dr.sc. Rudolf Scitovski, redoviti profesor
Odjel za matematiku, Sveučilište u Osijeku

# Metode optimizacije

Pripreme za predavanja iz kolegija *Metode optimizacije*
poslijediplomski studij matematike
Odsjek za matematiku, Univerzitet u Tuzli
Svibnja, 2005.

## Sadržaj

**1. Konveksne funkcije** **1**
1.1. Konveksni skupovi . . . 1
1.2. Konveksne funkcije . . . 1
1.3. Lokalni minimum . . . 3
1.4. Primjeri i motivacija . . . 5

**2. Metode jednodimenzionalne minimizacije** **7**
2.1. Metoda zlatnog reza . . . 7
2.2. Metoda parabole . . . 11
2.3. Newtonova metoda tangenti . . . 13

**3. Gradijentne metode** **14**
3.1. Gradijentna metoda i metoda najbržeg spusta . . . 15

**4. Metode Newtonovog tipa** **20**
4.1. Newtonova metoda . . . 20
4.1.1. Svojstva Newtonove metode . . . 21
4.2. Quasi–Newtonove metode . . . 25
4.2.1. Sustavi nelinearnih jednadžbi . . . 25
4.2.2. Minimizacija glatke funkcije . . . 30
4.2.3. Korekcije koje čuvaju pozitivnu definitnost . . . 33
4.3. Inverzne korekcije . . . 36

**5. Seminarski radovi** **38**

# 1. Konveksne funkcije

## 1.1. Konveksni skupovi

**Definicija 1.1.** *Kažemo da je skup $K \subseteq \mathbb{R}^n$* **konveksan** *ako za bilo koje dvije točke $x_1, x_2 \in K$ sadrži i segment određen tim točkama, tj.*

$$x_1, x_2 \in K \Rightarrow \lambda x_1 + (1-\lambda)x_2 \in K \quad \forall \lambda \in [0,1].$$

**Zadatak 1.1.** *Pokažite da konveksan skup sadrži svaku konveksnu kombinaciju od konačno svojih točaka, tj.*

$$(x_1, \dots, x_n \in K) \quad \wedge \quad \left(\lambda_1, \dots, \lambda_n \geq 0, \quad \sum_{i=1}^n \lambda_i = 1\right) \Longrightarrow \sum_{i=1}^n \lambda_i x_i \in K$$

**Definicija 1.2.** *Kažemo da je skup $K \subseteq \mathbb{R}^n$* **strogo konveksan** *ako*

$$x_1, x_2 \in K, \quad x_1 \neq x_2 \Rightarrow \lambda x_1 + (1-\lambda)x_2 \in \operatorname{Int} K \quad \forall \lambda \in (0,1).$$

**Definicija 1.3.** *Kažemo da je skup $K \subseteq \mathbb{R}^n$* **jako konveksan** *ako postoji takva konstanta $\gamma > 0$, tako da*

$$(x_1, x_2 \in K) \wedge \left(\|y\| \leq \gamma \|x_2 - x_1\|^2\right) \Longrightarrow \frac{x_1 + x_2}{2} + y \in K.$$

**Zadatak 1.2.** *Pokažite da je svaki jako konveksan skup ujedno i strogo konveksan, ali obrat ne vrijedi.*

## 1.2. Konveksne funkcije

**Definicija 1.4.** *Kažemo da je funkcija $f : \mathbb{R}^n \to \mathbb{R}$* **konveksna** *ako vrijedi*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) \tag{1.1}$$

$\forall x, y \in \mathbb{R}^n$ *i* $\forall \lambda \in [0,1]$.

**Definicija 1.5.** *Kažemo da je funkcija $f : \mathbb{R}^n \to \mathbb{R}$* **strogo konveksna** *ako $\forall x, y \in \mathbb{R}^n, \quad x \neq y$ vrijedi*

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y) \tag{1.2}$$

$\forall \lambda \in (0,1)$.

**Definicija 1.6.** *Kažemo da je funkcija $f : \mathbb{R}^n \to \mathbb{R}$* **jako konveksna** *ako postoji takav realni broj $\gamma > o$ da vrijedi*

$$f\left(\frac{x+y}{2}\right) \leq \frac{1}{2}(f(x) + f(y)) - \gamma \|x - y\|^2 \tag{1.3}$$

$\forall x, y \in \mathbb{R}^n$.

**Primjedba 1.1.** *Može se pokazati da da je funkcija $f : \mathbb{R}^n \to \mathbb{R}$ jako konveksna ako postoji realan broj $\kappa > 0$ tako da bude*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \kappa \lambda (1-\lambda) \|x - y\|^2 \tag{1.4}$$

$\forall x, y \in \mathbb{R}^n$ *i* $\forall \lambda \in [0,1]$.

Naime, ako u (1.4) stavimo $\lambda = \frac{1}{2}$ i $\gamma = \frac{1}{4}\kappa$, dobivamo (1.4).

**Primjer 1.1.** *Neka je $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^n$ simetrični linearni operator. Tada je kvadratna forma $f(x) := \frac{1}{2}(\mathcal{A}x, x)$ konveksna funkcijaonda i samo onda ako je $\mathcal{A} \geq 0$, tj. ako su sve svijstvene vrijednosti nenegativne – označimo ih s $\lambda_1 \geq \cdots \lambda_n \geq 0$. Kako je (vidi primjerice [9], [10])*

$$\lambda_n \|x\|^2 \leq (\mathcal{A}x, x) \leq \lambda_1 \|x\|^2 \text{ za sve } x \in \mathbb{R}^n,$$

*može se pokazati da vrijedi*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) - \frac{1}{2}\lambda_n \lambda(1-\lambda)\|x-y\|^2$$

*Dakle, ako je $\mathcal{A}$ pozitivno definitan ($\lambda_n > 0$), $f$ je jako konveksna funkcija*

**Primjer 1.2.** *Funkcija $f : \mathbb{R}^2 \to \mathbb{R}^2$, $\quad f(x_1, x_2) = x_1^2 + x_2^2$ je jako konveksna funkcija, a funkcija $g : \mathbb{R}^2 \to \mathbb{R}^2$, $\quad g(x_1, x_2) = x_1^2$ je konveksna, ali nije jako konveksna ($\lambda_1 = 1, \lambda_2 = 0$). Kako izgledaju plohe ovih funkcija?*

**Zadatak 1.3.** *Pokažite da je svaka jako konveksna funkcija ujedno i konveksna, ali da obrat ne vrijedi.*

Uvedimo sljedeće oznake. Za neprekidno diferencijabilnu funkciju $f : \mathcal{D} \to \mathbb{R}^n$, $\mathcal{D} \subset \mathbb{R}^n$ ($f \in C^1(\mathcal{D})$) s

$$f'(x) = \left[\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right]$$

označit ćemo derivaciju funkcije $f$ u točki $x = (x_1, \ldots, x_n)^T \in \mathcal{D}$, a s

$$\nabla f(x) := (f'(x))^T$$

gradijent funkcije $f$ u točki $x$.

Ako je $f$ dvostruko neprekidno diferencijabilna ($f \in C^2(\mathcal{D})$) za $x \in \mathcal{D}$ s

$$\nabla^2 f(x) = (f''(x) =) := \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\ldots,n}$$

označit ćemo Hesseovu matricu (hesijan) funkcije $f$ u točki $x$. Primijetite da je $\nabla^2 f$ simetrična matrica.

**Lema 1.1.** *Ako je $f \in C^1(K)$ konveksna funkcija na konveksnomn skupu $K$, onda vrijedi*

$$(f'(y), x-y) \leq f(x) - f(y) \leq (f'(x), x-y) \qquad \forall x, y \in K. \tag{1.5}$$

<u>Dokaz</u>. Oduzimajući $f(y)$ na obje strane nejednakosti (1.1) dobivamo

$$f(y + \lambda(x-y)) - f(y) \leq \lambda(f(x) - f(y)).$$

Primijenjujući na lijevoj strani Lagrangeov teorem o srednjoj vrijednosti

$$f(y + \lambda(x-y)) - f(y) = (f'(y + \vartheta\lambda(x-y)), \lambda(x-y)), \qquad \vartheta \in (0,1),$$

nakon dijeljenja nejednakosti s $\lambda$ i stavljanja $\lambda \to +0$, dobivamo lijevu stranu tražene nejednakosti. Pokušajte na sličan način dobiti i desnu stranu nejednakosti. ♣

**Lema 1.2.** *Neka je $K \subseteq \mathbb{R}^n$ konveksan skup. Funkcija $f \in C^1(K)$ je jako konveksna onda i samo onda ako postoji $m > 0$ takav da je*

$$(f'(y) - f'(x), y - x) \geq m\|y - x\|^2 \quad \text{za sve } x, y \in K. \tag{1.6}$$

**Teorem 1.1.** *Neka je $K \subseteq \mathbb{R}^n$ konveksan skup takav da je $\operatorname{Int} K \neq \emptyset$. Funkcija $f \in C^2(K)$ je jako konveksna onda i samo onda ako postoji $m > 0$ takav da je*

$$(f''(x)y, y) \geq m\|y\|^2 \quad \text{za sve } x \in K \text{ i za sve } y \in R^n. \tag{1.7}$$

## 1.3. Lokalni minimum

**Definicija 1.7.** *Neka je $f : \mathcal{D} \to \mathbb{R}$, $\mathcal{D} \subset \mathbb{R}^n$ proizvoljna funkcija. Kažemo da je $x^* \in \mathcal{D}$* točka lokalnog minimuma *funkcije $f$ na skupu $\mathcal{D}$ ako postoji okolina $\mathcal{O}$ točke $x^*$, tako da je $f(x) \geq f(x^*)$ za sve $x \in \mathcal{O} \cap \mathcal{D}$. Točku $x^*$ zovemo* točka strogog lokalnog minimuma *ako postoji okolina $\mathcal{O}$ točke $x^*$, tako da vrijedi $f(x) > f(x^*)$ za sve $x \in \mathcal{O} \cap \mathcal{D} \backslash \{x^*\}$. Točku $x^*$ zovemo* točkom globalnog minimuma *funkcije $f$ ako je $f(x) \geq f(x^*)$ za sve $x \in \mathcal{D}$.*

**Teorem 1.2.** *Ako je $f : K \to \mathbb{R}$, $K \subseteq \mathbb{R}^n$ jako konveksna neprekidna funkcija na zatvorenom konveksnom skupu $K$, onda:*

*(i) funkcija $f$ je ograničena odozdo na $K$, tj. $f^* := \inf f(x) > -\infty$,*

*(ii) postoji jedinstvena točka $x^* \in K$, takva da bude $f(x^*) = f^*$,*

*(iii) skup $M(y) = \{x \in K : f(x) \leq f(y)\}$ je ograničen za svaki $y \in K$.*

**Lema 1.3.** *Da bi konveksna funkcija $f \in C^1(K)$ dostigla svoj infimum na konveksnom skupu $K \subseteq \mathbb{R}^n$ u točki $x^* \in K$ nužno je i dovoljno da bude:*

$$(f'(x^*), x - x^*) \geq 0 \quad \text{za sve } x \in K. \tag{1.8}$$

*Ako je $x^* \in \operatorname{Int} K$, onda je ovaj uvjet ekvivalentan jednakosti $f'(x^*) = 0$.*

**Lema 1.4.** *neka je $x_0 \in \mathbb{R}^n$ proizvoljna točka i $f \in C^2(\mathbb{R}^n)$ jako konveksna funkcija. Tada je skup $Y = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ zatvoren, ograničen i jako konveksan.*

**Primjer 1.3.** *Razmotrimo kvadratnu funkciju $f : \mathbb{R}^2 \to \mathbb{R}$,*

$$f(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2.$$

Ova funkcija je pozitivno definitna kvadratna forma zadana pozitivno definitnim simetričnim linearnim operatorom $\mathcal{A}$ sa svojstvenim vrijednostima $\lambda_1 = 3, \lambda_2 = 1$, kome u bazi $(e_1, e_2)$ pripada simetrična kvadratna matrica $A = \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1 \end{bmatrix}$. Prema *Primjeru 1.1.* ova funkcija je jako konveksna na čitavom $\mathbb{R}^2$, a prema *Lemi 1.3.* i *Teoremu 1.2.* postoji jedinstvena točka minimuma $x^*$, koja se može pronaći tako da riješimo jednadžbu $f'(x) = 0$. Matrica drugih derivacija (Hessijan) u toj točki mora biti pozitivno definitna. Lako s može vidjeti da je točka minimuma ove funkcije $x^* = (0, 0)$.

* * * * *

Ako pretpostavimo da je $f : \mathcal{D} \to \mathbb{R}, \quad \mathcal{D} \subset \mathbb{R}^n$ dovoljno puta diferencijablina funkcija definirana na otvorenom skupu $\mathcal{D} \subset \mathbb{R}^n$, onda možemo koristiti jednostavne nužne i dovoljne uvjete egzistencije lokalnog minimuma, koji su dobro razrađeni u elementarnoj analizi. Točku $x^* \in \mathcal{D}$ za koju vrijedi $\nabla f(x^*) = 0$ zvat ćemo stacionarna točka funkcije $f$. Sljedeći teorem daje nužne uvjete lokalnog minimuma.

**Teorem 1.3.** *Neka je $f \in C^1(\mathcal{D})$ neprekidno diferencijabilna funkcija definirana na otvorenom skupu $\mathcal{D} \subset \mathbb{R}^n$ i neka je $x^*$ točka lokalnog minimuma od $f$. Tada vrijedi*

*(i) $x^*$ je stacionarna točka funkcije $f$, tj. $\nabla f(x^*) = 0$,*

*(ii) ako je $f \in C^2(\mathcal{D})$, njezin hesijan $\nabla^2 f(x^*)$ u točki $x^*$ je pozitivno semidefinitna matrica.*

Dokaz. (i) Pretpostavimo da je $f \in C^1(\mathcal{D})$ i da $x^*$ nije stacionarna točka funkcije $f$, tj. $\nabla f(x^*) \neq 0$ i definirjmo funkciju

$$\varphi(t) := f\left(x^* - t\nabla f(x^*)\right), \quad t \in \mathbb{R}.$$

Za maleni $|t|, t \in \mathbb{R}$ funkcija $\varphi$ je neprekidno diferencijabilna funkcija za koju vrijedi

$$\varphi'(0) = -f'(x^*)\nabla f(x^*) = -\|\nabla f(x^*)\|_2^2$$

Kako je $\nabla f(x^*) \neq 0$, mora biti $\varphi'(0) < 0$. Zato postoji $\epsilon > 0$, takav da bude $\varphi(\epsilon) < \varphi(0)$. To znač da $x^*$ nije točka lokalnog minimuma od $f$, što je u suprotnosti s polaznom pretpostavkom.

(ii) Pretpostavimo da je $f \in C^2(\mathcal{D})$ i da $\nabla^2 f(x^*)$ nije pozitivno semidefinitna matrica. To znači da postoji vektor $d \in \mathbb{R}^n, \quad d \neq 0$, takav da je $d^T\nabla^2 f(x^*)d < 0$. Prema Taylorovom teoremu, a zbog činjenice da je $\nabla f(x^*) = 0$, za maleni $t > 0$ postoji $\tau \in (0, t)$, takav da bude

$$f(x^* - td) = f(x^*) + \frac{1}{2}t^2 d^T\nabla^2 f(x^* - \tau d)d.$$

Odavde za dovoljno maleni $t > 0$ zbog neprekidnosti od $\nabla^2 f$ vrijedi $f(x^* - td) < f(x^*)$. To bi značilo da $x^*$ nije točka lokalnog minimuma od $f$, što je u suprotnosti s polaznom pretpostavkom. ♣

Sljedeći teorm daje dovoljne uvjete lokalnog minimuma.

**Teorem 1.4.** *Neka je $f \in C^2(\mathcal{D})$, $\mathcal{D} \subset \mathbb{R}^n$ otvoren skup, a $x^* \in \mathcal{D}$ stacionarna točka od $f$ u kojoj je hesijan $\nabla^2 f(x^*)$ pozitivno definitan. Tada je $x^*$ točka strogog lokalnog minimuma funkcije $f$.*

Dokaz. Kako je $x^* \in \operatorname{Int}\mathcal{D}$ i $f'(x^*) = 0$, prema Taylorovom teoremu za svaki dovoljno maleni vektor $d \in \mathbb{R}^n$ vrijedi

$$f(x^* + d) = f(x^*) + \frac{1}{2}d^T\nabla^2 f(x^* + \vartheta d)d, \qquad \vartheta \in (0, 1).$$

Kako je $\nabla^2 f(x^*)$ pozitivno definitan, postoji $\lambda > 0$, takav da za sve $d \in \mathbb{R}^n$ vrijedi

$$d^T\nabla^2 f(x^*)d \geq \lambda d^T d.$$

Zbog toga je

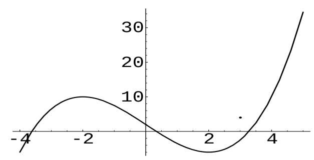
$$\begin{aligned} f(x^* + d) &= f(x^*) + \tfrac{1}{2}d^T\nabla^2 f(x^*)d - \tfrac{1}{2}d^T\left(\nabla^2 f(x^*) - \nabla^2 f(x^* + \vartheta d)\right)d, \\ &\geq f(x^*) + \tfrac{1}{2}\left(\lambda - \|\nabla^2 f(x^*) - \nabla^2 f(x^* + \vartheta d)\|\right)d^T d \end{aligned}$$

pri čemu je korištena Cauchyjeva nejednakost i svojstvo kompatibilnosti vektorskih i matričnih normi. Zbog $\lambda > 0$ i neprekidnosti od $\nabla^2 f$ slijedi $f(x^* + d) > f(x^*)$ za sve dovoljno malene vektore $d \neq 0$. Po *Definiciji* 1.7. To znači da je $x^*$ točka strogog lokalnog minimuma funkcije $f$. ♣

## 1.4. Primjeri i motivacija

Najprije ćemo navesti nekoliko primjera u kojima se pojavljuje problem minimizacije funkcije jedne ili fiše varijabli, koji će nam kasnije poslužiti kao test-primjeri kod raznih metoda minimizacije.

**Primjer 1.4.** *Treba izračunati $L_2$ udaljenost točke $T_0(3,4)$ do kubne parabole $q(x) = .5x^3 - 6x + 2$ (vidi* Sliku 1*).*




Slika 1. Udaljenost točke do krivlje

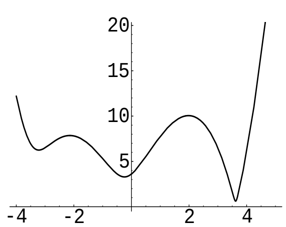
$L_2$ udaljenost točke $T_0(x_0, y_0)$ do neke točke $T(x, f(x))$ na grafu funkcije $q$ zadana je s

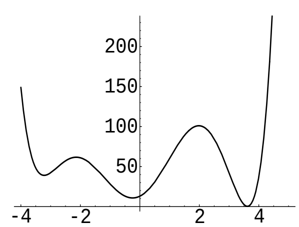
$$d_2(x) = d_2(T_0(x_0, y_0), T(x, q(x)) = \sqrt{(x - x_0)^2 + (q(x) - y_0)^2},$$

pa je određivanje udaljenosti točke $T_0(x_0, y_0)$ do grafa funkcije $q$ zadane na intervalu $[a, b]$ problem određivanja globalnog minimuma funkcije $d_2$ na segmentu $[a, b]$ (vidi *Sliku 2.a*).

Budući da je funkcija $x \mapsto \sqrt{x}$ monotono rastuća funkcija, onda se naš problem svodi na određivanje globalnog minimuma funkcije

$$f(x) = (x - x_0)^2 + (q(x) - y_0)^2.$$




Slika 2. a) Funkcija $d_2$ b) Funkcija $f$

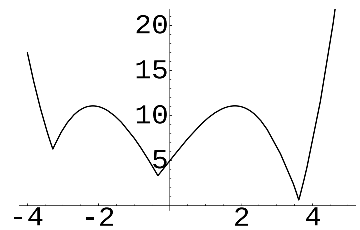
Dakle, u ovom primjeru radi se o problemu minimizacije "glatke" (višestruko derivabilne) funkcije jedne varijable (polinoma 6. stupnja) – tzv. problem jednodimenzionalne minimizacije. Iz slike se može uočiti da ova funkcija ima tri lokalna minimuma od kojih je jedan ujedno i globalni minimum.

**Primjer 1.5.** *Treba izračunati $L_1$ udaljenost točke $T_0(3,4)$ do kubne parabole $q(x) = .5x^3 - 6x + 2$ (vidi* Sliku 1*).*

$L_1$ udaljenost točke $T_0(x_0, y_0)$ do neke točke $T(x, f(x))$ na grafu funkcije $q$ zadana je s

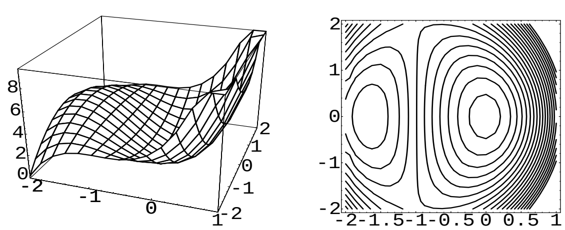
$$d_1(x) = d_1(T_0(x_0, y_0), T(x, q(x)) = |x - x_0| + |q(x) - y_0|,$$

pa je određivanje udaljenosti točke $T_0(x_0, y_0)$ do grafa funkcije $q$ zadane na intervalu $[a, b]$ problem određivanja globalnog minimuma funkcije $d_1$ na segmentu $[a, b]$ (vidi *Sliku 3.*). Dakle, i u ovom primjeru radi se također o problemu jednodimenzionalne minimizacije, ali ovaj puta je minimizirajuća funkcija nediferencijabilna (ima "špiceve"). Također iz slike se može uočiti da ova funkcija ima tri lokalna minimuma od kojih je jedan ujedno i globalni minimum.



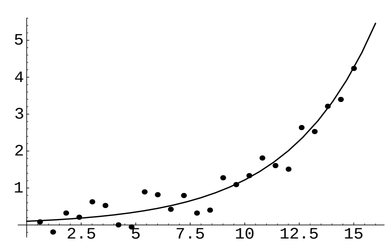

Slika 3 Funkcija $d_1$

**Primjer 1.6.** *Funkcija* $f(x_1,x_2)=2x_1^3+x_1x_2^2+5x_1^2+x_2^2$ *ima četiri stacionarne točke:* $T_1(0,0)$, $T_2(-\frac{5}{3},0)$, $T_3(-1,2)$, $T_4(-1,-2)$. *Može se pokazati da je* $T_1$ *točka minimuma,* $T_2$ *točka maksimuma, a da su* $T_3$ *i* $T_4$ *sedlaste točke. Na* Slici 4 *prikazana je ploha i* `contourPlot` *ove funkcije.*



Slika 4 Ploha i `ContourPlot` funkcije $f(x_1,x_2)=2x_1^3+x_1x_2^2+5x_1^2+x_2^2$

**Primjer 1.7.** *Zadani su podaci mjerenja* $(x_i,y_i)\; i=1,\ldots,m$. *Treba odrediti optimalne parametre* $b^*,c^*$ *eksponencijalne funkcije-modela* $f(t;b,c)=b\,e^{ct}$*, tako da suma kvadrata izmjerenih od teoretskih vrijednosti bude minimalna (vidi* Sliku 4*).*




Slika 5 Podaci i najbolja $L_2$ eksponencijalna funkcija-model

Optimalne parametre $b^*,c^*$ eksponencijalne funkcije-modela $f(t;b,c)=b\,e^{ct}$ odredit ćemo tako da minimiziramo funkcional

$$F(b,c)=\frac{1}{2}\sum_{i=1}^{m}\left(b\,e^{cx_i}-y_i\right)^2.$$

Primijetite da se u ovom slučaju radi o problemu minimizacije "glatke" (višestruko derivabilne) funkcije dviju varijabli – tzv. problem višedimenzionalne minimizacije.

# 2. Metode jednodimenzionalne minimizacije

Zadana je funkcija $f : [a, b] \to \mathbb{R}$, koja općenito nije derivabilna i koja u nepoznatoj točki $x^*$ postiže strogi lokalni minimum. Problem pronalaženja točke $x^*$ ne možemo riješiti primjenom metoda koje se zasnivaju na poznavanju derivacije funkcije $f$. Razmotrit ćemo dvije jednostavne metode za određivanje minimuma takve funkcije,metodu zlatnog reza i metodu parabole. Pretpostavimo još da poznajemo graf funkcije $f$. Od dodatnih svojstava na funkciju $f : [a, b] \to \mathbb{R}$ zahtijevat ćemo tzv. svojstvo unimodalnosti (vidi primjerice [10]).

**Definicija 2.1.** *Za funkciju $f : [a, b] \to \mathbb{R}$ kažemo da je* unimodalna na intervalu [a,b] *ako $f$ postiže minimum u nekoj točki $x^* \in [a, b]$ i ako za svake dvije točke $x_1 < x_2$ vrijede svojstva:*

*i) ako je $x_1 < x_2 \le x^*$ tada je $f(x_1) > f(x_2)$ tj. funkcija je padajuća,*

*ii) ako je $x^* \le x_1 < x_2$ tada je $f(x_1) < f(x_2)$ tj. funkcija je rastuća.*

Primijetite da poznavanjem grafa funkcije $f$ lako možemo provjeriti svojstvo unimodalnosti funkcije $f$ na intervalu $[a, b]$.

## 2.1. Metoda zlatnog reza

Neka je $f : [a, b] \to \mathbb{R}$ zadana funkcija te neka je $[a_{k-1}, b_{k-1}] \subset [a, b]$ interval takav da je $f : [a_{k-1}, b_{k-1}] \to \mathbb{R}$ unimodalna na $[a_{k-1}, b_{k-1}]$. Pretpostavimo nadalje da je $x^* \in [a_{k-1}, b_{k-1}]$ nepoznata točka u kojoj se postiže lokalni minimum funkcije $f$ na segmentu $[a_{k-1}, b_{k-1}]$. Vrijednost funkcije izračunat ćemo u dvije točke $y_k < z_k$, gdje su $y_k, z_k \in [a_{k-1}, b_{k-1}]$ za koje vrijede sljedeći uvjeti

I. $y_k$ i $z_k$ jednako su udaljeni od krajeva segmenta $[a_{k-1}, b_{k-1}]$ tj. vrijedi

$$y_k - a_{k-1} = b_{k-1} - z_k,$$

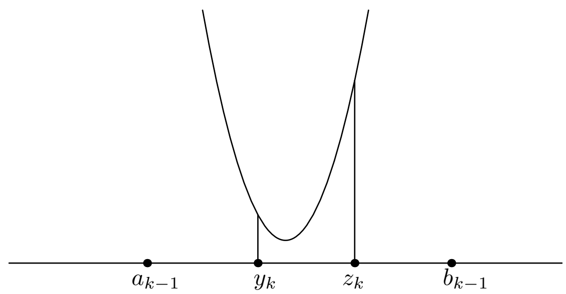
II. $y_k$ je "bliže" lijevom rubu segmenta $[a_{k-1}, b_{k-1}]$, a $z_k$ je "bliže" desnom rubu segmenta $[a_{k-1}, b_{k-1}]$, tj. za neki $c \in (0, \frac{1}{2})$ (koji ćemo kasnije precizno odrediti) vrijedi

$$y_k = a_{k-1} + c(b_{k-1} - a_{k-1}), \tag{2.1}$$

$$\begin{aligned} z_k &= b_{k-1} - c(b_{k-1} - a_{k-1}) \\ &= a_{k-1} + (1 - c)(b_{k-1} - a_{k-1}) \\ &= a_{k-1} + b_{k-1} - y_k \end{aligned} \tag{2.2}$$

III. Mogu pri tome nastupiti dva slučaja

Slučaj 1. (vidi *Sliku 6.*)




Slika 6. *Slučaj kada vrijedi* $f(y_k) \leq f(z_k)$

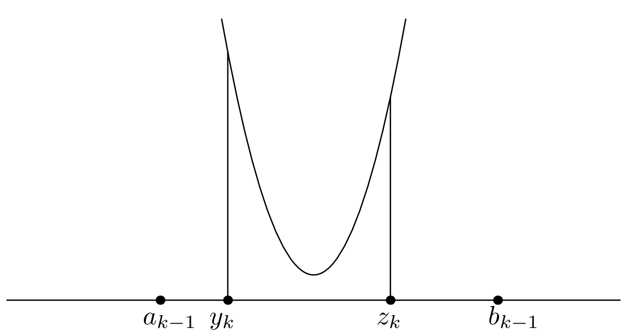
Ako je $f(y_k) \leq f(z_k)$ zbog unimodalnosti funkcije $f$ je $x^* \in [a_{k-1}, z_k]$ pa stavljamo

$$\begin{aligned} a_k &:= a_{k-1} \\ b_k &:= z_k \\ x_k &:= y_k. \end{aligned}$$

Sada je $x_k$ nova aproksimacija točke $x^*$, koja je bolja od prethodne $x_{k-1}$.

Slučaj 2. (vidi *Sliku 7.*) Ako je $f(y_k) > f(z_k)$ zbog unimodalnosti funkcije $f$ je $x^* \in [y_k, b_{k-1}]$, pa stavljamo

$$\begin{aligned} a_k &:= y_k \\ b_k &:= b_{k-1} \\ x_k &:= z_k. \end{aligned}$$




Slika 7. *Slučaj kada vrijedi* $f(y_k) > f(z_k)$

Sada smo došli do segmenta $[a_k, b_k]$ za koji očito vrijedi $[a_k, b_k] \subset [a_{k-1}, b_{k-1}]$. Opisani postupak možemo iterirati sve dok lijevi i desni kraj segmenta u kojem se nalazi točka $x^*$ ne postanu dovoljno blizu. Kako vrijednost funkcije ne bismo u svakom intervalu morali računati u dvije točke proširit ćemo III. i to na sljedeći način. U *slučaju 1.* (kada je $f(y_k) \leq f(z_k)$) stavljamo još

$$z_{k+1} := y_k, \tag{2.3}$$

a $y_{k+1}$ računamo iz formule (2.1) (zamjenom indeksa $k$ s indeksom $k+1$). Primijetite da smo ovdje pretpostavili još jedan uvjet, a to je da točka $y_k$ leži "bliže" točki $z_k$ nego točka $a_{k-1}$, što će biti zadovoljeno dobrim izborom broja $c$. U *slučaju 2.* (kada je $f(y_k) > f(z_k)$) stavljamo još

$$y_{k+1} := z_k, \tag{2.4}$$

a $z_{k+1}$ računamo iz formule (2.2) (zamjenom indeksa $k$ s indeksom $k+1$). Ovdje smo još pretpostavili da točka $z_k$ leži "bliže" točki $y_k$ nego točka $b_{k-1}$. Preostaje još jedino problem kako odrediti odgovarajući $c \in (0, \frac{1}{2})$, ako uopće postoji. Promatrajmo *slučaj 1.* iz III. Iz relacije (2.2) zamjenom indeksa $k$ s indeksom $k+1$ slijedi

$$z_{k+1} = a_k + (1-c)(b_k - a_k). \tag{2.5}$$

Redom iz relacije (2.3) i (2.1) i (2.2) imamo $z_{k+1} = y_k = a_{k-1} + c(b_{k-1} - a_{k-1}) = a_{k-1} + (z_k - a_{k-1})\frac{c}{1-c}$. Kako je u ovom slučaju $a_k = a_{k-1}$ i $b_k = z_k$ slijedi

$$z_{k+1} = a_k + \frac{c}{1-c}(a_k - b_k). \tag{2.6}$$

Uspoređuju relacije (2.5) i (2.6) dobivamo

$$\frac{c}{1-c} = 1 - c,$$

odakle zbog $c \in (0, \frac{1}{2})$ slijedi $c = \frac{3-\sqrt{5}}{2}$.

Provjerimo, zadovoljava li ovakav $c$ uvjet da točka $y_k$ leži "bliže" točki $z_k$ nego točka $a_{k-1}$. Prema (2.1) i (2.2) imamo $z_k - y_k = a_{k-1} + (1-c)(b_{k-1} - a_{k-1}) - a_{k-1} - c(b_{k-1} - a_{k-1}) = (1-2c)(b_{k-1} - a_{k-1}) = (\sqrt{5} - 2)(b_{k-1} - a_{k-1})$, dok je $y_k - a_{k-1} = a_{k-1} + c(b_{k-1} - a_{k-1}) - a_{k-1} = c(b_{k-1} - a_{k-1}) = \frac{3-\sqrt{5}}{2}(b_{k-1} - a_{k-1})$. Sada je očito $z_k - y_k < y_k - a_{k-1}$.

**Zadatak 2.1.** *Analizirajući* slučaj 2. *iz III. pokažite da je također* $c = \frac{3-\sqrt{5}}{2}$.

**Primjedba 2.1.** *Promatramo li segment* $[a, b]$ *te u njemu odredimo točku* $y = a + c(b-a)$*, gdje je* $c = \frac{3-\sqrt{5}}{2}$ *kažemo da točka* $y$ *dijeli segment* $[a, b]$ *u* omjeru zlanog reza.

**Zadatak 2.2.** *Dokažite da vrijedi* $\frac{b-a}{b-y} = \frac{b-y}{y-a}$.

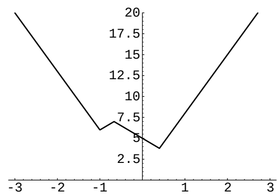
**Primjedba 2.2.** *Metoda zlatnog reza specijalan je slučaj Fibbonacijeve metode više o tome vidi u* [10], [21], [23].

Izgradimo sada odgovarajući *Mathematica* – modul za određivanje točke $x^*$ u kojoj se postiže minimum zadane funkcije $f$.

```
ln[1]:=ZlatniRez[f_, a1_, b1_, eps_] := Module[{c, i, y, z, x, a, b},
                a = a1; b = b1;
                 c = N[(3 - Sqrt[5])/2];
                   y = a + c(b - a);
                     z = a + b - y;
                If[f[y] <= f[z], a = a, a = y];
                If[f[y] <= f[z], b = z, b = b];
                  i = 1;
        Print["Iteracija     a     b      x       f[x]        greska"];
        While[b - a > eps,
          If[f[y] <= f[z], z = y; f[z] = f[y]; y = a + c(b - a); x = y, y = z;
             f[y] = f[z]; z = a + b - z; x = z];
```

```
If[f[y] <= f[z], a = a, a = y];
   If[f[y] <= f[z], b = z, b = b];
      greska = b - a;
 Print[i, "     ",a, "          ",b, "         ",x,"        ",f[x], "       ",greska];
 i++]]
```

**Primjer 2.1.** *Zadana je funkcija* $f : \mathbb{R} \to \mathbb{R}$ *zadana s* $f(x) = 5|x+1| - |3x+2| + |5x-2|$.




Graf funkcije $f(x) = 5|x+1| - |3x+2| + |5x-2|$

*Primjenom Mathematica-modula* `ZlatniRez[f,a,b,eps]` *odredit ćemo točku minimuma funkcije* $f$ *tako da je* $a = 0.1$, $b = 2$ *i* $eps = 10^{-3}$.

```
ln[2]:=f[t_]:=5 Abs[t+1]-3 Abs[3 t+2]+Abs[5 t-2];
       a=0.1;  b=2;  eps=10^(-3);
       ZlatniRez[f,a,b,eps];
```

Dobivamo sljedeću tablicu:

| Iteracija | a | b | x | f(x) | greška |
|---|---|---|---|---|---|
| 1 | 0.1 | 0.825735 | 0.54829 | 4.8397 | 0.725735 |
| 2 | 0.1 | 0.548529 | 0.377206 | 3.86838 | 0.448529 |
| 3 | 0.271323 | 0.548529 | 0.271323 | 4.166838 | 0.277206 |
| 4 | 0.271323 | 0.442646 | 0.442646 | 4.09852 | 0.171323 |
| 5 | 0.336762 | 0.442646 | 0.336762 | 3.98971 | 0.105883 |
| 6 | 0.377206 | 0.442646 | 0.402202 | 3.81541 | 0.0654395 |
| 7 | 0.377206 | 0.41765 | 0.41765 | 3.92355 | 0.0404438 |
| 8 | 0.392654 | 0.41765 | 0.392654 | 3.82204 | 0.0249957 |
| 9 | 0.392654 | 0.408103 | 0.408103 | 3.85672 | 0.0154482 |
| 10 | 0.392654 | 0.402202 | 0.398555 | 3.86433 | 0.0095475 |
| 11 | 0.396301 | 0.402202 | 0.396301 | 3.81111 | 0.00590068 |
| 12 | 0.398555 | 0.402202 | 0.399948 | 3.80016 | 0.00364682 |
| 13 | 0.398555 | 0.400809 | 0.400809 | 3.80566 | 0.00225386 |
| 14 | 0.399416 | 0.400809 | 0.399416 | 3.80175 | 0.00139296 |
| 15 | 0.399416 | 0.400277 | 0.400277 | 3.80194 | 0.000860897 |

Tablica 1. *Iterativni postupak minimizacije funkcije* $f$ *primjenom Mathematica-modula* `Zlatnirez[f,a,b,eps]`

Nakon 15 iteracija dobivamo da je $x^* \approx 0.400277$.

## 2.2. Metoda parabole

Razmotrit ćemo još jednu jednostavnu metodu za određivanje točke $x^*$ u kojoj se postiže minimum zadane funkcije $f : [a,b] \to \mathbb{R}$, $x^* \in [a,b]$. Odredimo najprije realni broj $c \in <a,b>$ tako da je $f(a) > f(c)$ i $f(c) < f(b)$. Za točke $A = (a, f(a)), C = (c, f(c))$ i $B = (b, f(b))$ sa zadanim svojstvom postoji jedinstveni polinom drugog stupnja $P_2(x) = \alpha x^2 + \beta x + \gamma$ čiji graf (kvadratna parabola) prolazi zadanim točkama (vidi primjerice [18]). Koeficijente $\alpha, \beta$ i $\gamma$ polinoma $P_2$ potražit ćemo primjenom programskog paketa *Mathematica*:

```
ln[1]:=p2=InterpolatingPolynomial[{{a, f[a]}, {b, f[b]}, {c, f[c]}}, x];
            koeficijent=CoefficientList[p2];
              alfa=koeficijent[[3]]
                beta=koeficijent[[2]]
                   gama=koeficijent[[1]]
```

Dobivamo

$$\alpha = \frac{f(a)}{(a-c)(a-b)} + \frac{f(c)}{(c-a)(c-b)} + \frac{f(b)}{(b-a)(b-c)},$$

$$\beta = -\left(\frac{bf(a)+cf(a)}{(a-c)(a-b)} + \frac{af(c)+bf(c)}{(c-a)(c-b)} + \frac{af(b)+cf(b)}{(b-a)(b-c)}\right),$$

$$\gamma = \frac{cbf(a)}{(a-c)(a-b)} + \frac{abf(c)}{(c-a)(c-b)} + \frac{acf(b)}{(b-a)(b-c)}.$$

Nakon što smo odredili koeficijente polinoma $P_2$, odredit ćemo apscisu tjemena parabole. Ona glasi

$$x_T = \frac{-\beta}{2\alpha} = -\frac{b^2f(a) - c^2f(a) - a^2f(b) + c^2f(b) + a^2f(c) - b^2f(c)}{2(-bf(a) + cf(a) + af(b) - cf(b) - af(c) + bf(c))}. \tag{2.7}$$

**Zadatak 2.3.** *Algebarski izvedite formulu (2.7).*

Točku $x_T$ uzimamo kao prvu aproksimaciju točke $x^*$. Zadamo li točnost $\varepsilon$ iterativni postupak nastavljamo na sljedeći način: Ako je $|f(x_T) - P_2(x_T)| < \varepsilon$ uzimamo $x^* = x_T$ i postupak je završen. U suprotnom između četiri broja $a, c, b, x_T$ za novi $c$ biramo onaj u kojem funkcija $f$ poprima najmanju vrijednost (primijetite da to mogu biti samo $x_T$ ili $c$), dok za novi $a$ i $b$ uzimamo prvi najveći broj lijevo od novog $c$ odnosno prvi najmanji broj desno od novog $c$. Novi $x_T$ sada dobivamo primjenom formule (2.7). Izgradimo odgovarajući *Mathematica-modul*.

```
ln[2]:=Parabola[f_, a1_, b1_, c1_, epsilon_] :=
               Module[{c, b, a, x1, x2, x3, p2, x, pogreska, xT, xt, greska, lista,
                       sortiranalista, pozicija},
        c = c1; b = b1; a = a1;
        p2[x_, x1_, x2_, x3_] :=
           InterpolatingPolynomial[{{x1, f[x1]}, {x2, f[x2]}, {x3, f[x3]}}, x];
        tjeme[x1_, x2_, x3_] := -(x2^2 f[x1] - x3^2 f[x1] - x1^2f[x2] +
                                 x3^2f[x2] + x1^2 f[x3] - x2^2 f[x3])/(2 (-x2 f[x1] + x3f[x1] +
                                 x1 f[x2] - x3 f[x2] - x1 f[x3] + x2 f[x3]);
        pogreska[xt_, x1_, x2_, x3_] := Abs[f[xt] - p2[xt, x1, x2, x3]];
            xT = tjeme[a, b, c] // N;
               greska = pogreska[xT, a, b, c];
        Print["Iteracija     a     c      b       xT      f[xT]     greska"];

       sl[1] = Plot[{f[x], p2[x, a, b, c]}, {x, a-0.5 (b-a), b+0.5 (b-a)},
                    PlotStyle -> {Hue[0.1], {Thickness[0.015], Hue[0.6]}},
                      DisplayFunction -> Identity];
        i = 1;
       While[Abs[f[xT] - p2[xT, a, b, c]] > epsilon,
              lista = {a, b, c, xT};
                 If[f[c] > f[xT], xT, c];
                   sortiranalista = Sort[lista];
```

```
                pozicija = Flatten[Position[sortiranalista,c]][[1]];
                  b=sortiranalista[[pozicija + 1]];
                   a = sortiranalista[[pozicija - 1]];
                    xT = tjeme[a, b, c] // N;
          greska = pogreska[xT, a, b, c];
    Print[i, "     ", a, "     ", c, "  ",  b, "   ", xT, "  ",f[xT], "   " ,greska];
  sl[i+1] = Plot[{f[x], p2[x, a, b, c]}, {x, a-0.5 (b-a), b+0.5 (b-a)},
                 PlotStyle -> {Hue[0.1], {Thickness[0.015], Hue[0.6]}},
                   DisplayFunction -> Identity];
         i++];
      ]
```

**Primjer 2.2.** *Zadana je funkcija $f : \mathbb{R} \to \mathbb{R}$ zadana s $f(x) = 5|x+1| - |3x+2| + |5x-2|$ iz* Primjera 2.1.. *Primjenom Mathematica-modula* `Parabola[f,a,b,c,eps]` *odredit ćemo točku minimuma funkcije $f$ tako da je $a = -2$, $c = 0$, $b = 2$ i $\varepsilon = 10^{-3}$. Primijetite da je zadovoljena relacija $f(a) > f(c)$ i $f(c) < f(b)$.*

```
ln[2]:=f[x_]:=5 Abs[x+1]-Abs[3 x+2]+Abs[5 x-2];
          a=-2;  c=0;  b=2;  eps=10^(-3);
            Parabola[f,a,b,c,eps];
```

Dobivamo sljedeću tablicu:

| Iteracija | a | c | b | $x_T$ | $f(x_T)$ | greška |
|---|---|---|---|---|---|---|
| 1 | −0.111111 | 0 | 2 | 0.340278 | 3.97917 | 0.0.582051 |
| 2 | 0 | 0.340278 | 2 | 0.481337 | 4.36936 | 0.486099 |
| 3 | 0 | 0.340278 | 0.481337 | 0.295353 | 4.11394 | 0.158951 |
| 4 | 0.295353 | 0.340278 | 0.481337 | 0.366197 | 3.90441 | 0.0569293 |
| 5 | 0.340278 | 0.366197 | 0.481337 | 0.38319 | 3.85043 | 0.0365175 |
| 6 | 0.366197 | 0.38319 | 0.481337 | 0.395534 | 3.81342 | 0.0260649 |
| 7 | 0.38319 | 0.395534 | 0.481337 | 0.404892 | 3.83245 | 0.0293058 |
| 8 | 0.38319 | 0.395534 | 0.404892 | 0.395589 | 3.81323 | 0.000164881 |

Tablica 2. *Iterativni postupak minimizacije funkcije $f$ primjenom Mathematica–modula* `Parabola[f,a,b,c,eps]`

Nakon 8 iteracija dobivamo da je $x^* \approx 0.395589$. Za svaku iteraciju iscrtajmo još graf funkcije $f$ i grafove odgovarajućih kvadratnih parabola.

```
ln[3]:=slike=Partition[Table[sl[j],{j,i}],3];
         Show[GraphicsArray[slike],DisplayFunction->$DisplayFunction];
```

Složimo li graf funkcije $f$ i grafove odgovarajućih kvadratnih parabola jednu ispod druge primjenom naredbe

```
ln[4]:=Table[Show[sl[j], DisplayFunction -> $DisplayFunction], {j, i}];
```

te selektiramo vanjsku ćeliju i pritisnemo tipke `Ctrl+Y` dobivamo animaciju funkcije $f$ i grafova odgovarajućih kvadratnih parabola.

**Primjedba 2.3.** *U primjenama često se koristi tzv.* Brentova metoda *za jednodimenzionalnu minimizaciju (vidi Brent, [8]) koja se zasniva kombinaciji metode zlatnog reza i metode parabole.*

## 2.3. Newtonova metoda tangenti

Pretpostavimo da je $f \in C^2[a,b]$ i da je $x^\star \in (a,b)$ točka lokalnog minimum funkcije $f$. To znači da je $f'(x^\star) = 0$ i $f''(x^\star) > 0$. I obrnuto, ako je $f'(x^\star) = 0$ i $f''(x^\star) > 0$, onda funkcija $f$ u točki $x^\star$ postiže lokalni minimum.

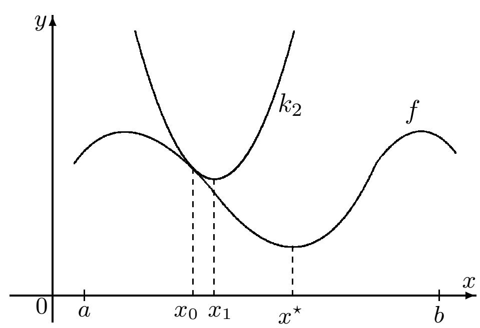
Iz navedenog slijedi da se lokalni minimum $x^\star$ funkcije $f$ može tražiti tako da najprije lokaliziramo interval u kome funkcija postiže lokalni minimum. Nakon toga na tom intervalu riješimo jednadžbu $f'(x) = 0$ primjenom neke od metoda za rješavanje nelinearne jednadžbe (vidi primjerice [18]). Točka $x^\star$ za koju vrijedi $f'(x^\star) = 0$ naziva se stacionarna točka funkcije $f$.

Ako za rješavanje jednadžbe $f'(x) = 0$ primijenimo Newtonovu metodu uz odgovarajući izbor početne aproksimacije $x^0$ (vidi primjerice [18]), dobivamo iterativnu metodu koju u literaturi možemo naći pod nazivom Newtonova metoda za jednodimenzionalnu minimizaciju

$$x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}, \quad n = 0, 1, \dots \tag{2.8}$$

Istu formulu mogli bi dobiti i na sljedeći način. Najprije u okolini točke $x_0 \in \Omega$) funkciju $f$ aproksimiramo kvadratnom funkcijom pomoću Taylorove formule

$$k_2(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2$$




*Newtonova metoda*

Sljedeću aproksimaciju $x_1$ tražene točke $x^\star$ birat ćemo tako da odredimo minimum navedene kvadratne funkcije $k_1$:

$$x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}.$$

Ponavljajući postupak dobivamo niz $x_0, x_1, \dots, x_n, \dots$ zadan rekurzivnom formulom (4.3), koji uz neke uvjete (vidi primjerice Dennis (1996)) konvergira prema $x^\star$.

U tu svrhu sagradit ćemo *Mathematica*-modul kome ćemo predati funkciju $f$, početnu aproksimaciju $x_0$, traženu točnost *eps*, te maksimalno dozvoljen broj iteracija *it*. Iterativni proces zaustavit ćemo za neki $k$ ako bude $|f'(x_k)| < eps$ ili ako $k > it$. Na taj način dobit ćemo zapravo stacionarnu točku funkcije $f$. Ako je $|f'(x_k)| < eps$ i ako je $|f''(x_k)| > 0$, smatrat ćemo da smo pronašli točku minimuma funkcije $f$ s točnošću *eps*. Kao rezultat možemo ispisivati svaku iteraciju, te konačni rezultat.

```
In[1]:= NewtonMin1[f_,x0_,eps_,it_]:=Module[{x=N[x0],k=0},
            While[ Abs[f'[x]] > eps && k < it,
                Print["x",k,"=",x,",   f'',k,"=",f[x]];
                x=x-f'[x]/f''[x];
```

```
            k=k+1
                ];
        If[f''[x]<0,Print["Pronadjena je stacionarna tocka"]];{x,f[x]}
                  ];
```

U ovom primjeru modul ćemo koristiti na sljedeći način

```
In[2]:= q[x_]:=.5 x^3-6 x+2; x0=3; y0=4;
        f[x_]:=(x - x0)^2+ (q[x] - y0)^2
        NewtonMin1[f,4,0.005,10]

        x0 = 4.         f0 = 37.
        x1 = 3.72544,   f1 = 2.77591
        x2 = 3.62874,   f2 = 0.40942
        x3 = 3.61692,   f3 = 0.382444
Out[4]:= {3.61676,0.382438}
```

Ako stavimo `NewtonMin1[f, -3.5, -005, 10]`, dobivamo drugi lokalni minimum `{-3.2173,39.0807}`, a za `NewtonMin1[f, 2, -005, 10]`, pronađena je stacionarna točka $\{1.98298\}$ u kojoj funkcija postiže lokalni maksimum.

Lokalni minimum iste funkcije možemo potražiti direktnom primjenom *Mathematica*-modula

```
In[4]:= FindMinimum[f[x], {x, 4}]
Out[4]:= {0.382438,{x->3.61676}}
```

Dakle, globalni minimum funkcije $f$ postiže se u točki $x^\star = 3.61676$, pri čemu je $f(x^\star) = 0.382438$. To znači da je udaljenost točke $T_0$ do grafa kubne parabole $q(x) = .5x^3 - 6x + 2$ jednaka $d(T_0, q) = \sqrt{f(x^\star)} = 0.618416$, pri čemu se taj minimum postiže u točki $T^\star(x^\star, q(x^\star)) = (3.61676, 3.95472)$.

# 3. Gradijentne metode

Iterativni proces za traženje točke lokalnog minimuma dovoljno puta neprekidno diferencijabilne funkcije $f : \mathbb{R}^n \to \mathbb{R}$ općenito je oblika

$$x_{k+1} = x_k + \alpha_k p_k, \quad k = 0, 1, \ldots , \tag{3.1}$$

gdje je $x_0$ početna aproksimacija, $p_k$ vektor smjera kretanja od točke $x_k$ u točku $x_{k+1}$, a $\alpha_k > 0$ duljina koraka u smjeru $p_k$.

Kretanje od točke $x_k$ u točku $x_{k+1}$ treba ostvariti tako da se postigne smanjenje vrijednosti funkcije, tj. tako da bude $f(x_{k+1}) < f(x_k)$. To se može postići tako da vektor $p_k$ izaberemo tako da bude (vidi primjerice [10])

$$(f'(x_k), p_k) < 0, \tag{3.2}$$

gdje je $f'(x_k)$ gradijent funkcije $f$ u točki $x_k$. Naime, ako za funkciju $f$ napišemo Taylorovu formulu u okolini točke $x_k$

$$f(x) = f(x_k) + (f'_k, x - x_k) + \frac{1}{2}\left(f''\left(x_k + \vartheta(x - x_k)\right)(x - x_k), x - x_k\right), \qquad \vartheta \in (0, 1)$$

i stavimo $x = x_k + \alpha_k p_k$, uz oznaku $\xi_k = x_k + \vartheta(x - x_k)$ dobivamo

$$f(x) = f(x_k) + \alpha_k (f'_k, p_k) + \frac{\alpha_k^2}{2}\left(f''(\xi_k) p_k, p_k\right), \qquad \vartheta \in (0, 1)$$

Ako je $(f'_k, p_k) < 0$, onda za male vrijednosti parametara $\alpha_k$ vrijedi $f(x_{k+1}) := f(x_k + \alpha_k p_k) < f(x_k)$.

## 3.1. Gradijentna metoda i metoda najbržeg spusta

Još je A. Cauchy 1845. godine uočio da je smjer najbržeg pada neprekidno derivabilne funkcije $f : \mathbb{R}^n \to \mathbb{R}$ u točki $\mathbf{x}$ u smjeru antigradijenta $(-f'(x))$. Naime, ako za neki vektor $d \in \mathbb{R}^n$ definiramo diferencijabilnu funkciju $\varphi : \mathbb{R} \to \mathbb{R}$ formulom

$$\varphi(t) := f(x + td),$$

onda zbog $\varphi'(0) = (f'(x), d)$:

- za $d := \frac{f'(x)}{\|f'(x)\|}$, vrijedi $\varphi'(0) = \|f'(x)\| > 0$, pa smjer gradijenta pokazuje smjer porasta funkcije $f$;
- za $d := -\frac{f'(x)}{\|f'(x)\|}$, vrijedi $\varphi'(0) = -\|f'(x)\| < 0$, pa smjer antigradijenta pokazuje smjer pada funkcije $f$.

Na toj ideji zasniva se obična gradijentna metoda za minimizaciju funkcije $f : \mathbb{R}^n \to \mathbb{R}$

$$x^{(k+1)} = x^{(k)} + \alpha_k p^{(k)}, \quad k = 0, 1, \dots , \tag{3.3}$$

gdje je $\alpha_k > 0$ *duljina koraka* u smjeru *vektora smjera kretanja* $p^{(k)} = -f'(x^{(k)})$, koji očigledno zadovoljava uvjet (4.2). Duljina koraka $\alpha_k$ računamo sljedećim algoritmom (koji će biti opravda u dokazu *Leme* 3.1. i *Teorema* 3.1.):

Algoritam za izračunavanje duljine koraka

Korak 0. Izabrati $0 < \varepsilon < 1$ i stavi $\alpha = 1$

Korak 1. Izračunati $x = x_k + \alpha_k p_k$ i provjeriti uvjet

$$f(x) - f(x^{(k)}) \le \varepsilon\alpha(f'(x^{(k)}), p^{(k)})) \tag{3.4}$$

Korak 2. Ako je uvjet (3.4) ispunjen, staviti $\alpha_k = \alpha$; inače staviti $\alpha := h\alpha, \quad h \in (0, 1)$ i prijeći na *Korak 1.*

Najprije ćemo sagraditi *Mathematica*-modul `Korak[n,f,d,pk,xk]`, koji će izračunavati duljinu koraka $\alpha_k$ iz točke $x^{(k)})$ u smjeru vektora $p^{(k)})$. Pri tome je $n$ broj varijabli, $f$ ime minimizirajuće funkcije, a $d$ njezin gradijent.

```
In[1]:= Korak[n_, f_, pk_, xk_] := Module[{epsilon = .5, h = .5, f0, f1, fd},
              al = 1;   w = Table[x[i] -> xk[[i]], {i, n}];
              d[x_] := Table[D[f[x], x[i]], {i, n}];
              f0 = f[x] /. w;
              fd = d[x] /. w;
              While[f1 = f[x] /. Table[x[i] -> xk[[i]] + al pk[[i]], {i, n}];
                    f1 - f0 > epsilon al Apply[Plus, fd pk], al = h al];
                                       al]
```

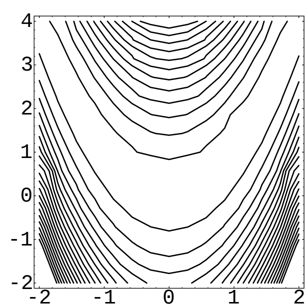
Gradijentna metoda dana je u *Mathematica*-modulu `MinGrad[n,f,d,x0,eps,it]`, gdje je $x_0$ početnu aproksimaciju, *eps* traženu točnost, a *it* maksimalno dozvoljen broj iteracija. Modul `MinGrad` poziva modul `Korak`, a iterativni postupak završava onda ako je $\|f'(x^{(k)})\|_\infty < eps$ za neki $k$ ili ako je $k > it$.

```
In[2]:= MinGrad[n_, f_, x0_, eps_, it_]:= Module[{k = 0, xs = x0, xn, gr, w},
                w = Table[x[i] -> x0[[i]], {i, n}];
                d[x_] := Table[D[f[x], x[i]], {i, n}];
        While[ w = Table[x[i] -> xs[[i]], {i, n}];  grad = Max[Abs[d[x] /. w]];
               grad > eps && k < it,
               pk = -d[x] /. w;
               alpha = Korak[n, f, pk, xs];
               Print["x", k, "=", xs, "   f=", f[x] /. Table[x[i] -> xs[[i]], {i, n}],
               "   |grad|=", grad, "   alpha=", alpha];
           xn = xs + alpha pk;
           xs = xn; k = k + 1;
              ];
                {xn, f[x] /. Table[x[i] -> xn[[i]], {i, n}]}]
```

Modul `MinGrad` testirati ćemo na poznatoj *Rosenbrockovoj* funkciji, koja postiže minimum u točki $x^\star = (0,0)$ i za koju ćemo odmah definiradi gradijent i nacrtati nivo krivulje u okolini minimuma, a djelomične rezultate izvođenja programa za $x_0 = (.5, .5)$ dati u *Tablici 1.*

```
In[3]:= n = 2;  f[x_] := 100(x[2] - x[1]^2)^2 + (1 - x[1])^2
                f[x] /. {x[1] -> x1, x[2] -> x2}
                d[x_] := Table[D[f[x], x[i]], {i, n}]
                d[x] /. {x[1] -> x1, x[2] -> x2}
        ContourPlot[f[x] /. {x[1] -> x1, x[2] -> x2}, {x1, -2, 2}, {x2, -2, 4},
                    Contours -> 20, ContourShading -> None];

Out[3]:= (1 - x1)^2 + 100(-x1^2 + x2)^2
         {-2(1 - x1) - 400 x1(-x1^2 + x2), 200(-x1^2 + x2)}
```




| $k$ | $(x^{(k)})$ | $f(x^{(k)})$ | $\|f'(x^{(k)})\|_\infty$ | $\alpha_k$ |
|---|---|---|---|---|
| 0 | (0.5,0.5) | 6.5 | 51 | 0.00195313 |
| 1 | (0.599609, 0.402344) | 0.343602 | 11.0691 | 0.00195313 |
| 2 | (0.621229, 0.38562) | 0.143477 | 0.681796 | 0.00195313 |
| 10 | (0.703733, 0.492322) | 0.0886257 | 0.58356 | 0.00195313 |
| 100 | (0.785646, 0.615674) | 0.0461927 | 0.313171 | 0.00195313 |
| 500 | (0.895571, 0.801784) | 0.0109123 | 0.114267 | 0.00195313 |
| 1000 | (0.984068, 0.968167) | 0.000258777 | 0.0556747 | 0.03125 |

Tablica 1.

Sljedeća lema pokazuje da gradijentna metoda (3.3) osigurava konvergenciju iterativnog procesa ili prema inf $f$ ili prema nekoj stacionarnoj točki funkcije $f$.

**Lema 3.1.** *Ako je*

*(i) funkcija* $f : \mathbb{R}^n \to \mathbb{R}$ *ograničena odozdo*

*(ii) gradijent* $f'$ *zadovoljava Lipschitzov uvjet*

$$\|f'(x) - f'(y)\| \leq R\|x - y\| \quad \text{za sve } x, y \in \mathbb{R}^n, \tag{3.5}$$

*(iii) parametri* $\alpha_k$ *biraju se iz uvjeta (3.4),*

*tada za gradijentnu metodu (3.3) vrijedi*

$$\|f'(x_k)\| \to 0, \text{ za } k \to \infty$$

*neovisno o izboru početne aproksimacije* $x_0$*, a parametri* $\alpha_k$ *uvijek se mogu izabrati prethodno opisanim Algoritmom za izračunavanje duljine koraka.*

Dokaz. Prema Lagrangeovom teoremu o srednjoj vrijednosti postoji $\vartheta \in (0,1)$, tako da bude

$$f(x_k + (x - x_k)) - f(x_k) = (f'(x_k + \vartheta(x - x_k)), x - x_k).$$

Dodavanjem i oduzimanjem $f'(x_k)(x - x_k)$ na desnoj strani, t označavanjem: $\xi := x_k + \vartheta(x - x_k)$, $f_k := f(x_k)$, $f'_k := f'(x_k)$, $f'_\xi := f'(\xi)$ ovu jednakost možemo zapisati:

$$f(x) - f_k = (f'_k, x - x_k) + (f'_\xi - f'_k, x - x_k)$$

Za $x = x_k - \alpha f'_k$ korištenjem Cauchy-Schwartzove nejednakosti i uvjeta (3.5) dobivamo

$$\begin{aligned} f(x) - f_k &= -\alpha(f'_k, f'_k) - \alpha(f'_\xi - f'_k, f'_k) \quad \left[-\alpha(f'_k, f'_k) + \alpha(f'_k - f'_\xi, f'_k)\right] \\ &\leq -\alpha\|f'_k\|^2 + \alpha\|f'_\xi - f'_k\|\|f'_k\| \leq -\alpha\|f'_k\|^2 + \alpha R\|\xi - x_k\|\|f'_k\| \quad [\xi - x_k = \vartheta(x - x_k)] \\ &\leq -\alpha\|f'_k\|^2 + \alpha R\|x - x_k\|\|f'_k\| = -\alpha\|f'_k\|^2 + \alpha R\alpha\|f'_k\|\|f'_k\| \\ &= \alpha\|f'_k\|^2(\alpha R - 1) \end{aligned}$$

Kako je $R < \infty$, za $\varepsilon \in (0,1)$ iz Algoritma za izračunavanje duljine koraka parametar $\alpha$ možemo uzeti tako da bude $\alpha > 0$ i $\alpha R - 1 \leq -\varepsilon$. Zato iz prethodne nejednakosti dobivamo

$$f(x) - f_k \leq \alpha(f'_k, f'_k)(-\varepsilon) \Longrightarrow f(x) - f_k \leq \varepsilon\alpha(f'_k, -f'_k), \tag{3.6}$$

što je upravo uvjet (3.4). Drugim riječima uvjet (3.4) bit će ispunjen za $\alpha \leq \frac{1-\varepsilon}{R}$.

Ako stavimo $x := x_{k+1}$, onda prethodna nejednakost pokazuje da je $f_{k+1} < f_k$ za svaki $k = 0, 1, \ldots$ uz uvjet da je $\|f'_k\| \neq 0$, tj niz $(f_n)$ je monotono padajuć. Kako je po pretpostavci funkcija $f$ ograničena odozdo, onda to znači da je i niz $(f_n)$ ograničen odozdo. Dakle, niz $(f_n)$ je konvergentan, a onda i Cauchyjev, pa možemo zaključiti da vrijedi

$$f_{k+1} - f_k \to 0 \quad \text{za } k \to \infty.$$

što uz nejednakost $\left[f_{k+1} - f_k \leq -\varepsilon\alpha\|f'_k\|^2\right]$

$$\|f'_k\|^2 \leq \frac{f_k - f_{k+1}}{\varepsilon\alpha_k}$$

koja slijedi iz nejednakosti (3.6) za $x = x_{k+1}$ dokazuje tvrdnju leme. ♣

**Teorem 3.1.** *Neka je* $f \in C^2(\mathbb{R}^n)$ *i*

$$m\|y\|^2 \leq (f''(x)y, y) \leq M\|y\|^2, \quad M > m > 0 \qquad \text{za sve } x, y \in \mathbb{R}^n \tag{3.7}$$

*Ako se niz* $(x_n)$ *gradi pomoću iterativnog procesa (3.3), a parametri* $\alpha_k$ *biraju iz uvjeta (3.4), tada neovisno o izboru početne aproksimacije* $x_0 \in \mathbb{R}^n$ *vrijedi*

$$x_k \to x^* \quad i \quad f(x_k) \to f(x^*),$$

*gdje je* $x^* \in \mathbb{R}^n$ *jedinstvena točka minimuma funkcije* $f$*. Pri tome vrijede ocjene:*

$$\|x - x^*\| \leq \frac{\|f'(x)\|}{m}, \qquad f_k - f^* \leq g^k(f_0 - f^*), \qquad \|x_k - x^*\| \leq C\, q^{\frac{k}{2}}, \tag{3.8}$$

*gdje je* $C < \infty, \quad q \in (0, 1)$.

Dokaz. Prema *Teoremu* 1.1. lijevi dio nejednakosti (3.7) ekvivalentan je činjenici da je $f$ jako konveksna funkcija. Nadalje, prema *Teoremu* 1.2. funkcija $f$ ograničena je odozdo i postoji jedinstvena točka minimuma $x^* \in \mathbb{R}^n$. Zato još treba dokazati da niz $(x_n)$ konvergira prema $x^*$ neovisno o izboru početne aproksimacije $x_0 \in \mathbb{R}^n$ i dokazati ocjene (3.8).

Pokažimo najprije da niz $(x_n)$ konvergira prema $x^*$. Primjenom Taylorove formule za funkciju $f$ u okolini točke $x$, za $x^*$ dobivamo

$$f(x^*) = f(x) + (f'(x), x^* - x) + \frac{1}{2}\,(f''(\xi)(x^* - x), x^* - x)$$

a odavde korištenjem (3.7) dobivamo

$$f(x) - f(x^*) = (f'(x), x - x^*) - \frac{1}{2}\,(f''(\xi)(x^* - x), x^* - x) \leq (f'(x), x - x^*) - \frac{m}{2}\|x - x^*\|^2,$$

što uz primjenu Cauchyjeve nejednakosti daje

$$f(x) - f(x^*) \leq \|f'(x)\|\|x - x^*\| - \frac{m}{2}\|x - x^*\|^2. \tag{3.9}$$

Primjenom Taylorove formule za funkciju $f$ u okolini točke $x^*$ i korištenjem činjenice da je $f'(x^*) = 0$, dobivamo

$$f(x) - f(x^*) = \frac{1}{2}\,(f''(\zeta)(x - x^*), x - x^*)\,.$$

Koristeći ovdje (3.7) dobivamo

$$\frac{m}{2}\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{M}{2}\|x - x^*\|^2. \tag{3.10}$$

Iz (3.9) i (3.10) slijedi

$$\|f'(x)\|\|x - x^*\| - \frac{m}{2}\|x - x^*\|^2 \geq f(x) - f(x^*) \geq \frac{m}{2}\|x - x^*\|^2, \Longrightarrow m\|x - x^*\|^2 \leq \|f'(x)\|\|x - x^*\|$$

a odavde

$$\|x - x^*\| \leq \frac{\|f'(x)\|}{m}. \tag{3.11}$$

Desna strana ove nejednakosti za $x = x_k$ prema *Lemi* 3.1. konvergira prema nuli, što znači da $\|x_k - x^*\| \to 0$ za $k \to \infty$, tj. $x_k \to x^*$.

Korištenjem nejednakosti (3.10) i (3.11) u (3.9) dobivamo

$$f(x) - f(x^*) \leq \|f'(x)\| \frac{\|f'(x)\|}{m} - \frac{m}{2}\frac{2}{M}(f(x) - f(x^*)),$$

iz čega slijedi $\left[\frac{1}{m}\|f'(x)\|^2 \geq (f(x) - f(x^*)) + \frac{m}{M}(f(x) - f(x^*))\right]$

$$\|f'(x)\|^2 \geq m\left(1 + \frac{m}{M}\right)(f(x) - f(x^*)). \tag{3.12}$$

Korištenjem ove nejednakosti u (3.6) za $x = x_k - \alpha f'_k$ $\left[f(x) - f(x_k) \leq -\varepsilon\alpha\|f'_k\|^2\right]$ dobivamo

$$f_{k+1} - f_k \leq -\varepsilon\alpha_k m\left(1 + \frac{m}{M}\right)(f(x_k) - f(x^*)). \tag{3.13}$$

S druge strane primjenom Taylorove formule za funkciju $f$ u okolini točke $x_k$, za $x = x_k - \alpha f'_k$ dobivamo

$$\begin{aligned} f(x) - f(x_k) &= (f'_k, x - x^*) + \tfrac{1}{2}\left(f''(\xi)(x - x_k), x - x_k\right) = -\alpha\|f'_k\|^2 + \tfrac{\alpha^2}{2}\left(f''(\xi)f'_k, f'_k\right) \\ \text{(3.7)} \quad &\leq -\alpha\|f'_k\|^2 + \tfrac{\alpha^2}{2}M\|f'_k\|^2 = -\alpha\left(1 - \tfrac{\alpha M}{2}\right)\|f'_k\|^2. \end{aligned}$$

Ako $\varepsilon > 0$ izaberemo tako da bude $1 - \frac{\alpha M}{2} > \varepsilon$, onda iz prethodne nejednakosti dobivamo uvjet (3.4) pomoću kojeg biramo duljinu koraka $\alpha_k$

$$f(x) - f(x_k) \leq -\alpha\varepsilon\|f'_k\|^2.$$

Dakle uvjet (3.4) bit će ispunjen za $\alpha_k \leq \overline{\alpha} = \frac{2(1-\varepsilon)}{M}$.

Iz (3.13) slijedi $\left[f_{k+1} - f_k \leq -\varepsilon\alpha_k m\left(1 + \frac{m}{M}\right)(f(x_k) - f(x^*))\right]$

$$f_{k+1} - f^* = (f_{k+1} - f_k) + (f_k - f^*) \leq \left(1 - \varepsilon\alpha_k m\left(1 + \frac{m}{M}\right)\right)(f(x_k) - f(x^*)) \leq q(f_k - f^*),$$

uz oznaku $q := 1 - \varepsilon\alpha_k m\left(1 + \frac{m}{M}\right)$, što više puta primijenjeno daje

$$f_k - f^* \leq q^k(f_0 - f^*). \tag{3.14}$$

Kako je $\alpha_k \leq \overline{\alpha} = \frac{2(1-\varepsilon)}{M}$, bit će $q \geq 1 - \frac{2\varepsilon(1-\varepsilon)m}{M}\left(1 + \frac{m}{M}\right)$. Ako sada $q$ shvatimo kao realnu funkciju varijable $\varepsilon$ i pronađemo njezin minimu, dobivamo $q_{min} = 1 - \frac{m}{2M}\left(1 + \frac{m}{M}\right)$ za $\varepsilon = \frac{1}{2}$. Sada se vidi da je u uvjet (3.4) najbolje staviti $\varepsilon = \frac{1}{2}$. To će osigurati najveću moguću brzinu konvergencije gradijentne metode.

Na kraju, iz (3.10) i (3.14) izlazi

$$\|x_k - x^*\| \leq \sqrt{\frac{2}{m}}\sqrt{f_k - f^*} \leq q^{\frac{k}{2}}\sqrt{\frac{2}{m}}\sqrt{f_0 - f^*} \leq Cq^{\frac{k}{2}}.$$

♣

# 4. Metode Newtonovog tipa

Iterativni proces za traženje točke lokalnog minimuma dovoljno puta neprekidno diferencijabilne funkcije $f : \mathbb{R}^n \to \mathbb{R}$ općenito je oblika

$$x_{k+1} = x_k + \alpha_k p_k, \quad k = 0, 1, \dots , \tag{4.1}$$

gdje je $x_0$ početna aproksimacija, $p_k$ vektor smjera kretanja od točke $x_k$ u točku $x_{k+1}$, a $\alpha_k > 0$ duljina koraka u smjeru $p_k$.

Kretanje od točke $x_k$ u točku $x_{k+1}$ treba ostvariti tako da se postigne smanjenje vrijednosti funkcije, tj. tako da bude $f(x_{k+1}) < f(x_k)$. To se može postići tako da vektor $p_k$ izaberemo tako da bude (vidi primjerice [10], [16])

$$(f'(x_k), p_k) < 0, \tag{4.2}$$

gdje je $f'(x_k)$ gradijent funkcije $f$ u točki $x_k$. Naime, ako za funkciju $f$ napišemo Taylorovu formulu u okolini točke $x_k$

$$f(x) = f(x_k) + (f'_k, x - x_k) + \frac{1}{2} \left(f'' \left(x_k + \vartheta(x - x_k)\right)(x - x_k), x - x_k\right), \qquad \vartheta \in (0, 1)$$

i stavimo $x = x_k + \alpha_k p_k$, uz oznaku $\xi_k = x_k + \vartheta(x - x_k)$ dobivamo

$$f(x_k + \alpha_k p_k) = f(x_k) + \alpha_k (f'_k, p_k) + \frac{\alpha_k^2}{2} \left(f''(\xi_k) p_k, p_k\right), \qquad \vartheta \in (0, 1)$$

Ako je $(f'_k, p_k) < 0$, onda za male vrijednosti parametara $\alpha_k$ vrijedi $f(x_{k+1}) := f(x_k + \alpha_k p_k) < f(x_k)$.

## 4.1. Newtonova metoda

Pretostavimo da je promatrana funkcija $f$ strogo konveksna i dovoljno glatka i da je na neki način izabrana početna aproksimacija $x_0$. Analogno Newtonovoj metodi za jednodimenzionalnu minimizaciju u okolini točke $x_0$ funkciju $f$ aproksimirat ćemo kvadratnom funkcijom dobivenom iz Taylorove formule

$$\psi(x) = f(x_0) + (f'(x_0), x - x_0) + \frac{1}{2}(f''(x_0)(x - x_0), x - x_0)$$

Funkcija $\psi$ također je strogo konveksna i njezin minimum postiže se u točki $x_1 = x_0 - (f''(x_0)^{-1} f'(x_0)$, koju ćemo smatrati sljedećom aproksimacijom. Primijetite da smo iz točke $x_0$ u točku $x_1$ "došli" preko vektora smjera kretanja definiranog s $p = -(f''(x_0)^{-1} f'(x_0)$. Ovaj vektor definira smjer spusta iz točke $x_0$ jer zbog konveksnosti funkcije $f$ ispunjava uvjet spusta (4.2)

$$(f'(x_0), p) = (f'(x_0), -(f''(x_0)^{-1} f'(x_0)) = -(f''(x_0)p, p) < 0$$
$$\left[(f'_0, p) = (f'_0)^T p = -(f'_0)^T \left((f''_0)^{-1} f''_0\right)(f''_0)^{-1} f'_0 = -p^T f''_0 p = -(f''_0 p, p)\right]$$

Ponavljajući navedeni postupak dolazimo do iterativne metode

$$x_{k+1} = x_k - (f''(x_k)^{-1} f'(x_k), \qquad k = 0, 1, \dots ,$$

koja se u literaturi obično naziva obična Newtonova ili Newton-Raphsonova metoda (vidi primjerice [6], [8], [10], [12], [15], [16], [20]). Pri tome u implementaciji metode računanje inverzne matrice $(f''(x_k)^{-1}$ izbjegava se tako da umjesto toga u svakom koraku rješavamo jedan sustav linearnih jednadžbi:

$$x_{k+1} = x_k + p_k, \quad \text{(gdje je } p_k \text{ rješenje sustava)}$$
$$f''(x_k) p = f'(x_k).$$

Uzevši u obzir da kretanje od točke $x_k$ u smjeru vektora $p_k$ donosi smanjenje vrijednosti funkcije, kao i u slučaju gradijentne metode, uvodimo duljinu koraka $\alpha_k$ i tako definiramo Newtonovu metodu s regulacijom koraka

$$x_{k+1} = x_k - \alpha_k f''(x_k)^{-1} f'(x_k), \qquad k = 0, 1, \dots, \tag{4.3}$$

koju u implementaciji definiramo kao

$$\begin{aligned} &x_{k+1} = x_k + p_k, \quad \text{(gdje je } p_k \text{ rješenje sustava)} \\ &f''(x_k)p = f'(x_k). \end{aligned} \tag{4.4}$$

Kao i u slučaju gradijentne metode parametar duljine koraka $\alpha_k$ možemo definirati na dva načina. Prvi način definiran je sljedećim algoritmom:

Algoritam za izračunavanje duljine koraka

Korak 0. Staviti $\alpha = 1$ i izabrati $0 < \varepsilon < \frac{1}{2}$;

Korak 1. Izračunati $x = x_k + \alpha_k p_k$ i provjeriti uvjet

$$f(x) - f(x_k) \leq \varepsilon\alpha(f'(x_k), p_k); \tag{4.5}$$

Korak 2. Ako je uvjet (4.5) ispunjen, staviti $\alpha_k = \alpha$; inače staviti $\alpha := h\alpha, \quad h \in (0,1)$ i prijeći na *Korak 1.*

Drugi način izbora duljine koraka $\alpha_k$ dobivamo postupkom jednodimenzionalne minimizacije

$$\alpha_k := \arg\min_{\alpha \geq 0} f\left(x_k - \alpha p_k\right). \tag{4.6}$$

### 4.1.1. Svojstva Newtonove metode

Pretpostavit ćemo da je promatrana funkcija $f : \mathbb{R}^n \to \mathbb{R}^n$ dvostruko neprekidno diferencijabilna i jako konveksna, tj.

$$m\|y\|^2 \leq (f''(x)y, y) \leq M\|y\|^2, \quad M > m > 0 \qquad \text{za sve } x, y \in \mathbb{R}^n. \tag{4.7}$$

Primijetite da su takve funkcije ograničene odozdo i da posjeduju jedinstvenu točku minimuma.

Može se pokazati da također vrijedi (vidi primjerice [10], [16])

$$\frac{m}{M^2}\|y\|^2 \leq ((f''(x))^{-1}y, y) \leq \frac{1}{m}\|y\|^2, \quad M > m > 0 \qquad \text{za sve } x, y \in \mathbb{R}^n. \tag{4.8}$$

iz čega slijedi ocjena

$$(f_k', p_k) = -(f_k', (f''(x_k))^{-1} f_k') \leq -\frac{m}{M^2}\|f_k'\|^2. \tag{4.9}$$

**Teorem 4.1.** *Neka funkcija $f \in C^2(\mathbb{R}^n)$ zadovoljava uvjet (4.7) i neka se parametri duljine koraka $\alpha_k$ biraju iz uvjeta (4.5).*

*Tada, neovisno o izboru početne aproksimacije $x_0$, iterativni proces (4.3) konvergira prema jedinstvenoj točki minimuma $x^*$ superlinearnom brzinom*

$$\|x_{k+1} - x^*\| \leq \lambda_k \|x_k - x^*\|, \tag{4.10}$$

*gdje $\lambda_k \to 0$ za $k \to \infty$.*

Dokaz. Za $x = x_k + \alpha p_k$, gdje je $p_k = -f_k''^{-1} f_k'$, korištenjem (4.7) u Taylorovoj formuli dobivamo

$$f(x) - f(x_k) = \alpha\,(f_k', p_k) + \frac{\alpha^2}{2}\,(f''(\xi)p_k, p_k) \leq \alpha\,(f_k', p_k)\left(1 + \frac{\alpha M \|p_k\|^2}{2(f_k', p_k)}\right)$$

Kako je s druge strane $\left[(f_k', p_k) = (f_k')^T p_k = (f_k')^T \quad (f_k'')^{-1} f_k'' \quad p_k = -p_k^T f_k'' p_k =\right]$

$$(f_k', p_k) = -(f''(x_k)p_k, p_k) \leq -m\|p_k\|^2,$$

dobivamo $-(f_k', p_k) \geq m\|p_k\|^2 \quad \Rightarrow \quad \left(1 + \frac{\alpha M \|p_k\|^2}{2(f_k', p_k)}\right) = \left(1 - \frac{\alpha M \|p_k\|^2}{-2(f_k', p_k)}\right) \leq \left(1 - \frac{\alpha M}{2m}\right)$

$$f(x) - f(x_k) \leq \alpha\,(f_k', p_k)\left(1 - \frac{\alpha M}{2m}\right).$$

Izaberimo $\varepsilon > 0$ tako da bude

$$1 - \frac{\alpha M}{2m} \geq \varepsilon. \tag{4.11}$$

Množenjem ove nejednakosti s negativnim brojem $\alpha(f_k', p_k) < 0$ dobivamo

$$\varepsilon\alpha(f_k', p_k) \geq \alpha(f_k', p_k)\left(1 - \frac{\alpha M}{2m}\right) \geq f(x) - f(x_k)$$

Odavde slijedi da će uvjet (4.5) biti ispunjen ako je ispunjen uvjet (4.11), odnosno ako je $\alpha \leq \overline{\alpha} = \frac{2(1-\varepsilon)m}{M}$. Time je opravdan način izbora duljine koraka $\alpha_k$.

Kako je $(f_k', p_k) < 0$ za $\|f_k'\| \neq 0$ iz uvjeta

$$f_{k+1} - f_k \leq \varepsilon\alpha_k (f_k', p_k) \tag{4.12}$$

slijedi $f_{k+1} < f_k$, tj. niz $(f_k)$ je monotono padajuć. Zbog uvjeta jake konveksnosti funkcije $f$ on je i ograničen odozdo, što znači da je konvergentan i da vrijedi $f_{k+1} - f_k \to \infty$. Korištenjem uvjeta (4.8) u (4.12) i činjenice da je $f''(x_k)$ simetrična matrica dobivamo

$$f_{k+1} - f_k \leq \varepsilon\alpha_k (f_k', p_k) = -\varepsilon\alpha_k ((f_k'')^{-1} f_k', f_k') \leq -\varepsilon\alpha_k \frac{m}{M^2}\|f_k'\|^2, \tag{4.13}$$

a odavde

$$\|f_k'\|^2 \leq \frac{M^2}{m\varepsilon\alpha_k}\,(f_k - f_{k+1}),$$

iz čega slijedi $\|f_k'\| \to 0$. Time je pokazano da su rezultati analogni rezultatima *Leme* 3.1. ostali sačuvani.

Nadalje, slično kao u dokazu *Teorema* 3.1. pokazat ćemo da niz definiran s (4.3) konvergira prema točki minimuma $x^*$. Budući da je $f$ jako konveksna funkcija, postoji jedinstvena točka minimuma $x^* \in \mathbb{R}^n$. Pokažimo još da niz $(x_n)$ konvergira prema $x^*$ neovisno o izboru početne aproksimacije $x_0 \in \mathbb{R}^n$.

Primjenom Taylorove formule za funkciju $f$ u okolini točke $x$, za $x^*$ dobivamo

$$f(x^*) = f(x) + (f'(x), x^* - x) + \frac{1}{2}\,(f''(\xi)(x^* - x), x^* - x)$$

a odavde korištenjem uvjeta jake konveksnosti (4.7) dobivamo $\quad (f''(\xi)(x^* - x), (x^* - x)) \geq m\|x^* - x\|^2$

$$f(x) - f(x^*) = (f'(x), x - x^*) - \frac{1}{2}\,(f''(\xi)(x^* - x), x^* - x) \leq (f'(x), x - x^*) - \frac{m}{2}\|x - x^*\|^2,$$

što uz primjenu Cauchyjeve nejednakosti daje

$$f(x) - f(x^*) \leq \|f'(x)\|\|x - x^*\| - \frac{m}{2}\|x - x^*\|^2. \tag{4.14}$$

Primjenom Taylorove formule za funkciju $f$ u okolini točke $x^*$ i korištenjem činjenice da je $f'(x^*) = 0$, dobivamo

$$f(x) - f(x^*) = \frac{1}{2}\left(f''(\zeta)(x - x^*), x - x^*\right).$$

Koristeći ovdje uvjet jake konveksnosti (4.7) dobivamo

$$\frac{m}{2}\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{M}{2}\|x - x^*\|^2. \tag{4.15}$$

Iz (4.14) i (4.15) slijedi

$$\|f'(x)\|\|x - x^*\| - \frac{m}{2}\|x - x^*\|^2 \geq f(x) - f(x^*) \geq \frac{m}{2}\|x - x^*\|^2,$$

a odavde $\quad [m\|x - x^*\|^2 \leq \|f'(x)\|\|x - x^*\|]$

$$\|x - x^*\| \leq \frac{\|f'(x)\|}{m}. \tag{4.16}$$

Kao što smo ranije utvrdili desna strana ove nejednakosti za $x = x_k$ konvergira prema nuli, što znači da $\|x_k - x^*\| \to 0$ za $k \to \infty$, tj. $x_k \to x^*$.

Prijeđimo sada na dokaz ocjene brzine konvergencije (4.10). U tu svrhu najprije primijetimo da vrijedi

$$(f'_k, p_k) = -(f''(x_k)p_k, p_k) \leq -m\|p_k\|^2, \tag{4.17}$$

Iz (4.12) $\quad [f_{k+1} - f_k \leq \varepsilon\alpha_k(f'_k, p_k)] \quad$ slijedi $(f'_k, p_k) \to 0$. Zato iz (4.17) slijedi

$$\|p_k\| \to 0, \qquad k \to \infty.$$

Pokažimo nadalje da su u općoj Newtonovoj metodi (4.7), počevši od neke iteracije, svi parametri $\alpha_k$ jednaki 1. Koristeći Taylorovu formulu, te dodavanjem i oduzimanjem broja $\frac{\alpha_k^2}{2}(f''_k p_k, p_k)$ dobivamo

$$\begin{aligned}
f_{k+1} - f_k &= \alpha_k(f'_k, p_k) + \tfrac{\alpha_k^2}{2}(f''_k p_k, p_k) + \tfrac{\alpha_k^2}{2}\left((f''(\xi) - f''_k)p_k, p_k)\right) \\
\stackrel{(4.17)}{\leq}\; & \alpha_k(f'_k, p_k)\left(1 - \tfrac{\alpha_k}{2} + \tfrac{\alpha_k(-1)}{2}\tfrac{\|(f''(\xi) - f''_k)p_k\|\cdot\|p_k\|}{(-1)(f'_k, p_k)}\right) \\
\stackrel{(4.17)}{\leq}\; & \alpha_k(f'_k, p_k)\left(1 - \tfrac{\alpha_k}{2} - \tfrac{\alpha_k}{2}\tfrac{\|f''(\xi) - f''_k\|\cdot\|p_k\|^2}{m\|p_k\|^2}\right),
\end{aligned}$$

gdje je $\xi = x_k + \vartheta(x_{k+1} - x_k), \quad \vartheta \in [0, 1]$.

Budući da

$$\|f''(\xi) - f''(x_k)\| \leq \|f''(\xi) - f''(x^*)\| + \|f''(x^*) - f''(x_k)\|,$$

i $\|x_k - x^*\| \to 0$ za $k \to \infty$, onda zbog neprekidnosti funkcije $f''$ vrijedi $\|f''(\xi) - f''(x_k)\| \to 0$ za $k \to \infty$. Zbog toga za proizvoljni $0 < \varepsilon < \frac{1}{2}$ postoji prirodan broj $N_0(\varepsilon) \in \mathbb{N}$ takav da za svaki $k \geq N_0(\varepsilon)$ vrijedi

$\alpha_k \geq 1 \Rightarrow -\frac{\alpha_k}{2} \leq -\frac{1}{2} \Rightarrow 1 - \frac{\alpha_k}{2} \leq \frac{1}{2}$

$$1 - \frac{1}{2} \geq \qquad 1 - \frac{\alpha_k}{2} - \frac{\alpha_k}{2}\frac{\|f''(\xi) - f''_k\|}{m} \geq \varepsilon,$$

uz uvjet da je $\alpha_k \geq 1$. Budući da je $\alpha_k$ u uvjetu (4.5) biran tako da bude $\alpha_k \leq 1$, to znači da će u svim iteracijama, čiji je redni broj veći ili jednak $N_0(\varepsilon)$, biti $\alpha_k = 1$.

Dakle, za $k \geq N_0(\varepsilon)$ vrijedi $\quad x_{k+1} = x_k - (f''_k)^{-1}f'_k$

$$(x_{k+1} - x^*, x_{k+1} - x^*) = \left(x_k - x^* - (f''_k)^{-1}f'_k, x_{k+1} - x^*\right).$$

Ako na desnoj strani ove jednakosti iskoristimo Lagrangeovu formulu

$$\begin{aligned}\left((f_k'')^{-1}f_k', x_{k+1}-x^*\right) \quad [f'(x^*)=0] \quad &= \left((f_k'')^{-1}(f_k'-f'(x^*)), x_{k+1}-x^*\right)\\ &= \left((f_k'')^{-1}f''(\xi)(x_k-x^*), x_{k+1}-x^*\right)\end{aligned}$$

gdje je $\xi = x^* + \vartheta(x_k - x^*), \quad \vartheta \in (0,1)$, dobivamo

$$\begin{aligned}\|x_{k+1}-x^*\|^2 &= \left(\left(I-(f_k'')^{-1}f''(\xi)\right)(x_k-x^*), x_{k+1}-x^*\right)\\ &= \left((f_k'')^{-1}\left(f_k''-f''(\xi)\right)(x_k-x^*), x_{k+1}-x^*\right)\\ (4.8) \quad &\le \tfrac{1}{m}\|f_k''-f''(\xi)\|\cdot\|x_k-x^*\|\cdot\|x_{k+1}-x^*\|,\end{aligned}$$

a odavde

$$\|x_{k+1}-x^*\| \le \lambda_k\|x_k-x^*\|, \qquad \lambda_k = \frac{1}{m}\|f_k''-f''(\xi)\|. \tag{4.18}$$

Budući da $\|f_k''-f''(\xi)\| \to 0$ za $k\to\infty$, teorem je dokazan. ♣

**Korolar 4.1.** *Neka funkcija $f \in C^2(\mathbb{R}^n)$ zadovoljava uvjet (4.7) i Lipschitzov uvjet*

$$\|f''(x)-f''(y)\| \le R\|x-y\| \quad \forall x,y\in\mathbb{R}^n.$$

*Tada niz (4.3)*

$$x_{k+1} = x_k - \alpha_k f''(x_k)^{-1}f'(x_k), \qquad k=0,1,\dots,$$

*gdje se parametri $\alpha_k$ biraju iz uvjeta (4.5), konvergira prema točki minimuma $x^*$ neovisno o izboru početne aproksimacije $x_0$ kvadratičnom brzinom konvergencije*

$$\|x_{k+1}-x^*\| \le \frac{R}{m}\|x_k-x^*\|^2.$$

Dokaz. Koristeći Lipschitzov uvjet broj $\lambda_k$ iz nejednakost (4.18) možemo pisati

$[\xi = x^* + \vartheta(x_k - x^*), \quad 0<\vartheta<1]$

$$\begin{aligned}\lambda_k = \tfrac{1}{m}\|f''(x_k)-f''(\xi)\| &\le \tfrac{R}{m}\|x_k-x^*-\vartheta(x_k-x^*)\|\\ &= \tfrac{R}{m}\|x_k(1-\vartheta)-x^*(1-\vartheta)\|\\ &< \tfrac{R}{m}\|x_k-x^*\|,\end{aligned}$$

što daje traženu ocjenu. ♣

Ako duljinu koraka $\alpha_k$ biramo iz uvjeta (4.6)

$$\alpha_k := \arg\min_{\alpha\ge 0} f\left(x_k-\alpha p_k\right),$$

tada za Newtonovu metodu (4.3)

$$x_{k+1} = x_k - \alpha_k f''(x_k)^{-1}f'(x_k), \qquad k=0,1,\dots,$$

vrijedi

**Korolar 4.2.** *Neka funkcija $f\in C^2(\mathbb{R}^n)$ zadovoljava uvjet (4.7) i neka se duljina koraka $\alpha_k$ bira iz uvjeta (4.6).*

*Tada Newtonov iterativni proces (4.3) konvergira prema rješenju $x^*$ bez obzira na izbor početne aproksimacije $x_0$. superlinearnom brzinom. Ako pored toga funkcija $f$ zadovoljava i Lipschitzov uvjet, onda je brzina kvadratična.*

Dokaz. Neka je

$$\overline{x}_{k+1} = x_k - (f''(x_k))^{-1} f'(x_k)$$
$$x_{k+1} = x_k - \alpha_k (f''(x_k))^{-1} f'(x_k),$$

gdje je $\alpha_k$ izabran iz uvjeta (4.6). Budući da je $f(x_{k+1}) \le f(\overline{x}_{k+1})$, koristeći nejednakost (4.15), dobivamo $\left[\frac{m}{2}\|x - x^*\|^2 \le f(x) - f(x^*) \le \frac{M}{2}\|x - x^*\|^2\right]$

$$\frac{m}{2}\|x_{k+1} - x^*\|^2 \le f(x_{k+1}) - f(x^*) \le f(\overline{x}_{k+1}) - f(x^*) \le \frac{M}{2}\|\overline{x}_{k+1} - x^*\|^2. \tag{4.19}$$

Za $\overline{x}_{k+1} = x_k - (f''(x_k))^{-1} f'(x_k)$ možemo primijeniti nejednakost (4.18)

$$\|\overline{x}_{k+1} - x^*\| \le \lambda_k \|x_k - x^*\|, \qquad \lambda_k = \frac{1}{m}\|f''_k - f''(\xi)\| = \frac{1}{m}\|f''_k - f''(x^* + \vartheta(x_k - x^*))\|$$

gdje $\lambda_k \to 0$ za $k \to \infty$. Ako ovu nejednakost iskoristimo u (4.19), dobivamo

$$\begin{aligned} \|x_{k+1} - x^*\| &\le \sqrt{\tfrac{M}{m}}\|\overline{x}_{k+1} - x^*\| \le \lambda_k \sqrt{\tfrac{M}{m}}\|x_k - x^*\| \\ &\le \gamma_k \|x_k - x^*\|, \end{aligned}$$

gdje $\gamma_k = \lambda_k \sqrt{\frac{M}{m}} \to 0$ za $k \to \infty$.

Ako još vrijedi i Lipschitzov uvjet, onda je $\lambda_k \le \frac{R}{m}\|x_k - x^*\|$, pa vrijedi

$$\|x_{k+1} - x^*\| \le C\|x_k - x^*\|^2, \qquad C = \frac{R}{m}\sqrt{\frac{M}{m}}.$$

♣

## 4.2. Quasi–Newtonove metode

Budući da problem minimizacije diferencijabilne funkcije $f : \mathbb{R}^n \to \mathbb{R}^n$ možemo gledati kao problem rješavanja jednadžbe

$$F(x) = 0, \quad F : \mathbb{R}^n \to \mathbb{R}^n$$

gdje je $F(x) = f'(x)$, najprije ćemo razmotriti problem rješavanja sustava nelinearnih jednadžbi.

Općenito se sustav nelinearnih jednadžbi, kao i problem minimizacije funkcije više varijabli rješava iterativnim metodama. Smatra se da je klasu novih "quasi-Newtonovih" metoda za rješavanje problema minimizacije funkcije više varijabli bez ograničenja uveo Davidon (1959), a za rješavanje sustava nelinearnih jednadžbi Broyden (1965). U literaturi su za klasu ovih metoda korišteni razni nazivi: *quasi-Newton, variable metric, variance, secant, update, modification methods.*

### 4.2.1. Sustavi nelinearnih jednadžbi

Za funkciju $F : \mathbb{R}^n \to \mathbb{R}^n$ promatramo problem rješavanja jednadžbe $F(x) = 0$, odnosno sustava nelinearnih jednadžbi

$$f_i(x_1, \ldots, x_n) = 0, \qquad i = 1, \ldots, n,$$

gdje su funkcije $f_i$ komponente vektorske funkcije $F(x) = (f_1(x), \ldots, f_n(x))^T$, $x = (x_1, \ldots, x_n)^T$. Pretpostavimo da funkcija $F$ ima sljedeća svojstava

(4.20)

- $(i)$ $F$ je neprekidno diferencijabilna na otvorenom konveksnom skupu $\mathcal{D}$, tj. $F \in C^1(\mathcal{D})$ ;
- $(ii)$ postoji $x^* \in \mathcal{D}$ tako da bude $F(x^*) = 0$, i da je pri tome $F'(x^*)$ nesingularna matrica ;

Dodatno, uz uvjete (4.20) možemo zahtijevati i ispunjenje jačeg *Lipschitzovog uvjeta*

$$(iii) \quad \|F'(x) - F'(X^*)\| \leq R\|x - x^*\|, \qquad x \in \mathcal{D}. \tag{4.21}$$

Primijetite da ako je $\mathcal{D}$ dovoljno malen, onda je ovaj uvjet ispunjen za dvostruko neprekidno diferencijabolnu funkciju u $x^*$.

Uz pretpostavku da imamo $k$-tu aproksimaciju $x_k$ rješenja $x^*$, sljedeću $(k+1)$-u aproksimaciju kod obične Newtonove metode dobit ćemo tako da funkciju $F$ u okolini točke $x_k$ pomoću Taylorove formule aproksimiramo linearnim aproksimandom

$$L(x) = F(x_k) + F'(x_k)(x - x_k)$$

i riješimo linearni sustav $L(x) = 0$. Tako dobivamo Newtonov iterativni postupak za rješavanje jednadžbe $F(x) = 0$

$$x_{k+1} = x_k - F'(x_k)^{-1}F(x_k).$$

Za praktičnu implementaciju obično koristimo sljedeći algoritam

Korak 1 za poznati $x_k$ izračunati $F(x_k)$; ako je $x_k$ prihvatljiv, stop. U protivnom izračunati $F'(x_k)$ i prijeći na Korak 2

Korak 2 riješiti sustav $F'(x_k)p = -F(x_k)$ za $p$, staviti $p_k := p$ i izračunati $x_{k+1} = x_k + p_k$.

Newtonova metoda poboljšava se uvođenjem parametra duljine koraka $0 < \alpha_k \leq 1$

$$x_{k+1} = x_k - \alpha_k F'(x_k)^{-1}F(x_k).$$

Karakteristike Newtonove metode dane su sljedećim teoremom (dokaz vidi u [15], str. 312)

**Teorem 4.2.** *Neka funkcija $F : \mathbb{R}^n \to \mathbb{R}^n$ zadovoljava pretpostavke (4.20). Tada postoji otvoren skup $\mathcal{S}$ koji sadržava $x^*$, takav da je za proizvoljni $x_0 \in \mathcal{S}$ niz $(x_n)$ definiran Newtonovim iterativnim postupakom dobro definiran, ostaje u $\mathcal{S}$ i konvergira prema $x^*$ superlinearnom brzinom, tj. postoji nul-niz $(\lambda_n)$, takav da je*

$$\|x_{k+1} - x^*\| \leq \lambda_k \|x_k - x^*\|, \qquad k = 0, 1, \ldots .$$

*Ako funkcija $F$ dodatno zadovoljava Lipschitzov uvjet, onda je brzina konvergencije kvadratična, tj. postoji konstanta $\beta$ takva da je*

$$\|x_{k+1} - x^*\| \leq \beta \|x_k - x^*\|, \qquad k = 0, 1, \ldots .$$

**Primjedba 4.1.** *Može se pokazati (vidi [7]) da za niz $(x_n)$, koji superlinearnom brzinom konvergira prema $x^*$, vrijedi*

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} = 1$$

*uz uvjet da je $x_k \neq x^*$ za sve $k = 0, 1, \ldots$.*

Kako je [ $||\alpha| - |\beta|| \leq |\alpha - \beta|$ ]

$$|\|x_{k+1} - x_k\| - \|x_k - x^*\|| \leq \|x_{k+1} - x^*\|,$$

nakon dijeljenja s $\|x_k - x^*\|$ i koristeći $\|x_{k+1} - x^*\| \leq \lambda_k \|x_k - x^*\|, \quad \lambda_k \to 0$, dobivamo

$$\left| \frac{\|x_{k+1} - x_k\|}{\|x_k - x^*\|} - 1 \right| \leq \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \lambda_k \to 0$$

Na osnovi *Primjedbe* 4.1. pored kriterija $\|F(x_k)\| < \varepsilon_1$ za zaustavljanje procesa ima smisla koristiti dodatni kriterij $\|x_{k+1} - x_k\| < \varepsilon_2 \|x_k\|$

Prethodnim teoremom iskazane su dvije osnovne prednosti Newtonove metode:

- postoji područje prihvaćanja $\mathcal{S}$ Nevtonove iterativne metode,
- visoka brzina konvergencije u području prihvaćanja $\mathcal{S}$.

Nedostaci Newtonove metode su

- potreba izbora dobre početne aproksimacije $x_0$,
- u svakom koraku treba računati matricu parcijalnih derivacija reda $n$.

U cilju prevladavanja naročito posljednjeg navedenog nedostatka, matricu $F'(x_k)$ (odnosno matricu $F'(x_k)^{-1}$) zamijenit ćemo nekom matricom $B_k$ (odnosno matricom $H_k$), koja se lako računa, a nalikuje na matricu $F'(x_k)$ (odnosno matricu $F'(x_k)^{-1}$). Tako promatramo iterativne procese

$$\begin{aligned} x_{k+1} &:= x_k - \alpha_k B_k^{-1} F(x_k), \\ \text{odnosno} \quad x_{k+1} &:= x_k - \alpha_k H_k F(x_k). \end{aligned} \tag{4.22}$$

Vrijedi sljedeći teorem (vidi [7])

**Teorem 4.3.** *Neka funkcija $F : \mathbb{R}^n \to \mathbb{R}^n$ zadovoljava uvjete (4.20), a niz $(B_k)$ neka je niz regularnih matrica. Pretpostavimo nadalje da za neki $x_0 \in \mathcal{D}$ niz $(x_n)$ definiran s (4.22) ostaje u $\mathcal{D}$, $x_k \neq x^*$ za sve $k = 0, 1, \ldots$ i konvergira prema $x^*$. Tada niz $(x_n)$ konvergira superlinearno prema $x^*$ onda i samo onda ako*

$$\lim_{k\to\infty} \frac{(B_k - F'(x^*))\,(\|p_k\|)}{\|p_k\|} = 0, \qquad p_k = x_{k+1} - x_k. \tag{4.23}$$

Uvjet (4.23) zahtijeva da niz matrica $(B_n)$ konvergira prema $F'(x^*)$ u smjeru vektora kretanja $p_k = x_{k+1} - x_k$.

**Primjedba 4.2.** *Uvjet (4.23) ekvivalentan je uvjetu (vidi primjerice [10])*

$$\lim_{k\to\infty} \frac{\|B_k p_k - y_k\|}{\|p_k\|} = 0, \qquad y_k := F(x_{k+1} - F(x_k), \tag{4.24}$$

*što znači da uvjet $B_k p_k \approx y_k$ također garantira superlinearnu konvergenciju.*

Ne može se očekivati da bude $B_k p_k = y_k$, ali zato ćemo matricu $B_{k+1}$ brati tako da bude ispunjen tzv. quasi-Newtonov uvjet

$$B_{k+1} p_k = y_k. \tag{4.25}$$

i da je za poznate $p_k, y_k, B_k$ lako izračunati $B_{k+1}$. Iterativna metoda (4.22) pri čemu niz matrica $(B_n)$ ispunjavaju quasi-Newtonov uvjet naziva se quasi-Newtonova metoda. Radi pojednostavljivanja uvedimo sljedeće oznake:

$$\begin{aligned} &B := B_k, \qquad \overline{B} := B_{k+1}, \qquad x := x_k, \qquad \overline{x} := x_{k+1} \\ &p := p_k = \overline{x} - x, \qquad y := y_k = F(\overline{x}) - F(x). \end{aligned}$$

Uz ove oznake quasi-Newtonov uvjet glasi

$$\overline{B}\, p = y.$$

Broyden je 1965. godine prvi puta u radu [3] uveo takav niz matrica, u literaturi poznat pod nazivom Broydenova korekcija ranga 1

$$\overline{B} = B + \frac{(y - Bp)p^T}{p^T p}. \tag{4.26}$$

**Zadatak 4.1.** *Provjerite da vrijedi*

$$\overline{B}p = y \text{ i } Bz = \overline{B}z \quad \text{za sve } z \text{ takve da je } p^T z = 0.$$

Sljedeći teorem pokazuje da je matrica $\overline{B}$ definirana s (4.26) jedinstvena matrica koja ispunjava quasi-Newtonov uvjet, a najmanje se razlikuje od matrice $B$.

**Teorem 4.4.** *Za danu matricu $B \in \mathbb{R}^{n\times n}$, $y \in \mathbb{R}^n$ i neki nenul vektor $p \in \mathbb{R}^n$ matrica $\overline{B}$ je jedinstveno rješenje problema minimizacije*

$$\min_{\hat{B}} \left\{ \|\hat{B} - B\|_F : \hat{B}p = y \right\},$$

*gdje je $\|\cdot\|_F$ Frobenijusova norma.*[1]

Dokaz. Naprije primijetimo da za proizvoljnu matricu $\hat{B}$, za koju vrijedi quasi-Newtonov uvjet $\hat{B}p = y$ vrijedi

$$\begin{aligned} \|\overline{B} - B\|_F &= \left\| \tfrac{(y-Bp)p^T}{p^T p} \right\|_F = \left\| \tfrac{(\hat{B}-B)pp^T}{p^T p} \right\|_F = \\ &\le \|\hat{B} - B\|_F \left\| \tfrac{pp^T}{p^T p} \right\|_F = \qquad \left[ \|pp^T\|_F = \sum_{ij} (p_i p_j)^2 = \left(\sum_i p_i^2\right)\left(\sum_j p_j^2\right) = (p^T p)^2 \right] \\ &= \|\hat{B} - B\|_F \end{aligned}$$

Kako je skup svih $\hat{B} \in \mathbb{R}^{n\times n}$, takvih da je $\hat{B}p = y$ konveksan, a funkcija $\Phi : \mathbb{R}^{n\times n} \mapsto \mathbb{R}$, $\Phi(A) = \|B - A\|_F$ strogo konveksna, matrica $\overline{B}$ na kojoj se postiže minimum je jedinstvena. ♣

Sljedeći teorem pokazuje superlinearnu konvergenciju Broydenove metode. Dokaz se može vidjeti u [7].

**Teorem 4.5.** *Uz uvjete (4.20), (4.21) postoji $\varepsilon > 0$, takav da za proizvoljne $x_0$, odnosno $B_0$, koji zadovoljavaju $\|x_0 - x^*\| \le \varepsilon$, odnosno $\|B_0 - F'(x^*)\| \le \varepsilon$, iterativna metoda (4.22) uz $\alpha_k \equiv 1$ i korekciju (4.26) definira niz $(x_n)$, koji superlinearnom brzinom konvergira prema $x^*$.*

[1] Može se pokazati (vidi primjerice [7]) da za proizvoljnu kvadratnu matricu $A$ reda $n$ vrijedi

$$\|A\|_F^2 = Tr(A^T A) = \sum_{i=1}^{n} \|Av_i\|_2^2,$$

za proizvoljni ortonormiran skup $v_1, \ldots, v_n$ u $R^n$.

Broydenova quasi-Newtonova metoda definirana je dakle, na sljedeći način

$$x_{k+1} = x_k - B_k^{-1} F(x_k), \qquad k = 0, 1, \dots , \tag{4.27}$$

gdje se matrice $B_k$ generiraju rekurzivnom formulom

$$B_{k+1} = B_k + \frac{(y_k - B_k p_k) p_k^T}{(p_k, p_k)}, \qquad k = 0, 1, \dots , \tag{4.28}$$

gdje je

$$y_k = F(x_{k+1}) - F(x_k), \qquad p_k = x_{k+1} - x_k. \tag{4.29}$$

**Primjedba 4.3.** *Navedena Broydenova metoda (4.27-4.29) na svakom koraku uključuje riješavanje sustav linearnih jednadžbi $B_k p_k = -F(x_k)$. To se može izbjeći tako da definiramo tzv.* inverznu Broydenovu metodu

$$\begin{aligned} x_{k+1} &= x_k - H_k F(x_k), \\ H_{k+1} &= H_k + \tfrac{(p_k - H_k y_k) p_k^T H_k}{(p_k, H_k y_k)}, \qquad (p_k, H_k y_k) \neq 0 \end{aligned} \tag{4.30}$$

Pri tome koristi se tzv. Sherman-Morrisonova lema

**Lema 4.1.** *Neka je $u, v \in \mathbb{R}^n$, a $A \in \mathbb{R}^{n\times n}$ regularna matrica. Tada je $A + uv^T$ regularna onda i samo onda ako je $\sigma = 1 + (v, A^{-1}u) \neq 0$. Ako je $\sigma \neq 0$, vrijedi*

$$\left(A + uv^T\right)^{-1} = A^{-1} - \frac{1}{\sigma} A^{-1} uv^T A^{-1} \tag{4.31}$$

Dokaz ove leme provodi se korištenjem niže navedene *Leme* 4.2. (vidi primjerice [7]) i dirktnom provjerom.

**Lema 4.2.** *Za zadane $v, w \in \mathbb{R}^n$ vrijedi*

$$\det(I + vw^T) = 1 + (v, w). \tag{4.32}$$

Dokaz. Označimo $P = I + vw^T$. Pretpostavimo da je $v \neq 0$; u protivnom dokaz je trivijalan.

Neka je $\lambda$ svojstvena vrijednost, a $x$ odgovarajući svojstveni vektor od $P$. Tada mora biti

$$Px = \lambda x \Rightarrow (I + vw^T)x = \lambda x \Rightarrow x + v(w^T x) = \lambda x.$$

Ova jednakost može biti ispunjena ili tako da je $w \perp x$ ili da je $v \parallel x$.

Dakle, svaki svojstveni vektor od $P$ je ili okomit na $w$ ili paralelan s $v$. 1. Ako je $w \perp x$, onda je $\lambda = 1$.

2. Ako je $v \parallel x$, tj. $x = \alpha v$, tada je

$$\alpha v + vw^T \alpha v = \lambda \alpha v \Rightarrow v(1 + w^T v) = \lambda v \Rightarrow \lambda = 1 + (v, w).$$

Budući da može postojati samo jedan svojstveni vektor kolinearan s $v$, to znači da matrica $P$ ima jednu svojstvenu vrijednost $1 + (v, w)$, dok su preostalih $(n - 1)$ jenake 1. Time je lema dokazana. ♣

Sada možemo računati

$$B_{k+1}^{-1} = \left(B_k + \tfrac{y_k - B_k p_k}{(p_k, p_k)} p_k^T\right)^{-1} = B_k^{-1} - \tfrac{1}{\sigma} B_k^{-1} \tfrac{y_k - B_k p_k}{(p_k, p_k)} p_k^T B_k^{-1}$$

$$\left[\sigma(p_k, p_k) = \left(1 + p_k^T B_k^{-1} \tfrac{y_k - B_k p_k}{(p_k, p_k)}\right)(p_k, p_k) = (p_k, p_k) + p_k^T B_k^{-1} y_k - p_k^T p_k = (p_k, B_k^{-1} y_k)\right]$$

$$= B_k^{-1} + \tfrac{p_k p_k^T B_k^{-1} - B_k^{-1} y_k p_k^T B_k^{-1}}{(p_k, B_k^{-1} y_k)} =$$

$$B_k^{-1} + \tfrac{(p_k - B_k^{-1} y_k) p_k^T B_k^{-1}}{(p_k, B_k^{-1} y_k)}$$

Odavde uz oznake: $H_k := B_k^{-1}$ i $H_{k+1} := B_{k+1}^{-1}$ dobivamo (4.30).

**Zadatak 4.2.** *Na osnovi* Leme *4.2. dokažite Sherman–Morrisonovu lemu (*Lema *4.1. )*

### 4.2.2. Minimizacija glatke funkcije

Promatramo sada problem minimizacije bez ograničenja dovoljno glatke funkcije $f : \mathbb{R}^n \to \mathbb{R}$. Pretpostavimo da je $x^*$ strogi lokalni minimum funkcije $f$. Tada je prema *Teoremu* 1.3. $f'(x^*) = 0$, a $f''(x^*)$ je pozitivno definitna Hesseova matrica. Uz oznaku $F(x) := f'(x)$, te slijedeći uvjete (4.20) i (4.20) definirat ćemo uvjete na funkciju $f : \mathbb{R}^n \to \mathbb{R}$:

(*i*) $f$ je dvostruko neprekidno diferencijabilna na otvorenom konveksnom skupu $\mathcal{D}$, tj. $f \in C^2(\mathcal{D})$ ;

(*ii*) postoji $x^* \in \mathcal{D}$ tako da bude $f'(x^*) = 0$, a $f''(x^*)$ je pozitivno definitna matrica ;

(*iii*) $f''$ je u $x^*$ zadovoljava Lipschitzov uvjet
$\|f''(x) - f''(x^*)\| \leq R\|x - x^*\|$ za sve $x \in \mathcal{D}$

Uz oznaku $g(x) := f'(x)$ i $g_k = g(x_k)$ iterativna procedura (4.22) postaje

$$
\begin{aligned}
& x_{k+1} := x_k - \alpha_k B_k^{-1} g_k, \\
\text{odnosno} \quad & x_{k+1} := x_k - \alpha_k H_k g_k,
\end{aligned}
$$

gdje je sada $B_k$ aproksimacija hessijana $f''(x_k)$, a $H_k = B_k^{-1}$, pri čemu duljinu koraka $\alpha_k$ dobivamo jednodimenzionalnom minimizacijom

$$f(x_{k+1}) \approx \min_{\alpha \geq 0} f(x_k + \alpha p_k), \qquad p_k := -B_k^{-1} g_k = -H_k\, g_k.$$

Budući da matrice $B_k$ trebaju aproksimirati hessijan $f''(x_k)$, za njih se mora pretpostaviti simetričnost. Slično kao i ranije uvedimo oznake

$$
\begin{aligned}
& B := B_k, \qquad \overline{B} := B_{k+1} \\
& p := p_k = x_{k+1} - x_k \qquad y := y_k = g_{k+1} - g_k \quad (\text{ zbog } g(x) = F(x))
\end{aligned}
$$

Jedna mogućnost za definiranje korekcije ranga 1 je simetrična formula ranga 1 koju je uveo Davidon (1959):

$$\overline{B} = B + \frac{(y - Bp)(y - Bp)^T}{(y - Bp, p)}, \qquad (y - Bp, p) \neq 0 \tag{4.33}$$

**Zadatak 4.3.** *Pokažite da ako postoje matrice:* $H := B^{-1}$ *i* $\overline{H} := overlineB^{-1}$ *i ako je* $B$ *simetrična, tada vrijedi*

$$\overline{H} = H + \frac{(p - Hy)(p - Hy)^T}{(p - Hy, y)} \tag{4.34}$$

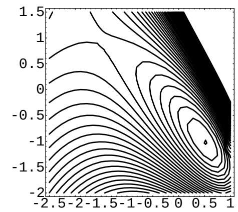
**Primjer 4.1.** *Potražimo lokalni minimum funkcije*

$$f(x_1, x_2) = e^{x_1}\left(4x_1^2 + 2x_2^2 + 4x_1 x_2 + 2x_2 + 1\right)$$

Gradijent, hessijan i `ContourPlot` dobit ćemo na sljedeći način

```
In[1]:= f[x_] := Exp[x[1]](4x[1]^2 + 2x[2]^2 + 4x[1]x[2] + 2x[2] + 1); n = 2;
        g[x_] := Table[D[f[x], x[i]], {i, n}]
        H[x_] := Table[D[g[x][[i]], x[j]], {i, n}, {j, n}]
             Print["f'=", Simplify[g[x] /. {x[1] -> x1, x[2] -> x2}] // MatrixForm,
                    "   H=", Simplify[H[x] /. {x[1] -> x1, x[2] -> x2}] // MatrixForm]
               ContourPlot[f[x] /. {x[1] -> x1, x[2] -> x2}, {x1, -2.5, 1}, {x2, -2, 1.5},
                             Contours -> 40, PlotPoints -> 30, ContourShading -> None];
```

$$f' = \begin{bmatrix} e^{x_1}(1+4x_1^2+6x_2+2x_2^2+4x_1(2+x_2)) \\ e^{x_1}(2+4x_1+4x_2) \end{bmatrix}, \qquad H = \begin{bmatrix} e^{x_1}(9+4x_1^2+10x_2+2x_2^2+4x_1(4+x_2)) & e^{x_1}(3+2x_1+2x_2) \\ e^{x_1}(3+2x_1+2x_2) & e^{x_1} \end{bmatrix}$$




Zadavanjem ulaznih podataka

```
In[2]:= n = 2; x0 = {1., 1}; k = 0; eps = 0.005;
          w = Table[x[i] -> x0[[i]], {i, n}]; B0 = IdentityMatrix[n];
            B0 = H[x] /. w
              g0 = g[x] /. w
```

i korištenjem programa (modul `Korak` definiran je ranije kod *Metode najbržeg silaska*)

```
In[3]:= While[
              p0 = -LinearSolve[B0, g0];
                alpha = Korak[n, f, p0, x0];
                  x1 = x0 + alpha p0;
                    w1 = Table[x[i] -> x1[[i]], {i, n}]; g1 = g[x] /. w1;
                      y0 = g1 - g0;
           B1 = B0 + Outer[Times, (y0 - B0.p0), (y0 - B0.p0)]/((y0 - B0.p0) . p0);
               f1 = f[x] /. w1;
                   Abs[f1] > eps,
                             B0 = B1; g0 = g1;
                                x0 = x1; k = k + 1;
          Print["x", k, "=", x1, "  fx=", f[x] /. w1, "  grad=", g1, "  alpha=", alpha]
               ];
        Print["x", k + 1, "=", x1, "  fx=",f[x] /. w1, "  grad=", g1, "  alpha=", alpha]
```

nakon 18 itracija dobivamo

```
{x_18 = {-11.43, 19.80},   fx=0.00479,   grad= {0.0047, 0.00038},  alpha=1}
```

**Zadatak 4.4.** *Slično kao u* Primjedbi *4.3. definirajte inverznu simetričnu formulu ranga* 1.

Slično *Teoremu* 4.4., vrijedi

**Teorem 4.6.** *Neka je $M \in \mathbb{R}^{n\times n}$ simetrična regularna matrica, $y, p \in \mathbb{R}^n \quad (p \neq 0)$ i $c := M^{-2}p$. Ako je $B$ simetrična matrica, tada je rješenje problema minimizacije*

$$\min_{\hat{B}} \left\{ \|M(\hat{B}-B)M\|_F : \hat{B} = \hat{B}^T, \quad \hat{B}p = y \right\}, \tag{4.35}$$

*zadano s*

$$\overline{B} = B + \frac{(y-Bp)c^T + c(y-Bp)^T}{c^Tp} - \frac{(y-Bp)^Tp}{(c^Tp)^2}cc^T. \tag{4.36}$$

Matricom $M$ definirana je težinska Frobeniusova norma, $\overline{B}$ je također simetrična matrica, koja ispunjava quasi-Newtonov uvjet $\overline{B}p = y$ i po težinskoj Frobeniusovoj normi najmanje se razlikuje od matrice $B$. Formulu (4.36) zovemo korekcija ranga 2 jer se $\overline{B}$ dobiva tako da matrici $B$ dodamo jednu matricu ranga 2

**Zadatak 4.5.** *Dva različita dokaza ovog teorema mogu se vidjeti u [10] i [7]. Izvedite i usporedite oba dokaza.*

Pokazat ćemo (vidi [7]) kako se izvodi quasi-Newtonova korekcija ranga 2 zadana s (4.36) jer je to način kako se mogu izvesti i druge poznate korekcije ranga 2. Polazimo od simetrične matrice $B \in \mathbb{R}^{n\times n}$, te slično Broydenovoj korekciji ranga 1 (4.26) kao mogućeg kandidata za $\overline{B}$ promatramo

$$C_1 = B + \frac{(y - Bp)c^T}{(c,p)}.$$

koji ispunjava quasi-Newtonov uvjet $C_1 p = y$. Budući da $C_1$ općenito nije simetrična matrica, definiramo simetričnu matricu

$$C_2 = \frac{1}{2}\left(C_1 + C_1^T\right).$$

Budući da $C_2$ ne ispunjava quasi-Newtonov uvjet, ponovimo proceduru i tako rekurzivno definiramo niz $(C_n)$

$$\begin{aligned} C_{2k+1} &= C_{2k} + \tfrac{(y - C_{2k}p)c^T}{(c,p)}, \\ C_{2k} &= \tfrac{1}{2}\left(C_{2k+1} + C_{2k+1}^T\right), \end{aligned} \tag{4.37}$$

gdje je $C_0 = B$. Vrijedi

**Lema 4.3.** *Neka je $B \in \mathbb{R}^{n\times n}$ simetrična matrica i $c, p, y \in \mathbb{R}^n$ sa svojstvom $(c,p) \neq 0.$. Tada niz $(C_n)$ zadan rekurzivnom formulom (4.37) uz $C_0 = B$ konvergira prema $\overline{B}$ zadanom s (4.36).*

Dokaz. Dokažimo da niz $(C_{2k})$ konvergira. Uz oznaku $G_k = C_{2k}$ iz (4.37) dobivamo

$$\begin{aligned} G_{k+1} = C_{2(k+1)} &= \tfrac{1}{2}\left(C_{2k+1} + C_{2k+1}^T\right) = \tfrac{1}{2}\left(G_k + \tfrac{(y - G_k p)c^T}{(c,p)} + G_k^T + \tfrac{c(y - G_k p)^T}{(c,p)}\right) \\ &= G_k + \tfrac{w_k c^T + c w_k^T}{2(c,p)}, \end{aligned}$$

gdje je $w_k = y - G_k p$. Koristeći dobivenu jednakost dobivamo

$$\begin{aligned} w_{k+1} = y - G_{k+1}p &= y - G_k p - \tfrac{1}{2}\tfrac{w_k c^T + c w_k^T}{(c,p)}p = w_k - \tfrac{1}{2}\tfrac{w_k c^T p}{(c,p)} - \tfrac{1}{2}\tfrac{c w_k^T p}{(c,p)} \\ &= \tfrac{1}{2} w_k - \tfrac{1}{2}\tfrac{c(p^T w_k)}{(c,p)}, \end{aligned}$$

što možemo pisati kao

$$w_{k+1} = P w_k, \qquad P = \frac{1}{2}\left[I - \frac{cp^T}{(c,p)} w_k\right]$$

Prema Lemi 4.2. matrica $P$ ima jednu svojstvenu vrijednost 0 $\left[\frac{1}{2}\left(1 - (\frac{c}{(c,p)}, p)\right) = 0\right]$ i svojstvenu vrijednost $\frac{1}{2}$ kratnosti $(n-1)$. Dakle, spektralni radijus matrice $P$ je $\rho(P) = \frac{1}{2} < 1$. Prema *Neumannovoj lemi* (vidi primjerice [15], str. 45) tada postoji matrica $(I - P)^{-1}$ i vrijedi

$$(I - P)^{-1} = \lim_{k\to\infty} \sum_{i=0}^{k} P^k.$$

Kako je s druge strane

$$w_k = P w_{k-1} = \cdots = P^k w_0 = P^k (y - Bp),$$

onda imamo

$$\sum_{k=0}^{\infty} w_k = \sum_{k=0}^{\infty} P^k(y - Bp) = (I - P)^{-1}(y - Bp).$$

Također vrijedi

$$\sum_{k=0}^{n}(G_{k+1} - G_k) = G_{n+1} - G_0 \qquad \Longrightarrow \qquad \sum_{k=0}^{\infty}(G_{k+1} - G_k) = \lim_{n\to\infty}(G_{n+1} - G_0) \qquad \Longrightarrow$$
$$\Longrightarrow \qquad \sum_{k=0}^{\infty}(G_{k+1} - G_k) = \lim_{k\to\infty} G_k - B$$

Zato vrijedi

$$\begin{aligned}
\lim_{k\to\infty} G_k &= B + \sum_{k=0}^{\infty}(G_{k+1} - G_k) = B + \sum_{k=0}^{\infty} \tfrac{1}{2}\tfrac{w_k c^T + c w_k^T}{(c,p)} = B + \tfrac{1}{2(c,p)}\sum_{k=0}^{\infty} w_k c^T + \tfrac{1}{2(c,p)}\sum_{k=0}^{\infty} c w_k^T \\
&= B + \tfrac{1}{2(c,p)}\left[(I-P)^{-1}(y-Bp)c^T + c(y-Bp)^T(I-P^T)^{-1}\right] = \\
&\quad \text{prema Lemi 4.1. :} \quad I - P = I - \tfrac{1}{2}\left[I - \tfrac{cp^T}{(c,p)}\right] = \tfrac{1}{2}\left[I + \tfrac{cp^T}{(c,p)}\right], \qquad \sigma = 1 + p^T\tfrac{c}{(c,p)} = 2 \\
&\quad (I-P)^{-1} = 2\left[I - \tfrac{1}{2}\tfrac{cp^T}{(c,p)}\right], \qquad (I-P^T)^{-1} = 2\left[I - \tfrac{1}{2}\tfrac{pc^T}{(c,p)}\right] \qquad , \\
&= B + \tfrac{1}{2(c,p)}\left\{2\left[I - \tfrac{1}{2}\tfrac{cp^T}{(c,p)}\right](y-Bp)c^T + c(y-Bp)^T 2\left[I - \tfrac{1}{2}\tfrac{pc^T}{(c,p)}\right]\right\} \\
&= B + \tfrac{1}{(c,p)}\left\{(y-Bp)c^T + c(y-Bp)^T - \tfrac{1}{2}\tfrac{cp^T}{(c,p)}(y-Bp)c^T - \tfrac{1}{2}c(y-Bp)^T\tfrac{pc^T}{(c,p)}\right\} \\
&= B + \tfrac{1}{(c,p)}\left\{(y-Bp)c^T + c(y-Bp)^T - \tfrac{c(y-Bp)^T pc^T}{(c,p)}\right\}
\end{aligned}$$

iz čega neposredno slijedi (4.36). ♣

Specijalno za $c = p$ dobiva se Powell–Broydenova simetrična korekcija (Powell Symetric Broyden - PSB)

$$\overline{B}_P SB = B + \frac{(y-Bp)p^T + p(y-Bp)^T}{(p,p)} - \frac{((y-Bp),p)pp^T}{(p,p)^2}. \tag{4.38}$$

### 4.2.3. Korekcije koje čuvaju pozitivnu definitnost

Do sada smo promatrali korekcije ranga 1 ili 2 koje su čuvale simeričnost matrice $B$ i zadovoljavale quasi-Newtonov uvjet. Dodatni zahtjev koji se postavlja bit će čuvanje pozitivne definitnosti. Quasi-Newtonov uvjet zajedno s uvjetom čuvanja pozitivne definitnosti matrice $B$ pobliže određuje kut između vektora $p$ i $y$:

$$(y,p) > 0 \qquad \left[(y,p) \overset{\overline{B}p=y}{=} (\overline{B}p,p)\right]$$

U cilju ispitivanja nasljednosti pozitivne definitnosti koristit ćemo sljedeću lemu (vidi primjerice [7])

**Lema 4.4.** *Neka je $A \in \mathbb{R}^{n\times n}$ simetrična matrica sa svojstvenim vrijednostima*

$$\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n,$$

*i neka je $A^T = A + \sigma uu^T$ za neki $u \in \mathbb{R}^n$. Ako je $\sigma \ge 0$, onda $A^T$ ima svojstvene vrijednosti $\lambda_i^T$, takve da je*

$$\lambda_1 \le \lambda_1^T \le \lambda_2 \le \cdots \le \lambda_n \le \lambda_n^T,$$

*a ako je $\sigma \le 0$, onda $A^T$ ima svojstvene vrijednosti $\lambda_i^T$, takve da je*

$$\lambda_1^T \le \lambda_1 \le \lambda_2^T \le \cdots \le \lambda_n^T \le \lambda_n.$$

**Teorem 4.7.** *Neka je $B \in \mathbb{R}^{n\times n}$ simetrična pozitivno definitna matrica i neka su $c, p, y \in \mathbb{R}^n$ takvi da je $(c,p) \neq 0$. Tada je $\overline{B}$ definiran s (4.36) pozitivno definitna matrica onda i samo onda ako je* $\det \overline{B} > 0$.

Dokaz. Ako je $\overline{B} > 0$, onda je iz opće teorije poznato da je $\det \overline{B} > 0$. Za dokaz obrata, matricu $\overline{B}$ napišimo najprije u obliku

$$\overline{B} = B + \tfrac{(y-Bp)c^T + c(y-Bp)^T}{c^Tp} - \tfrac{(y-Bp)^Tp}{(c^Tp)^2}cc^T$$
$$\overline{B} = B + vw^T + wv^T,$$

gdje je

$$v = \frac{y-Bp}{(c,p)} - \frac{1}{2}\frac{(y-Bp,p)}{(c,p)^2}c, \qquad w = c.$$

Zbog toga je $\left[(v+w)(v+w)^T - (v-w)(v-w)^T = \left(\|v\|^2 + \|w\|^2 + 2vw^T\right) - \left(\|v\|^2 + \|w\|^2 - 2wv^T\right)\right]$

$$\overline{B} = B + \frac{1}{2}\left((v+w)(v+w)^T - (v-w)(v-w)^T\right)$$

Na taj način smo $\overline{B}$ napisali kao zbroj od jedne pozitivno definitne matrice ($B$) i dvije simetrične matrice ranga 1. Matrica

$$C = B + \frac{1}{2}(v+w)(v+w)^T$$

prema Lemi 4.4. je pozitivno definitna jer je $\sigma = \frac{1}{2} > 0$. Nadalje, također prema Lemi 4.4., obzirom da je $C > 0$ i $\sigma = -\frac{1}{2} < 0$ matrica

$$\overline{B} = C - \frac{1}{2}(v-w)(v-w)^T$$

može imati najviše jednu svojstvenu vrijednost nepozitivnu. Kako je s druge strane

$$0 < \det \overline{B} = \lambda_1 \cdots \lambda_n$$

i ta jedna svojstvena vrijednost mora biti pozitivna. Dakle, $\overline{B} > 0$. ♣

U cilju ispitivanje za koji $c \in \mathbb{R}^n$ će rekurzivna formula (4.36) čuvati pozitivnu definitnost sljedeća lema ispituje $\det \overline{B}$.

**Lema 4.5.** *Neka je $u_i \in \mathbb{R}^n$, $i = 1,2,3,4$. Tada vrijrdi*

$$\det\left(I + u_1u_2^T + u_3u_4^T\right) = \left(1 + (u_1,u_2)\right)\left(1 + (u_3,u_4)\right) - (u_1,u_4)(u_2,u_3).$$

Dokaz. Ako trenutno pretpostavimo da je $(u_1,u_2) \neq -1$, tj. da je, sukladno Sherman-Morrisonovoj lemi, $I + u_1u_2^T$ regularna matrica, onda možemo pisati

$$I + u_1u_2^T + u_3u_4^T = \left(I + u_1u_2^T\right)\left(I + (I + u_1u_2^T)^{-1}u_3u_4^T\right)$$

Zato je $\det\left(I + u_1u_2^T + u_3u_4^T\right) = D_1 \cdot D_2$, gdje je

$$D_1 = \det\left(I + u_1u_2^T\right) = 1 + (u_1,u_2),$$
$$D_2 = \det\left(I + (I + u_1u_2^T)^{-1}u_3u_4^T\right).$$

Kako je $\left(I + (I + u_1u_2^T\right)^{-1} \overset{(4.1.)}{=} I - \frac{1}{1+(u_1,u_2)}u_1u_2^T$, onda

$$\begin{aligned} D_2 &= \det\left[I + \left(I - \tfrac{u_1u_2^T}{1+(u_1,u_2)}\right)u_3u_4^T\right] = \det\left[I + u_3u_4^T - \tfrac{u_1u_2^Tu_3u_4^T}{1+(u_1,u_2)}\right] \\ &= \det\left[I + \left(u_3 - \tfrac{(u_2,u_3)}{1+(u_1,u_2)}u_1\right)u_4^T\right] = 1 + \left(u_3 - \tfrac{(u_2,u_3)}{1+(u_1,u_2)}u_1, u_4\right) \\ &= 1 + (u_3,u_4) - \tfrac{(u_2,u_3)}{1+(u_1,u_2)}(u_1,u_4), \end{aligned}$$

iz čega slijedi tražena jednakost. ♣

Primjenom Leme 4.5. na (4.36) uz pretpostavku da je postoji $B^{-1}$ i oznaku $H := B^{-1}$, dobivamo

$$\det \overline{B} = \det\, B\left[\left((c,Hy)^2-(c,Hc)(y,Hy)+(c,Hc)(y,p)\right)\frac{1}{(c,p)^2}\right]. \tag{4.39}$$

U svrhu dokaza ove formule najprije ćemo (4.36)

$$\overline{B} = B + \frac{(y-Bp)c^T + c(y-Bp)^T}{c^Tp} - \frac{(y-Bp)^Tp}{(c^Tp)^2}cc^T$$

napisati u obliku [ $(y-Bp)^Tp$ je broj pa ga možemo "staviti unutar" $cc^T$ ]

$$\overline{B} = B + \frac{1}{(c,p)}\left[(y-Bp)c^T + c(y-Bp)^T\left(I-\frac{pc^T}{(c,p)}\right)\right]$$

Uvođenjem oznake $H := B^{-1}$, dobivamo

$$\overline{B} = B\left[I + \frac{H(y-Bp)}{(c,p)}c^T + Hc\frac{(y-Bp)^T}{(c,p)}\left(I-\frac{pc^T}{(c,p)}\right)\right].$$

Sada vrijedi

$$\det \overline{B} = \det B \det\left(I + u_1u_2^T + u_3u_4^T\right)$$

gdje je

$$\begin{aligned} u_1 &= \tfrac{Hy-p}{(c,p)}, \quad u_2 = c,\\ u_3 &= Hc, \qquad u_4 = \left(I-\tfrac{cp^T}{(c,p)}\right)\tfrac{(y-Bp)}{(c,p)}. \end{aligned}$$

Prema Lemi 4.5. je $\det\left(I+u_1u_2^T+u_3u_4^T\right) = (1+(u_1,u_2))\,(1+(u_3,u_4)) - (u_1,u_4)(u_2,u_3)$, gdje je

$$\begin{aligned} 1+(u_1,u_2) &= 1 + c^T\tfrac{Hy-p}{(c,p)} = 1 + \tfrac{c^THy}{(c,p)} - \tfrac{c^Ty}{(c,p)} = \tfrac{c^THy}{(c,p)}\\ 1+(u_3,u_4) &= 1 + c^T\left(H - \tfrac{Hcp^T}{(c,p)}\right)\tfrac{(y-Bp)}{(c,p)} = 1 + \tfrac{c^T}{(c,p)}\left(Hy - \tfrac{Hcp^Ty}{(c,p)} - s + \tfrac{Hcp^TBp}{(c,p)}\right) =\\ &= 1 + \tfrac{c^THy}{(c,p)} - \tfrac{cTHcp^Ty}{(c,p)^2} - 1 + \tfrac{cTHcp^TBp}{(c,p)^2},\\ (1+(u_1,u_2))\,(1+(u_3,u_4)) &= \tfrac{(c,Hy)^2}{(c,p)^2} - \tfrac{(c,Hy)(c,Hc)(p,y)}{(c,p)^3} + \tfrac{(c,Hy)(c,Hc)(p,Bp)}{(c,p)^3}\\ (u_1,u_4) &= \tfrac{y^TH-p^T}{(c,p)^2}\left(I-\tfrac{cp^T}{(c,p)}\right)(y-Bp) = \tfrac{y^THy}{(c,p)^2} - \tfrac{p^Ty}{(c,p)^2} - \tfrac{y^THcp^Ty}{(c,p)^3} + \tfrac{y^THcp^TBp}{(c,p)^3} \end{aligned},$$

iz čega slijedi

$$\det\left(I+u_1u_2^T+u_3u_4^T\right) = \frac{(c,Hy)^2}{(c,p)^2} - \frac{(c,Hc)(y,Hy)}{(c,p)^2} + \frac{(c,Hc)(y,p)}{(c,p)^2}$$

i tražena formula (4.39). Ako pretpostavimo da je $B>0$, onda je i $H := B^{-1} > 0$, pa možemo definirati $v = H^{1/2}y$ i $w = H^{1/2}c$. Time (4.39) postaje

$$\det \overline{B} = \det\, B\left[\left((v,w)^2 - \|v\|^2\|w\|^2 + \|w\|^2(y,p)\right)\frac{1}{(c,p)^2}\right]. \tag{4.40}$$

Na osnovi Teorema 4.7. zaključujemo da je $\overline{B} > 0$ onda i samo onda ako vrijedi

$$\|w\|^2(y,p) > \|v\|^2\|w\|^2 - (v,w)^2. \tag{4.41}$$

Jednostavan način kako se može zadovoljit ova nejednakost uveo je Davidon (1959) (kasnije objašnjeno i implementirano u radu Fletchera i Powella 1963). Naime ako je $w \parallel v$ (odnosno $c \parallel y$), onda je (4.41) zadovoljeno ako je $(y, p) > 0$. Zato se poznata Davidon – Fletchera – Powella korekcija (DFP) ranga 2 dobiva iz (4.36) uz $c = y$

$$\begin{aligned} \overline{B}_{DGP} &= B + \tfrac{(y-Bp)y^T + y(y-Bp)^T}{(y,p)} - \tfrac{(y-Bp,p)}{(y,p)^2} yy^T. \\ &= \left(I - \tfrac{yp^T}{(y,p)}\right) B \left(I - \tfrac{py^T}{(y,p)}\right) + \tfrac{yy^T}{(y,p)}. \end{aligned} \tag{4.42}$$

**Zadatak 4.6.** *Provjerite valjanost formula (4.42)!*

Vrijedi sljedeći teorem

**Teorem 4.8.** *Neka je $B \in \mathbb{R}^{n\times n}$ regularna simetrična matrica i neka je $\overline{B}_{DFP} \in \mathbb{R}^{n\times n}$ definiran sa (4.42) za neke vektore $y, p \in \mathbb{R}^n$, koji zadovoljavaju $(y, p) \neq 0$.*

*Tada je $\overline{B}_{DFP}$ regularna onda i samo onda ako je $(y, Hy) \neq 0$, gdje je $H := B^{-1}$.*

*Ako je $\overline{B}_{DFP}$ regularna, tada se $\overline{H}_{DFP} := \overline{B}_{DFP}^{-1}$ može zapisati kao*

$$\overline{H}_{DFP} = H + \frac{pp^T}{(y,p)} - \frac{Hyy^T H}{(y,Hy)}. \tag{4.43}$$

*Nadalje, ako je $B > 0$, onda je i $\overline{B}_{DFP} > 0$ onda i samo onda ako vrijedi $(y, p) > 0$.*

**Zadatak 4.7.** *Provjerite valjanost formula (4.42) i dokažite* Teorem *4.8.!*

## 4.3. Inverzne korekcije

U prethodnim razmatranjima nastojali smo dobiti rekurzivne formule za matrice–korekcije, koje što više nalikuju na hesijan minimizirajuće funkcije. Sada ćemo pokušati definirati rekurzivne formule za matrice–korekcije, koje će što više nalikovati na inverznu matricu hesijana minimizirajuće funkcije. Te matrice nazivat ćemo inverzne korekcije.

U tu svrhu pretpostavimo da posjedujemo neku aproksimaciju $H$ od $(\nabla^2 f(x))^{-1}$ i pomoću nje pokušajmo pronaći što bolju aproksimaciju $\overline{H}$ od $(\nabla^2 f(\overline{x}))^{-1}$, gdje je $\overline{x} = x + p$. Analogno quasi-Newtonovom uvjetu (4.25) $(\overline{B}p = y)$ postavimo inverzni quasi-Newtonov uvjet

$$\overline{H}y = p. \tag{4.44}$$

Odmah je jasno da opća inverzna korekcija ranga 1, koja zadovoljava inverzni quasi-Newtonov uvjet (4.44) mora biti oblika

$$\overline{H} = H + \frac{(p - Hy)d^T}{(d,y)}, \qquad \text{za neki} \quad d \in \mathbb{R}^n, \quad (d,y) \neq 0. \tag{4.45}$$

**Zadatak 4.8.** *Pokžite da korekcija ranga 1 (4.26) i inverzna korekcija ranga 1 (4.45) reprezentiraju istu korekciju ako je $c = B^T d$.*

Kao i u slučaju prethodno razmatranih korekcija koje aproksimiraju $(\nabla^2 f(x))$ i ovdje se najprije postavlja pitanje čuvanja nalјeđa simetričnosti, tj. vrijedi li

$$H \quad \text{je simetrična matrica} \qquad \Longrightarrow \qquad \overline{H} \quad \text{je simetrična matrica}$$

**Zadatak 4.9.** *Pokažite da je inverzna korekcija ranga* 1 *koja zadovoljava inverzni quasi-Newtonov uvjet (4.44) i čuva simetričnost također zadana formulom (4.34).*

**Zadatak 4.10.** *Slično kao u dokazu* Leme *4.3. izvedite formulu za inverznu quasi-Newtonova korekcija ranga* 2

$$\overline{H} = H + \frac{(p - Hy)d^T + d(p - Hy)^T}{(d, y)} - \frac{(p - Hy)^T y}{(d, y)^2} dd^T. \tag{4.46}$$

Najvažniju varijantu inverzne quasi-Newtonove korekcije ranga 2 (4.46), koja pri tome čuva i pozitivnu definitnost, nezavisno su napravili Broyden, Fletcher, Goldfarb i Shano 1970. godine, pa se ona u literaturi naziva Broyden–Fletcher–Goldfarb–Shano korekcija ili skraćeno BFGS-korekcija. Dobiva se tako da u (4.46) stavimo $d = p$. Može se pokazati da u tom služaju vrijedi

$$\overline{H}_{BFGS} = \left(I - \frac{py^T}{(y, p)}\right) H \left(I - \frac{yp^T}{(y, p)}\right) + \frac{pp^T}{(y, p)}. \tag{4.47}$$

kao i sljedći teorem

**Teorem 4.9.** *Neka je $H \in \mathbb{R}^{n \times n}$ regularna simetrična matrica i neka je $\overline{H}_{BFGS} \in \mathbb{R}^{n \times n}$ definirana sa (4.47) za neke vektore $y, p \in \mathbb{R}^n$, $(y, p) \neq 0$.*

*Tada je $\overline{H}_{BFGS}$ regularna onda i samo onda ako je $(p, Bp) \neq 0$, gdje je $B = H^{-1}$.*

*Ako je $\overline{H}_{BFGS}$ regularna, onda se $\overline{B}_{BFGS} := \overline{H}_{BFGS}^{-1}$ može zapisati kao*

$$\overline{B}_{BFGS} = B + \frac{yy^T}{(y, p)} - \frac{Bpp^T B}{(p, Bp)}.$$

*Nadalje, ako je $H > 0$, onda je $\overline{H}_{BFGS} > 0$ onda i samo onda ako je $(y, p) > 0$.*

**Zadatak 4.11.** *Provjerite valjanost formule (4.47) i dokažite* Teorem *4.9.!*

# 5. Seminarski radovi

1. Dva različita dokaza Teorema 4.6. mogu se vidjeti u [10] i [7]. Izvedite i usporedite oba dokaza.
   Literatura: [10], Ch. 6.6.2 (pp 178-179), [7], (pp 66-69)

2. Provjerite valjanost formula (4.42) i dokažite Teorem 4.8.!

3. Izradite zadatke: Zadatak 4.3., Zadatak 4.8., Zadatak 4.9..

4. Slično kao u dokazu Leme 4.3. izvedite formulu za inverznu quasi-Newtonova korekcija ranga 2 navedenu u (4.46). Pokažite da u slučaju $d = p$ vrijedi (4.47)

$$\overline{H}_{BFGS} = \left(I - \frac{py^T}{(y,p)}\right) H \left(I - \frac{yp^T}{(y,p)}\right) + \frac{pp^T}{(y,p)}.$$

5. Provjerite valjanost formule (4.47) i dokažite Teorem 4.9.!

6. Primijenite direktne i inverzne quasi-Newtonove korekcije na Gauss-Newtonovu metodu za rješavanje nelinearnog problema najmanjih kvadrata.
   Literatura: [10], Ch. 6.7.1 (pp 186-191), [19]

7. Najbolja $L_2$ aproksimacija funkxije $f : [a,b] \to \mathbb{R}$ na podprostoru svih polinoma $\mathcal{P}_n$ stupnja $\leq n$. Ortogonalni polinomi.
   Literatura: [4], Ch. 6.8 (pp 421-434), [14], Ch. 15 (401-411)

8. Najbolja $L_2$ aproksimacija funkxije $f : [a,b] \to \mathbb{R}$ na podprostoru svih trigonometrijskih polinoma $\mathcal{T}_n$ stupnja $\leq n$.
   Literatura: [4], Ch. 6.8 (pp 421-434), [18], Ch. 5 (93-103)

9. Trigonometrijska interpolacija i aproksimacija
   Literatura: [4], Ch. 6.12 (pp 477-490),

10. Najbolja $L_\infty$ aproksimacija funkxije $f : [a,b] \to \mathbb{R}$ na podprostoru svih polinoma $\mathcal{P}_n$ stupnja $\leq n$. Čebiševljevi polinomi.
    Literatura: [4], Ch. 6.8 (pp 434-449), [14], Ch. 15 (411-421)

11. Najbolja $L_p$, $\quad p = 1, 2, \infty$ aproksimacija podataka mjerenja $y_i$, $i = 1, \ldots, m$
    Literatura: radni papiri

12. Shepardova interpolacija i aproksimacija
    Literatura: [4], Ch. 6.10 (pp 459-462)

13. Moving least squares
    Literatura: [4], Ch. 6.10 (pp 465-467), [19]

14. Least Squares Spline 1. reda
    [13], [4], Ch. 6.14 (pp 493-498),

15. Linearno programiranje
    [4], Ch. 10 (pp 735-767),

16. Metoda zlatnog raza za traženje loklnog minimuma jednodimenzionalne funkcije bez korištenja derivacija

    Literatura: Jare, Stoer [10], Zlobec [23], Jukić, Scitovski [11]

17. Metoda parabole za traženje loklnog minimuma jednodimenzionalne funkcije bez korištenja derivacija

    Literatura: Jare, Stoer [10], Zlobec [23], Jukić, Scitovski [11]

18. Brentova metoda za traženje loklnog minimuma jednodimenzionalne funkcije bez korištenja derivacija

    Literatura: Brent [2], Ch. 5, Jare, Stoer [10], Zlobec [23]

19. Traženje globalnog minimuma jednodimenzionalne funkcije bez korištenja derivacija

    Literatura: Brent [2], Ch. 6

20. Powellova metoda za traženje minimuma višedimenzionalne funkcije bez korištenja derivacija

    Literatura: Brent [2], Ch. 7

21. Test funkcije za traženje minimuma jednodimenzionalne i višedimenzionalne funkcije

    Literatura: Brent [2], Ch. 7, Gill [8], Jare, Stoer [10], Kelley [12], Himelblau

## Literatura

[1] M. Alić, G. Nogo, *Optimizacija: Uvod u teoriju nužnih i dovoljnih uvjeta ekstrema*, Odjel za matematiku, Sveučili'vste u Osijeku, Osijek, 2004.

[2] R. P. Brent, *Algorithms for Minimization without Derivatives*, Dover Publications, New York, 1973.

[3] Broyden, *A class of methods for solving nonlinear simultaneous equation*, Math. Comput., **19**(1965), 577-593

[4] D. Kincaid,W. cheney, *Numerical Analysis*, Brooks/cole Publishing company, New York, 1996.

[5] W.C.Davidon, *Variable metric method for minimization*, Rep. ANL-5990 Rev, Argonne National Laboratories, Argonne, Ill., 1959.

[6] J. E. Dennis, Jr, R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, 1996

[7] J. E. Dennis Jr., J. J. More, *Quasi-Newton methods, motivation and theory*, SIAM Review, **19**(1977), 46-89

[8] P. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, Academic Press, 1981

[9] J. B. Hiriart-Urruty, C. Lemaréchal, *Fundamentals of Convex Analysis*, Springer-Verlag, Berlin - Heidelberg, 2001.

[10] F. Jare, J. Stoer, *Optimierung*, Springer-Verlag, Berlin, 2004

[11] D. Jukić, R. Scitovski, *Metode optimizacije – prdavanja*, Tuzla, 2005.

[12] C. T. Kelley, *Iterative Methods for Optimization*, SIAM, Philadelphia, 1999

[13] P. Lancaster, K. Šalkauskas, *Curve and Surface Fitting: An Introduction*, Academic Press, London, 1986.

[14] R. Plato, *Concise Numerical Mathematics*, Providence : American Mathematical Society, 2003.

[15] J. M. Ortega,W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970. (postoji i ruski prijevod)

[16] B.N.Psenicnij, J,M.Danilin, *Čislenie metodi v ekstremaljnih zadačah*, Nauka, Moskva, 1975

[17] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge, 1989

[18] R. Scitovski, *Numerička matematika*, Odjel za matematiku, Sveučilište u Osijeku, Osijek, 2004

[19] R. Scitovski, Š. Ungar, D. Jukić, *Approximating surfaces by moving total least squares method*, Applied Mathematics and Computation **93**(1998) 219-232

[20] J. Stoer, R. Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, New York, 1993

[21] F. P. Vasiljev, Lekcii po metodam rešenija ekstremalnih zadač, Moskovski univerzitet, Moskva, 1974.

[22] S. Wolfram, *The Mathematica Book*, Wolfram Media, Champaign, 1999.

[23] S. Zlobec, J. Petrić, *Nelinearno programiranje*, Naučna knjiga, Beograd, 1989.