# Optimizacija pomoću gradijentne metode

Marijan Čančarević[1], Šimun Zlopaša[2], Dragana Čulina[3]

## Uvod

U mnogim područjima (ekonomija, medicina, građevina, fizika...) često se rješenje nekog problema svodi na uporabu neke od metoda optimizacije odnosno određivanja lokalnih ili globalnih ekstrema funkcije. Također, zadnjih je godina došlo do naglog razvoja umjetne inteligencije koja nalazi primjenu u razvoju znanosti, olakšava poslovanje i čini svakodnevni život ugodnijim. Algoritmi dubokog učenja podloga su modernog pristupa umjetnoj inteligenciji, a većina njih sadrži neku vrstu optimizacije, najčešće optimizaciju gradijentnom metodom [2].

Metode optimizacije općenito mogu biti analitičke i numeričke, s ograničenjima i bez ograničenja, jednodimenzijske ili višedimenzijske. Poznate metode optimizacije su metoda zlatnog reza, parabole, Newtonova metoda, gradijentna metoda, simpleks metoda... , a sve one imaju i svoje podvarijante [5]. U ovom je članku opisana optimizacija pomoću gradijentne metode i ilustrirana na jednome primjeru. Ukratko je navedeno potrebno predznanje iz funkcija više varijabli. Na kraju je dan program napisan u matematičkom softwareu SageMath [4] u kojemu je implementirana ova metoda.

## Ekstremi funkcija više varijabli

Navedimo neke dobro poznate definicije i činjenice o ekstremima funkcije više varijabli [1,3].

Neka je zadana funkcija $f : S \to \mathbb{R},\ S \subseteq \mathbb{R}^n, n \geq 2,\ n \in \mathbb{N}$. Funkcija $f(x_1, x_2, \ldots, x_n)$ ima u točki $c = (c_1, c_2, \ldots, c_n)$ *lokalni minimum* (*lokalni maksimum*) ako postoji okolina $\mathcal{O}_c$ oko točke $c$ tako da za svaki $x = (x_1, x_2, \ldots, x_n) \in \mathcal{O}_c$ vrijedi $f(c) \leq f(x)$ $\left(f(c) \geq f(x)\right)$.

Funkcija $f(x_1, x_2, \ldots, x_n)$ je diferencijabilna na skupu $S$ (klase $C^1$) ako postoje sve prve parcijalne derivacije i one su neprekidne. Za proizvoljni $x = (x_1, x_2, \ldots, x_n)$, vektor

[1]Marijan Čančarević, Visoka škola za informacijske tehnologije
[2]Šimun Zlopaša, Veleučilište Velika Gorica
[3]Dragana Čulina, Visoka škola za informacijske tehnologije

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

ili često zapisan i u obliku

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

naziva se *gradijent funkcije f* u točki *x*. Iz definicije gradijenta vidi se da su njegove skalarne projekcije na koordinatne osi u pravokutnom koordinatnom sustavu parcijalne derivacije prvog reda zadane funkcije.

**Primjer 1.** Izračunajmo gradijent funkcije $f(x,y) = 3x^2 - 2xy$ u točki $(-1,2)$.

*Rješenje:*

$$\frac{\partial f}{\partial x}(x,y) = 6x - 2y, \quad \frac{\partial f}{\partial y}(x,y) = -2x,$$

$$\frac{\partial f}{\partial x}(-1,2) = 6\cdot(-1) - 2\cdot 2 = -10, \qquad \frac{\partial f}{\partial x}(-1,2) = -2\cdot(-1) = 2.$$

Gradijent zadane funkcije u zadanoj točki je vektor

$$\nabla f(-1,2) = (-10,2)^T = \begin{bmatrix} -10 \\ 2 \end{bmatrix}$$

Ako diferencijabilna funkcija $f(x_1, x_2, \ldots, x_n)$ u točki $c = (c_1, c_2, \ldots, c_n)$ ima lokalni ekstrem, onda je $\nabla f(c) = \vec{0}$. Točke u kojima je gradijent funkcije jednak nul-vektoru nazivaju se *stacionarne točke*. Dakle, nužan, ali ne i dovoljan uvjet lokalnih ekstrema funkcije je da se sve parcijalne derivacije u stacionarnoj točki poništavaju. Naime, funkcija može imati stacionarnu točku, a ne mora imati i ekstrem u toj točki.

**Primjer 2.** Nađimo stacionarne točke funkcije $f(x,y) = 3xy - 4x + xy^2$.

*Rješenje:*

$$\frac{\partial f}{\partial x}(x,y) = 3y - 4 + y^2, \qquad \frac{\partial f}{\partial y}(x,y) = 3x + 2xy .$$

Izjednačimo parcijalne derivacije s nulom:

$$y^2 + 3y - 4 = 0, \quad 3x + 2xy = 0 .$$

Riješimo sustav:

$$y_{1,2} = \frac{-3 \pm \sqrt{25}}{2} = \frac{-3 \pm 5}{2}, \quad y_1 = 1,\ y_2 = -4.$$

$$x(3 + 2y) = 0,\ x = 0 \ . \quad \left( y = -\frac{3}{2} \text{ otpada} \right).$$

Stacionarne točke funkcije su: $(0,1)$ i $(0,-4)$ (kandidati za lokalne ekstreme).

Matrica kojoj su elementi parcijalne derivacije drugog reda funkcije $f\left(x_1, x_2, \ldots, x_n\right)$ naziva se *Hesseova matrica* (*Hessijan*):

$$H\left(x_1, x_2, \ldots, x_n\right) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial {x_1}^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial {x_2}^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \dfrac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial {x_n}^2} \end{bmatrix}$$

**Primjer 3.** Funkciji $f\left(x, y\right) = 3xy - 4x + xy^2$ pridružimo pripadnu Hesseovu matricu.

*Rješenje:*

Imamo

$$\frac{\partial f}{\partial x} = 3y - 4 + y^2, \quad \frac{\partial f}{\partial y} = 3x + 2xy,$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 3 + 2y, \quad \frac{\partial^2 f}{\partial x^2} = 0, \quad \frac{\partial^2 f}{\partial y^2} = 2x$$

i prema tome je

$$H\left(x, y\right) = \begin{bmatrix} 0 & 3+2x \\ 3+2x & 2x \end{bmatrix}$$

U prethodnim smo primjerima našli gradijent funkcije, stacionarne točke i Hesseovu matricu. Time smo postavili temelje za nalaženje ekstrema funkcija više varijabli. Naime, ako je $c = \left(c_1, c_2, \ldots, c_n\right)$ stacionarna točka dva puta diferencijabilne funkcije $f\left(x_1, x_2, \ldots, x_n\right)$ i $D_k$, $k = 1, 2, \ldots, n$ glavne subdeterminante (minore) pridružene Hesseovoj matrici u točki $c$, onda funkcija $f\left(x_1, x_2, \ldots, x_n\right)$

a) ako su $D_k > 0$, $k = 1, 2, ..., n$, ima lokalni minimum u točki $c$,

b) ako su za parne $k$ $D_k > 0$, a za neparne $k$ $D_k < 0$, ima lokalni maksimum u točki $c$,

c) ako je za barem jedan paran $k$ $D_k < 0$ ili postoje neparni $k$ i $l$ takvi da je $D_k > 0$ i $D_l < 0$, nema lokalnih ekstrema.

**Primjer 4.** Odredimo ekstreme funkcije $f\left(x, y\right) = x^3 + y^3 - 3xy$.

*Rješenje:*

$$\frac{\partial f}{\partial x} = 3x^2 - 3y, \quad \frac{\partial f}{\partial y} = 3y^2 - 3x,$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = -3, \quad \frac{\partial^2 f}{\partial x^2} = 6x, \quad \frac{\partial^2 f}{\partial y^2} = 6y.$$

Izjednačimo li prve parcijalne derivacije s nulom, dobiva se

$$3(x^2 - y) = 0$$
$$3\left(y^2 - x\right) = 0 .$$

Iz prve jednadžbe nalazimo $y = x^2$ i uvrštavanjem u drugu jednadžbu imamo jednadžbu $x^4 - x = 0$ odnosno

$$x\left(x^3 - 1\right) = 0,\ x(x-1)\left(x^2 - x + 1\right) = 0 .$$

Zadnja jednadžba daje $x_1 = 0,\ x_2 = 1$, a kako je $y = x^2$, to su $y_1 = 0, y_2 = 1$, a time i stacionarne točke $(0,0),(1,1)$ odnosno kandidati za ekstreme funkcije.

Nadalje je

$$H(x,y) = \begin{bmatrix} 6x & -3 \\ -3 & 6y \end{bmatrix}$$

$$H(0,0) = \begin{bmatrix} 0 & -3 \\ -3 & 0 \end{bmatrix}$$

i

$$H(1,1) = \begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix}.$$

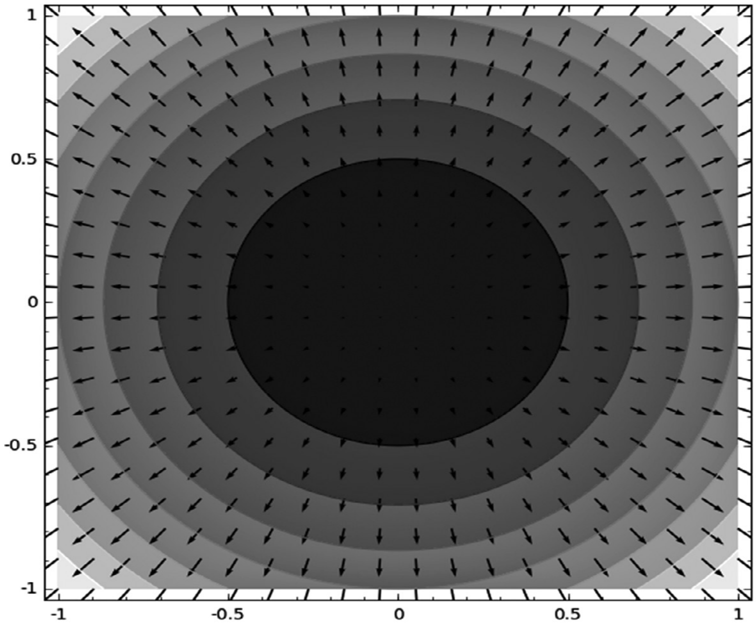
Kako je u točki $(0,0)$ vrijednost $D_2 = 0 \cdot 0 - (-3) \cdot (-3) = -9 < 0$, zaključujemo da funkcija nema ekstrema.

Za stacionarnu točku $(1,1)$ je $D_1 = 6 > 0$ i $D_2 = 6 \cdot 6 - (-3) \cdot (-3) = 27 > 0$. Dakle, funkcija $f(x,y) = x^3 + y^3 - 3xy$ ima lokalni minimum koji iznosi

$$f(1,1) = 1^3 + 1^3 - 3 \cdot 1 \cdot 1 = -1$$

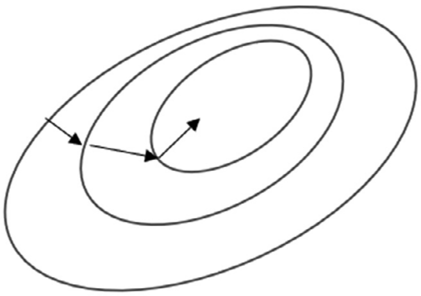
## Gradijentna metoda

Numeričke metode optimizacije u kojima se koristi gradijent nazivaju se *gradijentne metode*. Geometrijski, gradijent $\nabla f(x) \neq \vec{0}$ ima svojstvo da je okomit na nivo plohu $f(x) = a$ u proizvoljnoj točki te plohe. Iz tog svojstva proizlazi da gradijent funkcije $\nabla f(x)$ ima smjer i orijentaciju najbržeg rasta funkcije i, shodno tome, $-\nabla f(x)$ najbržeg pada funkcije. Na slici 1 prikazani su nivo krivulje i gradijenti funkcije $f(x,y) = x^2 + y^2$.




*Slika 1.*

Kod gradijentnih metoda cilj je pomoću zadane početne aproksimacije ekstrema, ili neke već izračunate aproksimacije, doći do sljedeće aproksimacije koristeći gradijent, kao što je idejno pokazano na slici 2.



*Slika 2.*

Korak aproksimacije dan je sljedećom formulom za maksimum

$$x^{(k+1)} = x^{(k)} + \lambda_k \nabla f\left(x^{(k)}\right)$$

odnosno za minimum

$$x^{(k+1)} = x^{(k)} - \lambda_k \nabla f\left(x^{(k)}\right) .$$

Prethodno uvedene iteracije omogućuju kretanje prema maksimumu i minimumu u smjeru gradijenta samo ako su sve vrijednosti $\lambda_k, k = 1,2,\ldots,n$ pozitivne. Postoje različite mogućnosti odabira koraka $\lambda_k$, a najveći je problem kako ga odabrati tako da se osigura konvergencija niza iteracija prema kritičnoj točki zadane funkcije. Izaberemo li čvrsti korak u svim iteracijama imamo metodu konstantnog spusta. Veliki problemi mogu nastati pri izboru koraka. Izabere li se preveliki korak, postoji mogućnost da se preskoči točka ekstrema, a opet premali korak iziskuje velik broj iteracija i velik broj izračuna. Izabranim korakom treba osigurati da je vrijednost funkcije kod računanja minimuma svakim korakom manja, odnosno gradijent funkcije sve bliži nuli. Tako jedan od kriterija zaustavljanja algoritma može biti

$$\sqrt{\sum_{i=1}^{n}\left(\frac{\partial f}{\partial x_i}(x^{(k)})\right)^2} \le \varepsilon,$$

gdje je $\varepsilon$ unaprijed zadani mali pozitivan realan broj.

Ako u $x^{(k+1)} = x^{(k)} - \lambda_k \nabla f(x^{(k)})$ korak $\lambda_k > 0$ biramo tako da minimiziramo funkciju jedne varijable

$$g(\lambda) = f\left(x^{(k)} - \lambda \nabla f\left(x^{(k)}\right)\right),\ \lambda \ge 0$$

dobiva se *metoda najbržeg spusta*. Minimizaciju funkcije $g(\lambda)$ može se provesti nekom od metoda jednodimenzijske optimizacije (npr. Newtonova metoda, metoda parabole, metoda bisekcije...).

**Primjer 1.** Metodom najbržeg spusta minimizirajmo funkciju $f(x,y) = x^3 + y^3 - 3xy$ počevši od točke $(2,2)$.

*Rješenje:*

Gradijent funkcije je

$$\nabla f(x,y) = \begin{bmatrix} 3x^2 - 3y \\ 3y^2 - 3x \end{bmatrix}, \quad \nabla f(2,2) = \begin{bmatrix} 3\cdot 2^2 - 3\cdot 2 \\ 3\cdot 2^2 - 3\cdot 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} x^{(1)} \\ y^{(1)} \end{bmatrix} = \begin{bmatrix} x^{(0)} \\ y^{(0)} \end{bmatrix} - \lambda \nabla f\left(x^{(0)}, y^{(0)}\right) = \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \lambda \begin{bmatrix} 6 \\ 6 \end{bmatrix} = \begin{bmatrix} 2-6\lambda \\ 2-6\lambda \end{bmatrix}$$

$$\begin{aligned} g(\lambda) = f\left(x^{(1)}, y^{(1)}\right) &= (2-6\lambda)^3 + (2-6\lambda)^3 - 3(2-6\lambda)(2-6\lambda) = \\ &= (2-6\lambda)^2(4-12\lambda-3) = (2-6\lambda)^2(1-12\lambda) \end{aligned}$$

$$\begin{aligned} g'(\lambda) &= 2(2-6\lambda)(-6)(1-12\lambda) + (2-6\lambda)^2(-12) = \\ &= -12(2-6\lambda)(1-12\lambda+2-6\lambda) = -36(2-6\lambda)(1-6\lambda). \end{aligned}$$

$$g'(\lambda)=0 \text{ za } \lambda_1=\frac{1}{6},\ \lambda_2=\frac{1}{3}$$

Izaberimo $\lambda=\frac{1}{6}$ .

$$\begin{bmatrix} x^{(1)} \\ y^{(1)} \end{bmatrix}=\begin{bmatrix} 2-6\cdot\frac{1}{6} \\ 2-6\cdot\frac{1}{6} \end{bmatrix}=\begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Provjerimo gradijent funkcije u točki $(1,1)$:

$$\nabla f(1,1)=\begin{bmatrix} 3\cdot1^2-3\cdot1 \\ 3\cdot1^2-3\cdot1 \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Očito je minimum funkcije u točki $(1,1)$ (to smo i znali jer smo ga našli analitički) i iznosi

$$f(1,1)=1^3+1^3-3\cdot1\cdot1=-1$$

**Napomena.** Osim analitički, točku minimuma funkcije jedne varijable $g(\lambda)$ može se naći nekom od metoda numeričkog rješavanja jednadžbe $g'(\lambda)=0$.

**Primjer 2.** Riješimo isti primjer bez jednodimenzijske optimizacije koristeći istu početnu točku i korak $\lambda=0.1$ (metoda konstantnog spusta).

**Rješenje:**

$$\nabla f(x,y)=\begin{bmatrix} 3x^2-3y \\ 3y^2-3x \end{bmatrix},\quad \nabla f(2,2)=\begin{bmatrix} 3\cdot2^2-3\cdot2 \\ 3\cdot2^2-3\cdot2 \end{bmatrix}=\begin{bmatrix} 6 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} x^{(1)} \\ y^{(1)} \end{bmatrix}=\begin{bmatrix} x^{(0)} \\ y^{(0)} \end{bmatrix}-\lambda\nabla f\left(x^{(0)},y^{(0)}\right)=\begin{bmatrix} 2 \\ 2 \end{bmatrix}-0.1\begin{bmatrix} 6 \\ 6 \end{bmatrix}=\begin{bmatrix} 2-0.6 \\ 2-0.6 \end{bmatrix}=\begin{bmatrix} 1.4 \\ 1.4 \end{bmatrix}$$

Usporedimo vrijednosti funkcije u početnoj i dobivenoj aproksimaciji:

$$f(2,2)=8+8-12=4,\qquad f(1.4,1.4)=1.4^3+1.4^3-3\cdot1.4\cdot1.4=-0.392$$

Očito se vrijednost funkcije smanjuje.
Prijeđimo na sljedeću aproksimaciju:

$$\begin{bmatrix} x^{(2)} \\ y^{(2)} \end{bmatrix}=\begin{bmatrix} x^{(1)} \\ y^{(1)} \end{bmatrix}-\lambda\nabla f\left(x^{(1)},y^{(1)}\right)=\begin{bmatrix} 1.4 \\ 1.4 \end{bmatrix}-0.1\begin{bmatrix} 3\cdot1.4^2-3\cdot1.4 \\ 3\cdot1.4^2-3\cdot1.4 \end{bmatrix}$$

$$=\begin{bmatrix} 1.4 \\ 1.4 \end{bmatrix}-0.1\begin{bmatrix} 1.68 \\ 1.68 \end{bmatrix}=\begin{bmatrix} 1.232 \\ 1.232 \end{bmatrix}$$

Usporedimo vrijednosti funkcije u zadnje dvije aproksimacije:

$$f\left(1.232{,}1.232\right)=1.232^3+1.232^3-3\cdot1.232\cdot1.232\approx-0.814$$

Vrijedi da je

$$f\left(x^{(2)},y^{(2)}\right)<f\left(x^{(1)},y^{(1)}\right)$$

Nadalje je

$$\begin{bmatrix}x^{(3)}\\y^{(3)}\end{bmatrix}=\begin{bmatrix}x^{(2)}\\y^{(2)}\end{bmatrix}-\lambda\nabla f\left(x^{(2)},y^{(2)}\right)=\begin{bmatrix}1.232\\1.232\end{bmatrix}-0.1\begin{bmatrix}3\cdot1.232^2-3\cdot1.232\\3\cdot1.232^2-3\cdot1.232\end{bmatrix}$$
$$\approx\begin{bmatrix}1.232\\1.232\end{bmatrix}-0.1\begin{bmatrix}0.860\\0.860\end{bmatrix}=\begin{bmatrix}1.232-0.086\\1.232-0.086\end{bmatrix}=\begin{bmatrix}1.146\\1.146\end{bmatrix}$$

Izračunajmo vrijednost funkcije u $\left(x^{(3)},y^{(3)}\right)$:

$$f\left(1.232{,}1.232\right)=1.146^3+1.146^3-3\cdot1.146\cdot1.146\approx-0.930$$

Četvrta aproksimacija je

$$\begin{bmatrix}x^{(4)}\\y^{(4)}\end{bmatrix}=\begin{bmatrix}x^{(3)}\\y^{(3)}\end{bmatrix}-\lambda\nabla f\left(x^{(3)},y^{(3)}\right)=\begin{bmatrix}1.146\\1.146\end{bmatrix}-0.1\begin{bmatrix}3\cdot1.146^2-3\cdot1.146\\3\cdot1.146^2-3\cdot1.146\end{bmatrix}$$
$$\approx\begin{bmatrix}1.146\\1.146\end{bmatrix}-0.1\begin{bmatrix}0.500\\0.500\end{bmatrix}=\begin{bmatrix}1.146-0.050\\1.146-0.050\end{bmatrix}=\begin{bmatrix}1.096\\1.096\end{bmatrix}$$

Analognim postupkom nalaze se:

$$\begin{bmatrix}x^{(5)}\\y^{(5)}\end{bmatrix}=\begin{bmatrix}1.064\\1.064\end{bmatrix},\quad\begin{bmatrix}x^{(6)}\\y^{(6)}\end{bmatrix}=\begin{bmatrix}1.044\\1.044\end{bmatrix},\quad\begin{bmatrix}x^{(7)}\\y^{(7)}\end{bmatrix}=\begin{bmatrix}1.030\\1.030\end{bmatrix},\quad\begin{bmatrix}x^{(8)}\\y^{(8)}\end{bmatrix}=\begin{bmatrix}1.021\\1.021\end{bmatrix},\ldots.$$

Uvrstimo li vrijednosti 8. aproksimacije u analitički zadanu funkciju, dobiva se:

$$f\left(1.021{,}1.021\right)=1.021^3+1.021^3-3\cdot1.021\cdot1.021\approx-0.9987$$

(prisjetimo se da je točna vrijednost minimuma funkcije $-1$) .

Iz provedenog računa vidi se da je gradijent u svakoj sljedećoj aproksimaciji sve bliži i bliži nuli (nulvektoru). Zanimljivo bi bilo vidjeti što se događa s nizom $x^{(k)}$ ako se nakon nekoliko iteracija mijenja korak $\lambda$.

## SageMath program za optimizaciju pomoću gradijentne metode

Ako se poslužimo računalom i primijenimo neki matematički software, onda se problem optimizacije primjenom gradijentne metode bitno pojednostavnjuje. Tako se pomoću SageMath programa prethodno navedeni primjer može riješiti na sljedeći način:

```
#gradijentna metoda, korekcija koraka
var('x,y')
f = x^3+y^3-3*x*y
x0=2.;y0=2.
a=0.1
n=0
G11=diff(f,x).subs(x=x0,y=y0)
G21=diff(f,y).subs(x=x0,y=y0)
G0=Matrix(RR,2,1,[G11,G21])
while norm(G0)>10.^-3:
    G11=diff(f,x).subs(x=x0,y=y0)
    G21=diff(f,y).subs(x=x0,y=y0)
    G0=Matrix(RR,2,1,[G11,G21])
    r0=Matrix(RR,2,1,[x0,y0])
    r=r0-a*G0
    x1=r[0,0];y1=r[1,0]
    if f(x=x1,y=y1)<f(x=x0,y=y0):
        x0=x1;y0=y1
        n=n+1
    else:
        a=0.5*a
x0,y0,n,G0, f(x=x0,y=y0)
([0.000587521417269699] 1.00013704999222, 1.00013704999222, 22,
[0.000587521417269699], -0.999999943646751)
```

Nakon 22 iteracije dobiva se točka minimuma

$$(1.00013704999222, 1.00013704999222),$$

vrijednost minimuma

$$-0.999999943646751$$

i gradijent u točki minimuma

$$\begin{bmatrix} 0.000587521417269699 \\ 0.000587521417269699 \end{bmatrix}.$$

Isti primjer izborom malog čvrstog koraka $(a = 0.1)$ može se riješiti na sljedeći način:

```
# gradijentna metoda bez korekcije koraka
vars = var('x y')
f = x^3+y^3-3*x*y
x0=2;y0=2.
a=0.1
n=0
G11=diff(f,x).subs(x=x0,y=y0)
G21=diff(f,y).subs(x=x0,y=y0)
G0=Matrix(RR,2,1,[G11,G21])
while norm(G0)>10.^-3:
   r0=Matrix(RR,2,1,[x0,y0])
   r=r0-a*G0
   x0=r[0,0];y0=r[1,0]
   G11=diff(f,x).subs(x=x0,y=y0)
   G21=diff(f,y).subs(x=x0,y=y0)
   n=n+1
   G0=Matrix(RR,2,1,[G11,G21])
x0,y0,n,G0, f(x=x0,y=y0)
```

([0.000587521417269699] 1.00019580213395, 1.00019580213395, 21,
[0.000587521417269699], -0.999999884969560)

Nakon 21 iteracije dobiva se točka minimuma

$$(1.00019580213395, 1.00019580213395),$$

vrijednost minimuma

$$-0.99999988496956$$

i gradijent u točki minimuma

$$\begin{bmatrix} 0.000587521417269699 \\ 0.000587521417269699 \end{bmatrix}.$$

Smanjenje koraka ($a = 0.5a$) realizira se u trenutku kada vrijednost funkcije u tekućoj aproksimaciji nije manja nego u prethodnoj. To se dogodilo, prema dobivenim rezultatima iz drugog koda, u jednom slučaju.

Ako optimiziramo korak $\lambda$ (u programu $t$, jednodimenzijska optimizacija):

```
vars = var('x y t')
f = x^3+y^3-3*x*y
```

```
x0=2.;y0=2.
n=0
G11=diff(f,x).subs(x=x0,y=y0)
G21=diff(f,y).subs(x=x0,y=y0)
G0=Matrix(RR,2,1,[G11,G21])
while norm(G0)>10.^-3:
   assume(t>0)
   g(t)=f(x=x0-t*G11,y=y0-t*G21)
   a=find_local_minimum(g,0,0.5)
   r0=Matrix(RR,2,1,[x0,y0])
   r=r0-a[1]*G0
   n=n+1
   x0=r[0,0]
   y0=r[1,0]
   G11=diff(f,x).subs(x=x0,y=y0)
   G21=diff(f,y).subs(x=x0,y=y0)
   G0=Matrix(RR,2,1,[G11,G21])
r,n

array([[ 1.],
       [ 1.]]), 1)
```

Nakon izvršenog koda vidimo da se točka minimuma podudara s rješenjem koje se dobilo u 1. primjeru nakon prve iteracije. Pri jednodimenzijskoj optimizaciji koraka $\lambda$ korištena je numerička metoda find_local_minimum$(g,a,b)$ softwarea SageMath koja kao rezultat vraća dvodimenzijski vektor s komponentama vrijednosti minimuma funkcije $g$ i točke minimuma.

## Literatura


1. Čulina, B.; Zlopaša, Š. 2015. *Matematika za tehničke visoke škole, peti dio: Višedimenzionalna matematika promjena.* Veleučilište Velika Gorica.
2. Goodfellow, I.J.; Bengio, Y.; Courville, A. 2016. *Deep Learning*. MIT Press.
3. Kurepa, S. 1989. *Matematička analiza 3: Funkcije više varijabli*, Golden marketing – Tehnička knjiga.
4. *SageMath*, https://www.sagemath.org/
5. Scitovski, R.; Truhar, N.; Tomljanović, Z. 2014. *Metode optimizacije*. Odjel za matematiku Sveučilišta J. J. Strossmayera u Osijeku.