čili

$$\beta_k = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} = \frac{(g_k)^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}. \quad (9)$$

U nelineárních úloh tyto vztahy nedávají stejnou hodnotu, liší se následovně: (detailní popis obou metod viz [3, 9, 11])

- Fletcher-Reevesova metoda

$$\beta_k = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} = \frac{g_{k+1}^T - g_{k+1}}{g_k^T g_k}. \quad (10)$$

- Polak-Ribierova metoda

$$\beta_k = \frac{(g_k)^T (g_k - (g_{k-1}))}{\|g_{k-1}\|^2} = \frac{(g_{k+1} - g_k)^T - g_{k+1}}{g_k^T g_k}. \quad (11)$$

### 3.3 Line-Search metody

Line-search metody vyžadují, aby vektory  $d_i \in \mathbb{R}^n$ ,  $i \in N$  byly spádové (viz definice 9) a délka kroku  $\alpha_i \in \mathbb{R}^n$ ,  $i \in N$  byla zvolena tak, aby splnila podmínku  $\alpha_i \geq 0$  a zároveň byla dodržena nerovnice

$$F_{i+1} - F_i \leq \epsilon_1 \alpha_i g_i^T d_i \quad (12)$$

a dále musí být splněna nerovnice

$$g_{i+1}d_i \geq \epsilon_2 g_i^T d_i, \quad (13)$$

kde konstanty  $\epsilon_1, \epsilon_2$  splňují  $0 \leq \epsilon_1 \leq \frac{1}{2}$  a  $\epsilon_1 \leq \epsilon_2 \leq 1$ . Takto byly shrnuty tzv. **Wolfeho podmínky**, důležité pro line-search minimalizaci. Následující věta charakterizuje globální konvergenci line-search metod. [8]

**Věta 1** *Nechť funkce  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  je zdola omezená a má omezené derivace druhého řádu. Uvažujeme line-search metody s iteračním procesem (viz rovnice (3)) a  $d_i, \alpha_i$  vyhovující nerovnicím (12) a (13).*

*Pokud*

$$\sum_{i=N} c_i^2 = \infty, \quad (14)$$

*potom  $\liminf_{i \rightarrow \infty} \|g_i\| = 0$ .*

### 3.4 Rychlost konvergence

Rychlost konvergence iteračních metod můžeme srovnávat pomocí jistého jejího asymptotického vyjádření. Označme odchylku od řešení v  $k$ -té iteraci jako  $h_k$ , tj.

$$h_k = x_k - x^*.$$

Pokud  $h_k \rightarrow 0$ , pak řekneme, že konvergence metody je  $p$ -tého řádu, když

$$\frac{\|x_{k+1}\|}{\|x_k\|^p} \rightarrow a, a > 0.$$

Z praktického pohledu jsou důležité jen dva případy, kdy  $p = 1$  (lineární konvergence) a  $p = 2$  (kvadratická konvergence), protože nejsou známy žádné obecně použitelné numerické metody s vyšším řádem konvergence. Podle definice pro každou lineárně konvergentní metodu existuje konstanta  $\alpha > 0$  tak, že

$$\frac{\|h_{k+1}\|}{\|h_k\|} < \alpha$$

neboli

$$h_{k+1} = O(\|h_k\|).$$

Obdobně pro metodu s kvadratickou konvergencí platí tvar

$$\frac{\|h_{k+1}\|}{\|h_k\|^2} < \alpha$$

neboli

$$h_{k+1} = O(\|h_k\|^2).$$

Co se týče použití těchto metod, tak metody s lineární konvergencí jsou obvykle používány za dostatečně malé konstanty  $\alpha$ , například  $\alpha < 1/4$ .

Řada minimalizačních metod, které nejsou kvadraticky konvergentní, konverguje rychleji než jak to zaručuje lineární konvergence. Podíl po sobě jdoucích odchylek  $h_k$  jde pro  $k \rightarrow \infty$  k nule, tj.

$$\frac{\|h_{k+1}\|}{\|h_k\|} \rightarrow 0$$

neboli

$$h_{k+1} = O(\|h_k\|).$$

V takovém případě hovoříme o **superlineární** konvergenci. K získání dalších informací o spádových metodách slouží [6, 7].

## 4 Metody s proměnnou metrikou

### 4.1 Obecný úvod

Metody s proměnnou metrikou patří mezi jedny z neefektivnějších metod pro řešení optimalizačního problému v úlohách s menšími rozměry. Tyto metody jsou založeny na principu aktualizace matic, které aproximují Hessovu matici, popřípadě její inverzi. Tato metoda snaží konstruovat inverzi Hessovy matice nebo aproximaci inverze Hessovy matice pomocí informací získaných během iteračního procesu.

**Poznámka 5** *Úlohou o menších rozměrech se rozumí úloha do 250 proměnných.*

Metody s proměnnou metrikou jsou založeny na lokálním kvadratickém modelu

$$Q_i(d) = \frac{1}{2}d^T B_i d + g_i^T d, \quad (15)$$

kde  $B_i$  je nějaká pozitivně definitní aproximace Hessovy matice  $H_i$ . Matice  $B_i, i \in N$  jsou konstruovány iteračně, takže  $B_1$  je libovolná pozitivně definitní matice a  $B_{i+1}$  je určena na základě  $B_i$  tak, že je pozitivně definitní, co možná nejpodobnější  $B_i$  a vyhovuje kvazi-Newtonovým podmínkám

$$B_{i+1}s_i = y_i, \quad (16)$$

kde  $s_i = x_{i+1} - x_i$  a  $y_i = g_{i+1} - g_i$ .

Metody s proměnnou metrikou mají některé výhody oproti modifikovaným Newtonovým metodám. Matice  $B_i$  jsou pozitivně definitní, tím pádem metody s proměnnou metrikou jsou globálně konvergentní. Navíc, můžeme aktualizovat inverzi

$$H_i = B_i^{-1} \quad (17)$$

nebo pomocí tzv. Choleskyho rozkladu

$$L_i D_i L_i^T = B_i \quad (18)$$

místo samotné matice  $B_i$  a bude nám na to stačit pouze  $O(n^2)$  operací na iteraci. Ačkoli metody s proměnnou metrikou využívají více iterací než modifikované Newtonovy metody, tak jsou více vhodné pro malé a středně velké problémy obsahující husté matice.

### 4.2 Metoda DFP

Metoda pojmenována po autorech Davidon, Fletcher, Powell. Updatujeme pomocí matice  $H_0$  hodnoti 2.

$$H_{k+1} = H_k + \frac{p_k p_k^T}{p_k^T q_k} - \frac{H_k q_k q_k^T H_k}{q_k^T H_k q_k}. \quad (19)$$

**Algoritmus metody DFP:**

1. Je dáno počáteční  $x_0$ , symetrická pozitivně definitní matice  $H_0$ , určíme  $g_0, k = 0$ .

(a)  $d_k = -H_k g_k$

- (b) musíme určit vhodné  $\alpha_k$ , které minimalizuje  $f(x_k + \alpha_k d_k)$
- (c)  $x_{k+1} = x_k + \alpha_k d_k$ , kde  $\alpha_k$  splňuje Wolfeho podmínky (viz podkapitola 3.3)
- (d) vypočteme  $g_{k+1}$
- (e)  $p_k = \alpha_k d_k$
- (f)  $g_k = g_{k+1} - g_k$

2. provedeme update  $H_{k+1}$  pomocí metody DFP.

Platí zde implikace o dodržení vlastností: Pokud  $H_k$  je pozitivně definitní  $\Rightarrow H_{k+1}$  je také pozitivně definitní.

### 4.3 Metoda BFGS

Nejpopulárnější kvazi-Newtonova metoda, pojmenovaná po svých autorech Broyden, Fletcher, Goldfarb a Shannon. Myšlenkou této metody je updatem pomocí matice  $H_0$  hodnoti 2 aproximovat přímo Hessovu matici, nikoli její inverzi. [1]

Iterační formule pro výpočet matice, pomocí níž aproximujeme Hessovu matici, vypadá takto:

$$B_{k+1} = B_k + \frac{q_k q_k^T}{q_k^T p_k} - \frac{B_k p_k p_k^T B_k}{p_k^T B_k p_k}. \quad (20)$$

Aproximace inverze Hessovy matice je definována jako:

$$H_{k+1}^{BFGS} = B_{k+1}^{-1}. \quad (21)$$

#### Algoritmus metody BFGS:

1. Je dáno počáteční  $x_0$ , symetrická pozitivně definitní matice  $H_0$ , úkolem spočítat  $g_0$ ,  $k = 0$ .
  - (a)  $d_k = -H_k g_k$
  - (b) line search  $f(x_k + \alpha_k d_k)$
  - (c)  $x_{k+1} = x_k + \alpha_k d_k$ , kde  $\alpha_k$  splňuje Wolfeho podmínky (viz podkapitola 3.3)
  - (d) vypočteme  $g_{k+1}$
  - (e)  $p_k = \alpha_k d_k$
  - (f)  $g_k = g_{k+1} - g_k$
2. provedeme update  $H_{k+1}$  pomocí metody BFGS.

I zde platí implikace o pozitivní definitnosti jako u metody DFP.



#### 4.4 Broydenovy metody

Tyto metody také patří do metod s proměnnou metrikou. Broydenovy metody zachycují myšlenku, že obě dvě předchozí metody, BFGS (viz podkapitola 4.3) a DFP (viz podkapitola 4.2), používají symetrický update.

Existuje více typů Broydenových metod. V naší práci zmíníme typ, který je jakousi lineární kombinací mezi dříve zmíněnými metodami, a to metodou BFGS a metodou DFP

$$H^\theta = (1 - \theta)H^{DFP} + \theta H^{BFGS}, \quad (22)$$

přičemž parametr  $\theta$  nemusí být v každém kroku stejný, lze ho libovolně měnit. Více informací k Broydenovým metodám k nalezení v [1, 7, 9].

## 5 Nelineární nejmenší čtverce

V této části práce bude čerpáno výhradně z článku od autorů L. Lukšana a E. Spedicata. [8]

### 5.1 Základní myšlenka

Předpokládáme, že cílová funkce

$$f : \mathbb{R}^n \rightarrow \mathbb{R}$$

má tvar

$$f(x) = \frac{1}{2} f^T(x) f(x) = \frac{1}{2} \sum_{k=1}^m f_k^2(x), \quad (23)$$

kde jsou funkce  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , při hodnotě  $k$ , ležící v intervalu  $1 \leq k \leq m$  dvakrát spojitě diferencovatelné.

Tato cílová funkce je často využívána v úlohách nelineární regrese a pro řešení systému nelineárních rovnic. Gradient a Hessovu matici můžeme vyjádřit ve formě

$$g(x) = J^T(x) f(x) = \sum_{k=1}^m f_k(x) g_k(x), \quad (24)$$

$$H(x) = J^T(x) J(x) + C(x) = \sum_{k=1}^m g_k(x) g_k^T(x) + \sum_{k=1}^m f_k(x) G_k(x), \quad (25)$$

kde  $g(x)$  a  $H(x)$  jsou gradient a Hessova matice funkce

$$f_k : \mathbb{R}^n \rightarrow \mathbb{R},$$

$$1 \leq k \leq m, f^T(x) = [f_1(x), \dots, f_m(x)], J^T(x) = [g_1(x), \dots, g_m(x)],$$

kde  $J^T(x)$  je Jacobiho matice (viz definice 4) zobrazující funkci  $f$  v bodě  $x$ .

### 5.2 Gauss-Newtonova metoda

Gauss-Newtonova metoda patří mezi nejznámější metodu pro nelineární nejmenší čtverce. Tato metoda vznikne z Newtonovy metody tak, že z výrazu (25) pro Hessovu matici  $H(x)$  zanedbá člen  $C(x)$ , protože pokud se nacházíme v blízkosti minima, blíží se tento člen k nule. [9, 11] Předpis pro Hessovu matici vypadá takto:

$$B_k = J_k^T J_k = \sum_{k=1}^m g_k(x) g_k^T(x), \forall k \in N \quad (26)$$

#### Algoritmus Gauss-Newtonovy metody

1. Je dána diferencovatelná funkce  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , počáteční bod  $x_0 \in \mathbb{R}^n$  a zastavovací podmínka  $\epsilon > 0$ . Dále je z funkce  $f$  určena Jacobiho matice, která bude sloužit pro aproximaci Hessovy matice.

2. **while**  $\|x_{k+1} - x_k\| > \epsilon$ 
  - (a)  $g_k = J^T(x_{k+1})f(x_{k+1})$
  - (b)  $-g_k = B_k s_k$ , kde  $s_k$  je směrový vektor
  - (c)  $x_{k+1} = x_k + s$
  - (d)  $k = k + 1$
  - (e) **end**
3. Přibližné řešení  $x^* = x_k$ .

### 5.3 Vliv metod s proměnnou metrikou

Pro velké funkce může být předpis pro aproximaci Hessovy matice z rovnice (26) aproximací špatnou, nepřesnou. Z těchto důvodů mohou být výhodnější metody kombinované, a to mezi Gauss-Newtonovou metodou a nějakou metodou s proměnnou metrikou. Kombinace těchto metod mohou být velmi efektivní. Vybraná metoda s proměnnou metrikou musí splňovat kvazi-Newtonovy podmínky a dále musí být upřesněno její kombinování s Gauss-Newtonovou metodou. [8, 9]

Hlavní strategie kombinování Gauss-Newtonovy metody a metody s proměnnou metrikou je vysvětlena následující nerovnicí, kde dochází k porovnání funkčních hodnot dvou po sobě jdoucích iterací  $x_k$  a  $x_{k+1}$ .

$$F(x_k) - F(x_{k+1}) \leq \Theta F(x_k), \quad (27)$$

kde parametr  $\Theta$  je libovolná hodnota vyhovující intervalu  $0 < \Theta < 1$ .

Pokud je tato nerovnice splněna, následující iterace  $x_{k+1}$  je vypočtena pomocí metody s proměnnou metrikou, v opačném případě Gauss-Newtonovou metodou.[9] V článku [8] je popsána mj. kumulativní korekce, kdy je aproximace Hessovy matice  $B_{k+1}$  při splnění nerovnice (27) řešena metodou s proměnnou metrikou (v případě našich numerických experimentů se jedná o metodu BFGS), v opačném případě je následující aproximace Hessovy matice řešena již zmíněnou Gauss-Newtonovou, a to

$$B_k = J_k^T J_k, \forall k \in N.$$

Nyní si popíšeme algoritmus výše popsaného kombinování metod (kumulativní korekci):

#### Algoritmus Gauss-Newtonovy v kombinaci s BFGS

Je dána funkce  $f$ , dále počáteční bod  $x_0 \in \mathbb{R}^n$ , uživatelem zvolená hodnota  $\Theta$  a zastavovací podmínka  $\epsilon$ . První iterace Hessovy matice  $H_k$  je vypočtena Gauss-Newtonovou metodou.

1. **while**  $\|x_{i+1} - x_i\| > \epsilon$ 
  - (a)  $g_k = J^T(x_{k+1}) * f(x_{k+1})$

(b)  $s = x_{k+1} - x_k$ , kde  $s$  je směrový vektor

(c) **if**  $F(x_k) - F(x_{k+1}) \leq \Theta F(x_k)$  - přepínání metod

BFGS metoda

- $B_{k+1} = B_k + (1 + \frac{g_k^T B_k g_k}{g_k^T s})(\frac{ss^T}{s^T g_k}) - (\frac{sg_k^T B_k + B_k g_k s^T}{g_k^T s})$  (BFGS)
- $s = -B_{k+1} g_k$
- $x_{k+1} = x_k + \alpha * s^T$ , délka kroku  $\alpha > 0$  odhadnuta metodou bisekce

**else**, v opačném případě, pokud nerovnice (27) není vyhovující:

Gauss-Newtonova metoda

- $B_{k+1} = J_k^T J_k$  (Gauss-Newton)
- $-g_k = B_k s_k$
- $x_{k+1} = x_k + s^T$

**end**

2. Přibližné řešení  $x^* = x_k$ .

**Poznámka 6** *Kritérium pro výběr metody (nerovnice (27)) probíhá pro každou novou iteraci znovu.*

Tímto jsme popsali jednotlivé optimalizační metody. Vybrané metody byly implementovány a v následující kapitole budou podrobeny numerickým experimentům na testovacích úlohách.

## 6 Numerické experimenty

V této části podrobíme vybrané gradientní metody a metody s proměnnou metrikou numerickým experimentům. Vybrané metody byly implementovány v prostředí matematického softwaru Matlab a budou testovány na jednotlivých úlohách. Výsledky budou okomentovány a dojde k hodnocení a porovnání daných metod.

V první části bude testována jednodušší funkce dvou proměnných na metodě DFP (viz podkapitola 4.2), metodě největšího spádu (STD, viz podkapitola 3.1), dále na metodě sdružených gradientů (CGM, podkapitola 3.2) a také na její Fletcher-Reevesově modifikaci (FR, podkapitola 3.2). Výsledky tohoto testování budou porovnány s interní funkcí matematického softwaru Matlab, a to konkrétně s funkcí *fminsearch*, jejíž charakteristické vlastnosti a popis syntaxe jsou popsány zde [10].

V druhé části experimentování se zaměříme na problematiku nejmenších čtverců, kde je předmětem minimalizace reziduí, tedy rozdíl naměřených hodnot od očekávaných. Předmětem testování bude populační funkce, kde je v čase  $t$  měřena živočišná populace. V této části také budou testovány implementované metody, a to konkrétně Gauss-Newtonova metoda, dále kombinace Gauss-Newtonovy metody a metody BFGS (přepínání metod, viz podkapitola (5.3) a výsledky tohoto měření budou porovnávány s interní funkcí z prostředí Matlab, a to funkcí *lsqnonlin*. [12]

Ve třetí části bude na vybraných metodách jako v části druhé testována funkce týkající se oceňování opcí za využití Black-Scholesových rovnic.

### 6.1 Testování 1

Jak již bylo v úvodu této kapitoly řečeno, předmětem prvního testování bude jednodušší kvadratická funkce dvou proměnných, konkrétně funkce

$$f(x, y) = 1.5x^2 + 8x + 100y^2 - 2y,$$

jejíž prostorový graf můžeme vidět na obrázku 1.

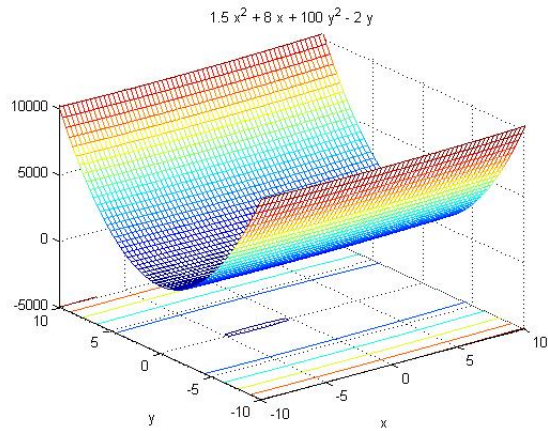
V prostředí Matlabu proběhlo testování se vstupními parametry, a to konkrétně funkcí  $f(x, y)$ , dále počátečním bodem  $x_0$  a zastavovací podmínkou  $\epsilon$ .

Konkrétně byly vstupní hodnoty nastaveny takto:

$$x_0 = [1, 1], \epsilon = 0, 01.$$

Mezi kritéria měření patřily hodnoty x-ové a y-ové souřadnice nalezeného minima, dále funkční hodnota, počet iterací a čas výpočtu, měřený pomocí funkce *tictoc*.

Z výsledků prvního testování v tabulce 1 můžeme vidět, že nejdále od řešení vypočteného funkcí *fminsearch* byly výsledky vyhodnocené metodou největšího spádu, která, ačkoli potřebovala nejvyšší počet iterací, a to 10, tak paradoxně trvala nejkratší dobu. Výsledky metody sdružených gradientů a Fletcher-Reevesova modifikace této metody vyhodnotily stejné výsledky i na několik desetinných míst. V čem se tyto metody lišily, byl ovšem čas, kdy byla Fletcher-Reevesova modifikace o skoro celou sekundu pomalejší, což je v tomto měřítku velká hodnota.

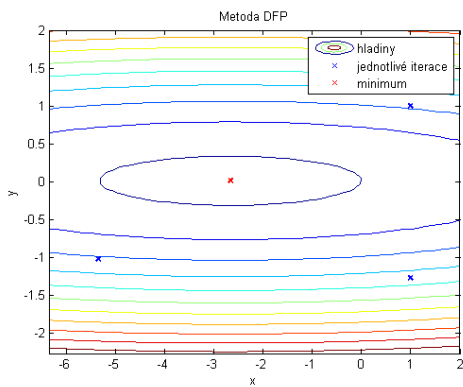


Obrázek 1: Funkce 1

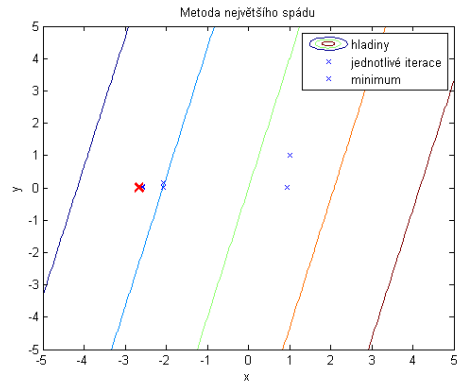
Funkce 1	x-minimum	y-minimum	Funkční hodnota	Počet iterací	Čas výpočtu
fminsearch	-2.666672521393033	0.010003594952835	-10.676666665322880	-	0,038778
DFP	-2.666666666666665	0.010000000000000	-10.676767676767676	3	1,265044
STD	-2.663953312329464	0.009997738871390	-10.676655622717758	10	0,125479
CGM	-2.666666666666667	0.010000000000000	-10.676666666666666	2	0,224142
FR	-2.666666666666667	0.010000000000000	-10.676666666666666	2	1,218367

Tabulka 1: Testování 1

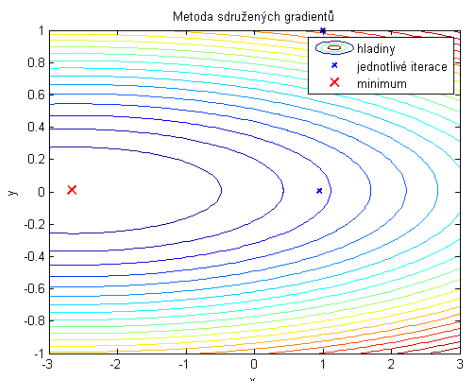
Na následujících obrázcích můžeme vidět cestu k dosažení naměřené hodnoty z tabulky 1. V grafech jsou znázorněny jednotlivé iterace od počátečního bodu, až k dosažení minimální hodnoty. Dále můžeme na obrázcích vidět také jednotlivé hladiny funkce  $f(x, y)$ .



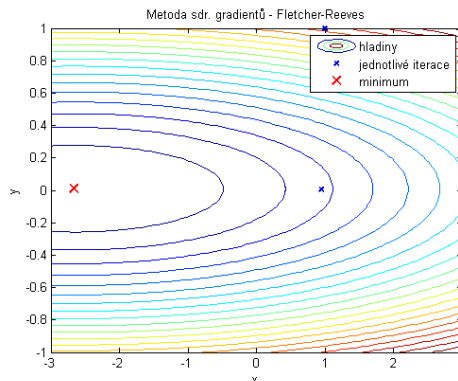
Obrázek 2: Metoda DFP



Obrázek 3: Metoda nejv. spádu



Obrázek 4: Metoda sdr. gradientů



Obrázek 5: Fletcher-Reeves

Na těchto čtyřech obrázcích jsme mohli vidět znázornění metod a jednotlivých iterací, kterak se od počátečního bodu  $x_0$  dostali na naměřenou hodnotu, která byla zaznamenána do tabulky 1.

## 6.2 Testování 2

V druhé části našich numerických experimentů se zaměříme na minimalizaci součtu čtverců funkcí, jak bylo zmíněno v podkapitole 5.1. Inspirací pro testování nám byla populační úloha [13].

Pro příklad máme k dispozici tato data  $\{(t_i, y_i)\}_{i=1}^m$ , obsahující velikost živočišné populace v různém čase. Předpokládejme data:

$$\begin{array}{l} t_i: \quad 1 \quad 2 \quad 4 \quad 5 \quad 8 \\ y_i: \quad 3 \quad 4 \quad 6 \quad 11 \quad 20 \end{array}$$

čas  $t$  je zde měřen v letech a velikost populace  $y$  ve stovkách. Je obvyklé použít na tuto populační problematiku exponenciální model, jako vhodné proložení bodů v grafu. Proto i zde byl použit exponenciální model, konkrétně

$$y_i \approx x_1 e^{x_2 t_i},$$

pro odpovídající výběr parametrů  $x_1$  a  $x_2$ .

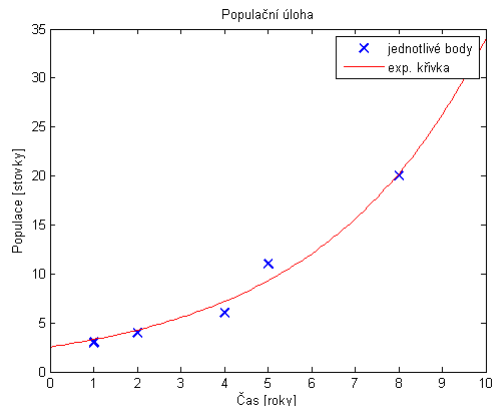
V modelu nejmenších čtverců budeme minimalizovat reziduální součet, čili odchylku naměřených hodnot od očekávaných. Podle kapitoly 5.1 bude minimalizační funkce vypadat takto:

$$\min_{x_1, x_2} f(x_1, x_2) = \sum_{k=1}^5 (x_1 e^{x_2 t_i} - y_i)^2 \quad (28)$$

a jednotlivá  $i$ -tá funkce takto

$$f(x_1, x_2) = x_1 e^{x_2 t_i} - y_i. \quad (29)$$

Na následujícím obrázku je ilustrován exponenciální model populační funkce, kde jednotlivé body pocházejí z dat, která jsou obsažena v tabulce v této kapitole.



Obrázek 6: Populační úloha

Minimalizace této funkce byla provedena interní matlabovskou funkcí *lsqnonlin*, jejíž syntaxe a vlastnosti jsou popsány zde [12].

Výsledky vypočtené touto funkcí budou porovnány s výsledky, které nám vynesou Gauss-Newtonova metoda a také s výsledky kombinované metody, kde se na základě porovnávání funkčních hodnot přepíná Gauss-Newtonova metoda s metodou BFGS, popsané v sekci 5.3.

Metodou *lsqnonlin* jsme řešili minimalizaci funkce z předpisu (28), která je součtovou funkcí jednotlivých funkcí pocházejících ze vzorce (29), a obsahující data z tabulky z této kapitoly. Počátečním bodem  $x_0$  byl zvolen bod o souřadnicích  $[1, 1]$ .

Výsledné minimum má tyto souřadnice:

$$x_{min} = (2.541043505623090, 0.259504925945667).$$

Funkční hodnota v minimu  $x_{min}$  vyšla

$$f(x_{min}) = 4.494261250455289.$$

K tomuto minimu došla funkce v celkově 19. iteraci. Funkční hodnoty jednotlivých iterací jsou pro lepší přehled uloženy v textovém souboru *"lsqnonlinvysledky.txt"*, který je součástí CD.

Výsledky vypočtené funkcí *lsqnonlin* porovnáme s Gauss-Newtonovou metodou, dále s metodou BFGS a také s kombinací těchto metod, speciální metodou (s kumulativní korekcí), která přepíná pro výpočty jednotlivých iterací právě mezi Gauss-Newtonovou metodou a metodou BFGS, na základě porovnání posledních funkčních hodnot, viz sekce 5.3.

Výsledky měření při vstupních parametrech

$$x_0 = [1, 1], \epsilon = 1e - 06$$

nejlépe shrnuje následující tabulka:



Testování 2	x_min	y_min	Počet iterací	Čas výpočtu
lsqnonlin	2.541043505623090	0.259504925945667	19	0.222944
Gauss-Newton	2.541045661755548	0.259504802435437	12	0.010899
GN&BFGS	2.541045661736715	0.259504802436516	13	1.795268

Tabulka 2: Testování 2, populační funkce

Z výsledků v tabulce 2 vidíme, že námi naměřené hodnoty pomocí Gauss-Newtonovy metody a dále metody založené na přepínání metod, vyšly takřka totožné, lišily se až na dvanáctém desetinném místě, což bylo jistě způsobeno hodně malou zastavovací hodnotou  $\epsilon$ . V čem se metody lišily, byl čas výpočtu, kde metoda kombinovaná trvala skoro o dva desetinné řády déle.

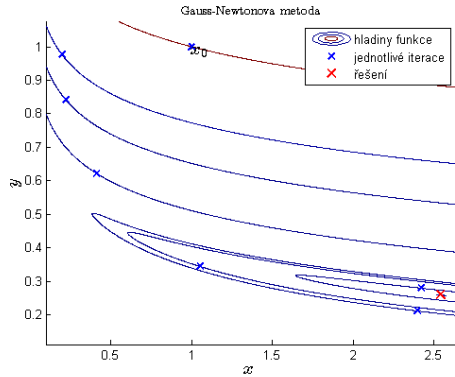
Nyní bude zajímavé si ukázat fungování naší hlavní metody (metoda s přepínáním, dále jen GN&BFGS) během jednotlivých iterací, kde bude znázorněno, ve které iteraci byla jaká metoda používána, viz následující tabulka. Volitelná hodnota  $\theta$  byla nastavena na 0,1.

Iterace	Používaná metoda	x-min	y-min	Funkční hodnota
1	Gauss-N	0.197370674798469	0.976166782723002	2.17614101e+05
2	Gauss-N	0.220839916829515	0.841441258389290	27306.315735
3	Gauss-N	0.411622453031570	0.621943297668367	1585.146835
4	Gauss-N	1.051325658235525	0.344581268979388	47.268807
5	Gauss-N	2.396176769511444	0.211887118022670	65.260799
6	Gauss-N	2.421998504437998	0.279771452427857	10.858248
7	BFGS	2.421999888523000	0.279775091133585	10.861802
8	Gauss-N	2.541579385636456	0.260001839294120	4.502416
9	Gauss-N	2.541381456756746	0.259486603279795	4.494262
10	Gauss-N	2.541033065738562	0.259505522387114	4.494261
11	Gauss-N	2.541046180968061	0.259504772704801	4.494261
12	Gauss-N	2.541045661736715	0.259504802436516	4.494261

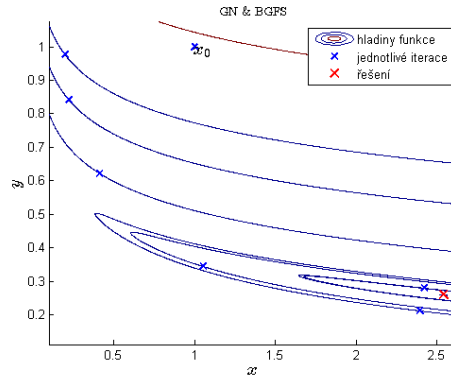
Tabulka 3: Testování 2,  $x_0 = [1, 1], \theta = 0.1$

V tabulce 3 bylo znázorněno testování naší kombinované metody. Je nutno podotknout, že volba vstupních parametrů je vysoce citlivá. Pokud zvolíme hodnotu  $\theta$  ještě nižší (0.0001), ukončí se iterační proces dříve, na 8. iteraci metodou BFGS. Naopak při větší hodnotě  $\theta$  dochází k výpočtu jen přes Gauss-Newtonovu metodu. Dá se říci, že čím větší přesnost vyžadujeme, čili zmenšujeme hodnotu  $\epsilon$ , tím více se v iteracích objevuje Gauss-Newtonova metoda. Všechna tato tvrzení mohou být předmětem dalšího zkoumání.

Na dalších dvou obrázcích si výpočty ještě graficky znázorníme, bude možno vidět hladiny součtové funkce a k tomu jednotlivé iterace a zobrazený finální bod.



Obrázek 7: Gauss-Newtonova metoda



Obrázek 8: Komb. metoda GN& BFGS

Tímto bychom zakončili testování populační úlohy, kde nám jednotlivé metody moc rozdílů nepřinesly, snad jen časové, kombinovaná metoda přepínání trvala jednoznačně nejdéle dobu, což bylo překvapivé zjištění, jelikož jsme čekali, že přepínání metody má hledání urychlit. Časový rozdíl jednotlivých metod však může být způsoben neefektivní implementací. Podle našeho experimentu nemá výrazný vliv na hledání minima ani hodnota  $\theta$ , kde hodnoty prakticky v celém jejím definičním oboru upřednostňují hledání pomocí Gauss-Newtonovy metody.

### 6.3 Příklad z praxe

Nyní si ukážeme prakticky příklad použití metody na nelineární nejmenší čtverce. Konkrétně se bude jednat o oceňování opcí na největším německém burzovním indexu DAX, konkrétně z dat z 15. května 2015, která jsou součástí příloženého CD.

Evropská opce je smlouva mezi prodávajícím a kupujícím, která dává kupujícímu právo prodat nebo koupit od prodávajícího konkrétní aktivum za konkrétní cenu  $K$  (strike) kdykoliv až do data vypršení kontraktu  $T$  (maturity). [23]

Cena evropské opce je daná Black-Scholesovou formulí

$$C_{BS}(\sigma) = \Phi(d_1)S_{t_0} - \Phi(d_2)Ke^{-r\tau},$$

$$d_1 = \frac{1}{\sigma\sqrt{\tau}} \left[ \ln\left(\frac{S_{t_0}}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)\tau \right],$$

$$d_2 = d_1 - \sigma\sqrt{\tau},$$

kde  $\Phi$  značí kumulativní distribuční funkci normálního rozdělení

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2},$$

$r$  značí rizikově neutrální úrokovou míru,  
 $\tau$  je čas do expirace (maturity time)  
a  $\sigma$  značí volatilitu.

Volatilita označuje míru kolísání hodnoty aktiva nebo jeho výnosové míry (obvykle jako směrodatnou odchylku těchto změn během určitého časového úseku). [25]

Roční volatilita  $\sigma$  je směrodatná odchylka logaritmu výnosů aktiva v průběhu 1 roku. Obecná volatilita  $\sigma_T$  pro časový úsek  $T$  v letech je vyjádřena jako:

$$\sigma_T = \sigma\sqrt{T}.$$

Implikovaná volatilita je taková nezáporná hodnota  $\sigma$  (určena jednoznačně), pro kterou je tržní cena opce  $C$  rovna modelové ceně s touto hodnotou parametru  $\sigma$ , tj.

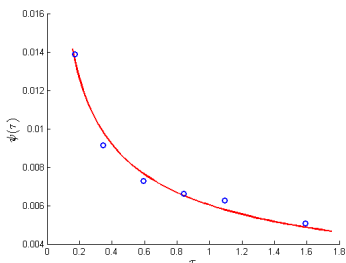
$$C = C_{BS}(\sigma). \quad (30)$$

V první části se podíváme na testování podobné úlohy jako v případě populační funkce v podkapitole 6.2. Opět se budeme snažit co nejlépe proložit křivku jednotlivými body, značící tržní data. Jedná se o šikmost (skew) tzv. volatility povrchu (surface) pro hodnoty  $K$  rovny (ATM - at the money) aktuální ceně podkladového aktiva (spot).

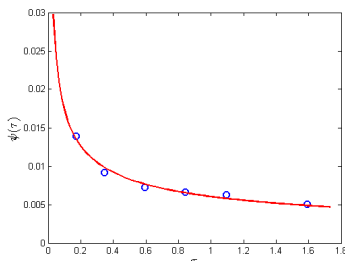
Podle [24] může být ATM šikmost aproximována funkcí

$$\psi(\tau) = A\tau^{-\alpha},$$

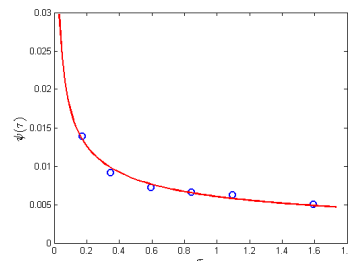
kde  $\tau$  je zbývající čas do doby expirace  $A$  a  $\alpha$  jsou neznámé parametry. Naším úkolem bude z dostupných dat odhadnout hodnoty parametrů  $A$  a  $\alpha$ . Opět budeme porovnávat Gauss-Newtonovu metodu, metodu založenou na přepínání mezi metodami BFGS a Gauss-Newtonovou a interní funkci *lsqnonlin*. Vstupní parametry jsou stejné jako u předchozího testování.



Obrázek 9: Řešení metodou *lsqnonlin*



Obrázek 10: Řešení G-N metodou



Obrázek 11: Řešení pomocí GN&BFGS

Na obrázcích můžeme pro ilustraci vidět jednotlivé ATM volatility v čase  $T$ . Z tabulky 4 vyčteme naměřené hodnoty parametrů  $\alpha$ , dále hodnoty reziduí a počet iterací.

Metoda	alpha	rezidua	počet iterací
<i>lsqnonlin</i>	0.459265	1.000476e-06	4
Gauss-Newton	0,458010	0.0000005	8
GNBFGS	0,458010	0.0000005	8

Tabulka 4: Testování ATM volatility

Z výsledků v tabulce 4 vidíme, že hodnoty parametrů naměřené námi implementovanými funkcemi byly prakticky totožné, obě potřebovaly osm iterací k dosažení této hodnoty.

V druhé části testování praktické úlohy se podíváme na SVI model [24] volatily povrchu, charakterizovaného funkcí

$$f(a, b, \rho, m, \sigma) = a + b \left\{ \rho(k - m) + \sqrt{(k - m)^2 + \sigma^2} \right\}, \quad (31)$$

kde

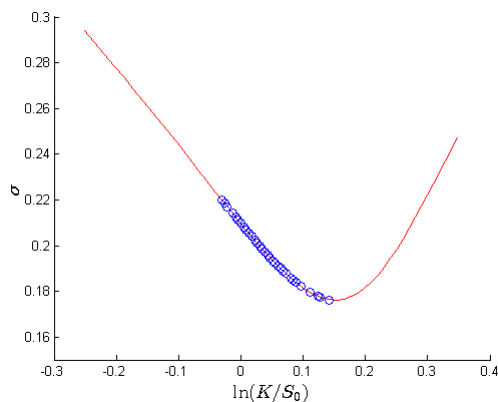
$$k = K/S_o,$$

je hodnota tzv. moneyness vyjadřující vztah mezi realizační cenou opce a cenou podkladového aktiva. [26]

Parciální derivace podle jednotlivých proměnných  $(a, b, \rho, m, \sigma)$  vypadají takto:

$$\begin{aligned} \frac{\partial f}{\partial a} &= 1, \\ \frac{\partial f}{\partial b} &= \left\{ \rho(k - m) + \sqrt{(k - m)^2 + \sigma^2} \right\}, \\ \frac{\partial f}{\partial \rho} &= b(k - m), \\ \frac{\partial f}{\partial m} &= -b\rho - 2b(k - m), \\ \frac{\partial f}{\partial \sigma} &= \frac{b\sigma}{\sqrt{(k - m)^2 + \sigma^2}}. \end{aligned}$$

Dvě námi implementované metody nešly použít, protože jsou zde omezení pro jednotlivé proměnné  $(a, b, \rho, m, \sigma)$  (viz článek [24]) a my řešíme nepodmíněnou optimalizaci. Původní představa se nepotvrdila, rezidua už v druhé iteraci vyšla záporná, což nedává smysl. Proto na obrázku 12 je vyobrazená závislost parametrů pouze pomocí funkce *lsqnonlin*.



Obrázek 12: SVI metodou *lsqnonlin*

Na obrázku 12 je zobrazena hodnota volatily  $\sigma$  v závislosti na hodnotě logaritmu moneyness.

## 7 Závěr

V této práci jsme se zabývali nepodmíněnou optimalizací. V teoretickém úvodu jsme si zařadili pojmy jako gradient nebo Hessova matice, se kterými bylo v průběhu celé práce pracováno. Dále jsme provedli úvodní rešerši gradientních metod a metod s proměnnou metrikou.

V kapitole 3 jsme se zaměřili na klasické gradientní metody a jejich vlastnosti. Po úvodu do této skupiny jsme se zaměřili na metodu největšího spádu a popsali její algoritmus. Na tuto metodu jsme navázali metodou sdružených gradientů, kde jsme po vysvětlení algoritmu uvedli dvě modifikace, a to Fletcher-Reevesovu a Polak-Ribierovu. Na závěr této kapitoly jsme se zabývali konvergencí iteračních metod.

Hlavním cílem kapitoly 4 bylo popsat metody s proměnnou metrikou. Po popsání specifik těchto metod jsme se zabývali jednotlivými metodami DFP a BFGS, kde jsme u obou popsali algoritmus a hlavní vlastnosti. Na závěr jsme zmínili příklad Broydenových metod, konkrétně jakousi lineární kombinaci dvou zmíněných metod s proměnnou metrikou.

V následující kapitole jsme se zaměřili na úlohu nelineárních nejmenších čtverců. Nejdříve jsme popsali základní myšlenku a cílovou funkci, poté jsme uvedli nejznámější metodu na nelineárních nejmenších čtvercích, a to Gauss-Newtonovu metodu a popsali její algoritmus. V další části této kapitoly jsme se zaměřili na kombinaci Gauss-Newtonovy metody a metody BFGS, určené pro úlohu nelineárních nejmenších čtverců. Vysvětlili jsme princip přepínání metod a popsali jejich algoritmy.

V kapitole 6 jsme podrobili vybrané metody numerickým experimentům. V prvním testování jsme na jednoduché kvadratické funkci porovnali metodu DFP, největšího spádu, sdružených gradientů a její Fletcher-Reevesovu modifikaci. Porovnávali jsme dosažené výsledky s interní matlabovskou metodou *fminsearch*. Souřadnice dosaženého minima i funkční hodnoty se lišily jen velmi drobně, přesto nejvzdálenější hodnoty od hodnot metody *fminsearch* měla metoda největšího spádu. Časově na tom byly nejlhův metody DFP a Fletcher-Reevesova modifikace. Pro přesnější hodnocení by bylo zapotřebí více testovacích funkcí, protože některé metody mohly být naimplementovány méně efektivněji.

Předmětem druhého testování byla úloha na nelineární nejmenší čtverce. Konkrétně se jednalo o praktickou úlohu týkající se živočišné populace. Zde jsme porovnávali výsledky Gauss-Newtonovy metody, dále metody založené na přepínání Gauss-Newtonovy a BFGS metody s interní funkcí *lsqnonlin*. Výsledky obou našich implementovaných metod byly takřka stejné, lišily se až na dvanáctém desetinném místě. V čem se ale výrazně lišily, byl čas, kde metoda s přepínáním trvala mnohem delší čas, což bylo oproti prvotnímu očekávání, že princip přepínání hledání minima urychlí. Ve výsledku našeho experimentu nám přepnutí na BFGS metodu spíše uškodilo, což jsme mohli vidět v tabulce 2. Jak již bylo řečeno, pro důvěryhodnější tvrzení by bylo potřeba více testů.

V třetí části našich experimentů jsme se zaměřili na praktickou úlohu z finančního prostředí týkající se oceňování opcí. Po krátkém úvodu do problematiky jsme se snažili odhadnout z dostupných dat neznámé parametry aproximující funkce. Opět jsme srovnali výsledky vy-

hodnocené metodou *lsqnonlin*, Gauss-Newtonovy a přepínací metody. Výsledky posledních dvou zmíněných se takřka nelišily, jen metoda *lsqnonlin* provedla o polovinu iterací méně. V poslední části jsme se podívali na SVI model volatility povrchu a snažili jsme se charakterizovat závislost mezi volatilitou a hodnotou moneyness. Rozumné výsledky jsme však dostali pouze pomocí funkce *lsqnonlin*. Námi implementované metody byly vytvořeny pro nepodmíněnou optimalizaci a zde bylo mnoho omezujících podmínek.

Určitě by se dala naznačit úprava metod pro úlohu hledání všech lokálních extrémů na uvedené oblasti, ale vzhledem k velkému rozsahu již implementovaných metod může být toto téma příkladem navázání na práci.

Během práce nás napadlo mnoho dalších námětů k případnému rozšíření práce. Jedním z nich je například další testování jednotlivých metod, které by odhalilo více důkazů pro jejich hodnocení. Dalším nápadem je větší zaměření na jiné typy Broydenových metod. Všechny tyto náměty by mohly sloužit k případnému dalšímu zkoumání.

## Reference

- [1] TICHÝ, Petr. *Numerické metody optimalizace* [online]. Plzeň: Západočeská univerzita v Plzni. Dostupné z WWW: [http://home.zcu.cz/~kopincov/MNO\\_Petr\\_Tichy\\_slidy\\_ZCU.pdf](http://home.zcu.cz/~kopincov/MNO_Petr_Tichy_slidy_ZCU.pdf)
- [2] MAŘÍK, Robert. *Parciální derivace* [online]. Mendelova univerzita v Brně, 2006. Dostupné z WWW: <https://www.math.muni.cz/~pribylova/parc-der-cz.pdf>
- [3] DOSTÁL, Z., BEREMLIJSKI, P. *Metody optimalizace* [online]. VŠB-TU Ostrava a ZČU v Plzni. Dostupné z WWW: [http://mi21.vsb.cz/sites/mi21.vsb.cz/files/unit/metody\\_optimalizace.pdf](http://mi21.vsb.cz/sites/mi21.vsb.cz/files/unit/metody_optimalizace.pdf)
- [4] *Gradient(matematika)*. Wikipedia: the free encyclopedia. [online]. 2001- [cit. 2016-04-7]. Dostupné z: [https://cs.wikipedia.org/wiki/Gradient\\_\(matematika\)](https://cs.wikipedia.org/wiki/Gradient_(matematika))
- [5] *Základní pojmy a tvrzení nepodmíněné matematické optimalizace*. Katedra mechaniky, ZČU v Plzni. Dostupné z: <http://www.kme.zcu.cz/download/predmety/405-dso2.pdf>
- [6] VAŘEKOVÁ SVOBODOVÁ, Radka. *Optimalizace bez omezení* [online prezentace]. Brno : Národní centrum pro výzkum biomolekul, MU, [cit. 2016-04-24]. Dostupný z WWW: <http://www.ncbr.muni.cz/~svobodova/vyuka/optimalizace/ppt/uncon2.ppt>.
- [7] LUKŠAN, L. *Nepodmíněná minimalizace* [online]. Academy of Sciences of the Czech Republic Institute of Computer Science. Dostupné z WWW: <http://www.cs.cas.cz/luksan/lekce5.pdf>
- [8] LUKŠAN, L., SPEDICATO, E. *Variable metric methods for unconstrained optimization and nonlinear least squares*. [online]. Journal of Computational and Applied Mathematics, 2000, p. 61–95. Dostupný z WWW <http://www.sciencedirect.com/science/article/pii/S0377042700004209> . ISSN 0377-0427.
- [9] LUKŠAN, L. *Numerické optimalizační metody*. [online]. Institute of Computer Science, Academy of Sciences of the Czech Republic, 2015. Dostupný z WWW <http://www.cs.cas.cz/luksan/lekce4.pdf>
- [10] *fminsearch*. The MathWorks, Inc. [online]. © 1994-2016 [cit. 2016-05-02]. Dostupné z: <http://www.mathworks.com/help/matlab/ref/fminsearch.html>
- [11] NOCEDAL, J., WRIGHT, S.J. *Numerical Optimization*. [online]. Springer, 2006. Dostupný z WWW [http://home.agh.edu.pl/~pba/pdfdoc/Numerical\\_Optimization.pdf](http://home.agh.edu.pl/~pba/pdfdoc/Numerical_Optimization.pdf)
- [12] *lsqnonlin*. The MathWorks, Inc. [online]. © 1994-2016 [cit. 2016-05-09]. Dostupné z: <http://http://www.mathworks.com/help/optim/ug/lsgnonlin.html>
- [13] *Nonlinear Least Squares Data Fitting*. Fakulta elektrotechnická, ČVUT. [online]. Datum nevedeno [cit. 2016-05-13]. Dostupné z: [https://cw.fel.cvut.cz/wiki/\\_media/courses/a4b33opt/cviceni/nls.pdf](https://cw.fel.cvut.cz/wiki/_media/courses/a4b33opt/cviceni/nls.pdf)

- [14] FLETCHER,R.,REEVES, C.M.: *Function minimization by conjugate gradients*. Computer J. 7 (1964) 149- 154.
- [15] POLAK,E., RIBIÉRE,G.: *Note sur la convergence de directions conjuguées*. Rev. Française Informat. Recherche Opertionelle, 3e Annee 16 (1969) 35-43.
- [16] DAVIDON,W.C.: *Variable metric method for minimization*. SIAM J. Optimization, 1 (1991), 1-17
- [17] BROYDEN,C.G.: *The convergence of a class of double rank minimization algorithms, Part 1 – general considerations, Part 2 – the new algorithm*, J. Inst. Math. Appl. 6 (1970) 76 –90, 222–231.
- [18] FLETCHER,R.: *A new approach to variable metric algorithms*, Comput. J. 13 (1970) 317–322.
- [19] GOLDFARB,D.: *A family of variable metric algorithms derived by variational means*, Math. Comput. 24 (1970) 23–26.
- [20] SHANNO, D.F.: *Conditioning of quasi-Newton methods for function minimization*, Math. Comp. 24 (1970) 647–656
- [21] *Metoda nejmenšich čtverců*. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2016-05-14]. Dostupné z: [https://cs.wikipedia.org/wiki/Metoda\\_nejmenšich\\_čtverců](https://cs.wikipedia.org/wiki/Metoda_nejmenšich_čtverců)
- [22] DENNIS,J.E.,GAY D.,WELSCH, R.E.: *An adaptive nonlinear least squares algorithm*, ACM Trans. Math. Software 7 (1981) 348–368.
- [23] *Opce(finance)*. Wikipedia: the free encyclopedia. [online]. 2001- [cit. 2016-05-17]. Dostupné z: [https://cs.wikipedia.org/wiki/Opce\\_\(finance\)](https://cs.wikipedia.org/wiki/Opce_(finance))
- [24] GATHERAL, J., JACQUIER, A.:*Arbitrage-free SVI volatility surfaces*. Quantitative Finance, 2014.,14,1,s. 59-71.
- [25] *Volatilita*. Wikipedia: the free encyclopedia. [online]. 2001- [cit. 2016-05-18]. Dostupné z: <https://cs.wikipedia.org/wiki/Volatilita>
- [26] *ATM forward strike*. SAXO Bank, Slovník finančních pojmů. [online].[cit. 2016-05-19]. Dostupné z: <http://cz.saxobank.com/support/slovník-pojmu/atm-forward-strike>



## A Příloha

Jako přílohu přikládám CD-ROM, na kterém se nachází:

- plný text práce (.pdf)
- kódy jednotlivých metod (.m)
- soubor s výsledky měření (.txt)
- data pro praktickou část (.mat)