Osječki matematički list xx (20xx), 1-13

Lucijana Grgić [*] Kristian Sabo [†]

Sažetak

U radu je opisana poznata Nelder-Meadova metoda, koja se smatra jednom od najpopularnijih lokalnih metoda direktne bezuvjetne optimizacije. Zbog jednostavnosti, analiziran je specijalni slučaj optimizacije u $\mathbb{R}^2$, jer se tada Nelder-Meadova metoda svodi na niz elementarnih geometrijskih transformacija u ravnini, te je za njezino potpuno razumijevanje dovoljno znanje srednjoškolske matematike. U svrhu ilustracije metode, dano je nekoliko numeričkih primjera koji su izrađeni u programskom paketu *Mathematica.*

Ključne riječi: *Bezuvjetna optimizacija, Lokalna optimizacija, Direktna metoda, Nelder-Meadova metoda, Simpleks algoritam*

Nelder-Mead's Method-Local Method for Direct Unconstrained Optimization

Abstract

In this paper the popular Nelder-Mead's method is described, which is considered to be one of the best known local method of the direct unconstrained optimization. Because of simplicity, a special case of optimization in $\mathbb{R}^2$ has been analyzed since then Nelder-Mead's method is coming down to a sequence of elementary geometric transformations in plane so, for its complete understanding, the high-school mathematical knowledge is enough. In order to illustrate the method, it is given several numerical examples which are made in software named *Mathematica.*

Kexwords: *Unconstrained optimization, Local optimization, Direct method, Nelder-Mead method, Simplex search algorithm*

[*] studentica Odjela za matematiku, Sveučilišta J. J. Strossmayera u Osijeku, email: lucijana.grgic@mathos.hr

[†] Odjel za matematiku, Sveučilište J. J. Strossmayera u Osijeku, email: ksabo@mathos.hr

1 Uvod

Mnogi problemi koji potječu iz različitih područja primjena svode se na traženje globalnog minimuma ili maksimuma funkcije $f : \mathbb{R}^n \to \mathbb{R}$ na skupu $\mathbb{R}^n$, odnosno točke iz $\mathbb{R}^n$ u kojoj se taj minimum, odnosno maksimum, postižu. Podsjetimo se kako za $x^\star \in \mathbb{R}^n$ kažemo da je točka globalnog minimuma funkcije f na $\mathbb{R}^n$ ako vrijedi $f(x^\star) \leq f(x)$, za svaki $x \in \mathbb{R}^n$. U tom slučaju pišemo

$$x^\star = \operatorname*{argmin}_{x \in \mathbb{R}^n} f(x). \tag{1}$$

Vrijednost funkcije f u točki $x^\star$ zovemo globalni minimum funkcije te pišemo

$$f(x^\star) = \min_{x \in \mathbb{R}^n} f(x). \tag{2}$$

Slično, za $x^\star \in \mathbb{R}^n$ kažemo da je točka globalnog maksimuma funkcije f na $\mathbb{R}^n$ ako vrijedi $f(x^\star) \geq f(x)$, za svaki $x \in \mathbb{R}^n$ i pišemo

$$x^\star = \operatorname*{argmax}_{x \in \mathbb{R}^n} f(x). \tag{3}$$

Pri tome, vrijednost funkcije f u točki $x^\star$ zovemo globalni maksimum funkcije f i pišemo

$$f(x^\star) = \max_{x \in \mathbb{R}^n} f(x). \tag{4}$$

Problemi (1)-(2), odnosno (3)-(4), u literaturi se obično nazivaju i problemima bezuvjetne globalne optimizacije.

Kako je

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

bez smanjenja općenitost, umjesto oba problema (1)-(2), odnosno (3)-(4), dovoljno je analizirati samo problem traženja globalnog minimuma funkcije, odnosno točke u kojoj se taj minimum postiže. Nažalost, određivanje točke globalnog minimuma funkcije f, općenito je vrlo ozbiljan i zahtjevan posao i ponekad zahtjeva poznavanje svih točaka lokalnih minimuma funkcije f na $\mathbb{R}^n$. Sjetimo se kako za $x^\star \in \mathbb{R}^n$ kažemo da je točka lokalnog minimuma funkcije f na $\mathbb{R}^n$, ako postoji otvorena kugla $K(x^\star, \delta)$ sa središtem u točki $x^\star$ polumjera $\delta > 0$ takva da vrijedi $f(x^\star) \leq f(x)$, za svaki $x \in K(x^\star, \delta)$. Pri tome vrijednost $f(x^\star)$ zovemo lokalni minimimum funkcije f. Metode za traženje točke lokalnog minimuma nazivaju se lokalne optimizacijske metode.

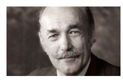
Ako funkciju f, umjesto na $\mathbb{R}^n$, promatramo na nekoj omeđenoj i zatvorenoj domeni $\mathcal{D} \subset \mathbb{R}^n$, u cilju traženja točke globalnog minimuma mogu se koristiti neke poznate globalne optimizacijske metode. Ovdje ćemo spomenuti dvije takve metode. Prva metoda poznata pod nazivom `DIRECT` metoda za globalnu optimizaciju ([3]), koristi se ako je funkcija f Lipshitz-neprekidna na $\mathcal{D}$, te ako $\mathcal{D}$ ima specijalni oblik tzv. hiperkvadra. U općem slučaju, mogu se koristiti globalni genetički algoritmi te odgovarajuće modifikacije (vidjeti primjerice [6]).

U ovom radu govorit ćemo o jednoj lokalnoj optimizacijskoj metodi. Kao što je već ranije navedeno, takve metode traže točku lokalnog minimuma funkcije f na $\mathbb{R}^n$ i općenito ne postoji jamstvo da je dobiveno rješenje ujedno točka globalnog minimuma. Treba spomenuti da neke specijalne funkcije, kao što su konveksne funkcije, imaju svojstvo da točka lokalnog minimuma, koja je određena nekom lokalnom metodom, ujedno predstavlja točku globalnog minimuma te funkcije (vidjeti [1]).

Ovisno o uvjetima koje zadovoljava funkcija f, za traženje točke lokalnog minimuma, mogu se koristiti različite lokalne optimizacijske metode. Ako je, primjerice, funkcija f diferencijabilna i znamo izračunati njezin gradijent u svakoj točki domene, u svrhu lokalne minimizacije može se koristiti gradijentna metoda te njezine modifikacije ([9]). Ako je funkcija f dva puta neprekidno diferencijabilna, te nam je poznat Hessijan funkcije f, koriste se Newtonova metoda i njezine modifikacije ([9]). Navedene metode međusobno se značajno razlikuju, kako po brzini konvergencije, tako i u točnosti.

Ako funkcija f nije derivabilna, ili je derivabilna, no gradijent i Hessijan se vrlo teško računaju, za traženje točaka lokalnih minimuna smisleno je koristiti metode koje su zasnovane isključivo na izračunavanju vrijednosti funkcije u točkama domene. Takve metode nazivaju se direktnim metodama ([4]).

U ovom radu opisat ćemo jednu od najpoznatijih lokalnih direktih metoda bezuvjetne optimizacije, poznatu pod nazivom Nelder-Meadova metoda. Ime metode potječe od njezinih autora Johna Neldera i Rogera Meada, koji su metodu konstruirali i objavili prije 50 godina, točnije 1965. godine u radu [5]. Neleder-Meadova metoda je jednostavna i u mnogim slučajevima vrlo efikasna, posebno u optimizacijskim problemima manjih dimenzija. Kao što će se vidjeti, Nelder-Meadovu metodu moguće je primijeniti na funkciju za koju je dovoljno poznavati isključivo vrijednosti u svim točkama domene. Nažalost, kod optimizacijskih problema u velikim dimenzijama metoda se ne pokazuje korisnom (vidi primjerice [4],[8]). Zanimljivo je da općenito ne postoji dokaz konvergencije ove metode. U radu [4] dani su uvjeti konvergencije za specijalni tip funkcije u jednoj i dvije dimenzije. Bez obzira na to, metoda je često korištena i citirana u mnoštvu različitih primjena ([7]). Tako se, primjerice, naredba `fminsearch` za traženje točke lokalnog minimuma u programskom sustavu *Matlab* zasniva na



John Nelder (8. listopada 1924.-7. kolovoza 2010.)

Roger Mead

Nelder-Meadovoj metodi ([10]). Više detalja o ovoj metodi može se naći u pregledu [11].

U ovom radu, zbog jednostavnosti, analiziramo specijalni slučaj optimizacije u $\mathbb{R}^2$, jer se onda Nelder-Meadova metoda svodi na niz elementarnih geometrijskih transformacija trokuta u ravnini, te je za njezino potpuno razumijevanje dovoljno znanje srednjoškolske matematike. Potpuno analogno, metoda se može proširiti i na više dimenzija, pri čemu se u višim dimenzijama umjesto trokuta promatra prirodna generalizacija trokuta, koja je poznata kao simpleks, pa se zbog toga metoda često naziva i Nelder-Meadova simpleks metoda.

Članak je organiziran na sljedeći način. U drugom odjeljku opisana je jedna iteracija Nelder-Meadove metode, kao i geometrijske transformacije na kojima je iteracija zasnovana. Također, naveden je jedan kriterij zaustavljanja kojim je određen broj iteracija Nelder-Meadove metode. U trećem odjeljku dano je nekoliko numeričkih primjera kojima su ilustrirane mogućnosti, kao i nedostaci ove metode.

2 Jedna iteracija Nelder-Meadove metode u dvije dimenzije

U ovom odjeljku detaljno ćemo opisati jednu iteraciju Nelder-Meadove metode u ravnini. Pretpostavimo da je zadana funkcija $f : \mathbb{R}^2 \to \mathbb{R}$. Promatrajmo trokut $\triangle ABC \subset \mathbb{R}^2$ s vrhovima $A = (x_1, y_1), B = (x_2, y_2)$ i $C = (x_3, y_3)$. Bez smanjenja općenitosti možemo pretpostaviti da je

$$f(A) \leq f(B) \leq f(C).$$

Pri tome ćemo točku A zvati najbolji vrh, točku C najlošiji vrh, dok ćemo točku B zvati drugi najlošiji vrh trokuta $\triangle ABC$.

Osnovna ideja jedne iteracije Nelder-Meadove metode sastoji se u tome da trokut $\triangle ABC$ zamijenimo novim trokutom u čijim vrhovima funkcija f postiže manju vrijednost od one koja se postiže u vrhovima trokuta $\triangle ABC$. U tu svrhu, s $M = \frac{1}{2}(A + B)$ označimo polovište vrhova A i B. Sada najlošiji vrh C preslikamo centralno-simetrično u odnosu na točku M. Na taj način dobivamo vrh

$$R = M + (M - C) = 2M - C = A + B - C.$$

Transformaciju koja je trokut $\triangle ABC$ preslikala u trokut $\triangle ABR$ zvat ćemo refleksija trokuta (vidi Sliku 2 (b)). Pri tome nas zanima je li $f(R) < f(A)$, odnosno je li vrh R bolji od najboljeg vrha A?

- Ako je odgovor pozitivan, naslućujemo da se krećemo u pravom smjeru, te se pomaknemo u istom smjeru do točke

$$
\begin{aligned}
E &= R + (R - M) = 2R - M = 2A + 2B - 2C - \frac{1}{2}A - \frac{1}{2}B \\
&= \frac{3}{2}A + \frac{3}{2}B - 2C.
\end{aligned}
$$

  Transformaciju koja trokut $\triangle ABC$ preslikava u trokut $\triangle ABE$ zovemo ekspanzija trokuta (Slika 2 (c)). Točku C zamijenit ćemo točkom E ako je $f(E) < f(R)$, odnosno s točkom R ako je $f(E) \geq f(R)$.

- Ako je odgovor negativan, odnosno ako je $f(R) \geq f(A)$, točku R uspoređujemo s drugom najlošijom točkom B, odnosno postavljamo sljedeće pitanje: je li $f(R) < f(B)$?
  - Ako je odgovor pozitivan, točku C zamijenjujemo s točkom R.
  - Ako je odgovor negativan, postavljamo novo pitanje: je li $f(R) < f(C)$?
    * Ako je odgovor pozitivan, točku C zamijenjujemo s točkom R te definiramo novu točku

$$
\begin{aligned}
P &= M + \frac{1}{2}(R - M) = \frac{1}{2}M + \frac{1}{2}R \\
&= \frac{3}{4}A + \frac{3}{4}B - \frac{1}{2}C.
\end{aligned}
$$

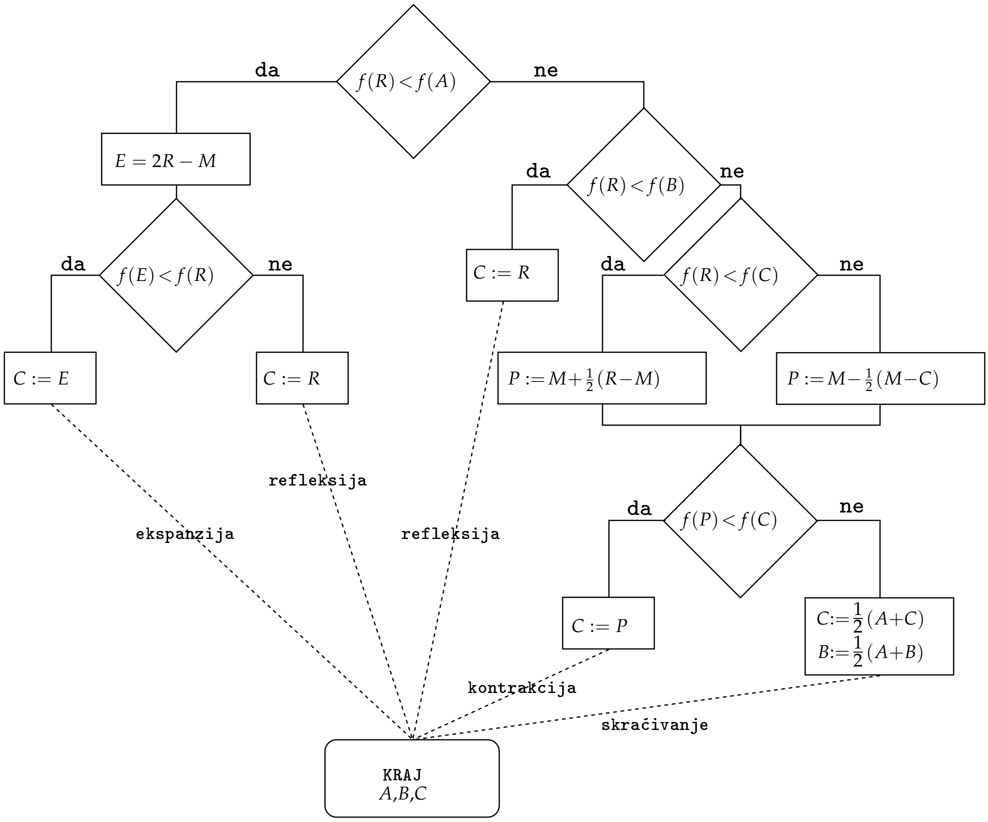
    * Ako je odgovor negativan, definiramo novu točku

$$
\begin{aligned}
P =& M - \frac{1}{2}(M - C) = \frac{1}{2}M + \frac{1}{2}C \\
&= \frac{1}{4}A + \frac{1}{4}B + \frac{1}{2}C.
\end{aligned}
$$

    Transformaciju koja trokut $\triangle ABC$ preslikava u trokut $\triangle ABP$ zovemo kontrakcija trokuta (Slika 2 (d) i (e)). Konačno, pitamo se je li $f(P) < f(C)$.
    * Ako je odgovor pozitivan, točku C zamijenjujemo s točkom P.
    * Ako je odgovor negativan, definiramo $H = \frac{1}{2}(A + C)$ te točku B zamijenimo s M, a točku C zamijenimo s H.

    Transformaciju koja trokut $\triangle ABC$ preslikava u trokut $\triangle AMH$ zovemo skraćivanje trokuta (Slika 2 (f)).

Dijagram toka jedne iteracije Nelder-Meadove metode u $\mathbb{R}^2$ prikazan je na Slici 1.[1]



Slika 1: Dijagram toka jedne iteracije Nelder-Meadeove metode u $\mathbb{R}^2$

Pseudo-kod jedne iteracije Nelder-Meadove metode u ravnini dan je u Algoritmu 1.

Iteracije Algoritma 1 se ponavljaju sve dok se ne zadovolji neki kriterij zaustavljanja. Jedna mogućnost je da se Algoritam 1 ponavlja sve dok udaljenost najboljeg vrha do težišta trokuta $\triangle ABC$ ne postane manja od unaprijed zadana točnosti $\varepsilon > 0$ (vidjeti primjerice [5]). Treba spomenuti da se još mogu i kombinirati različite strategije u provedbi geometrijskih transformacija trokuta. O takvim strategijama, kao i o pitanjima konvergen-

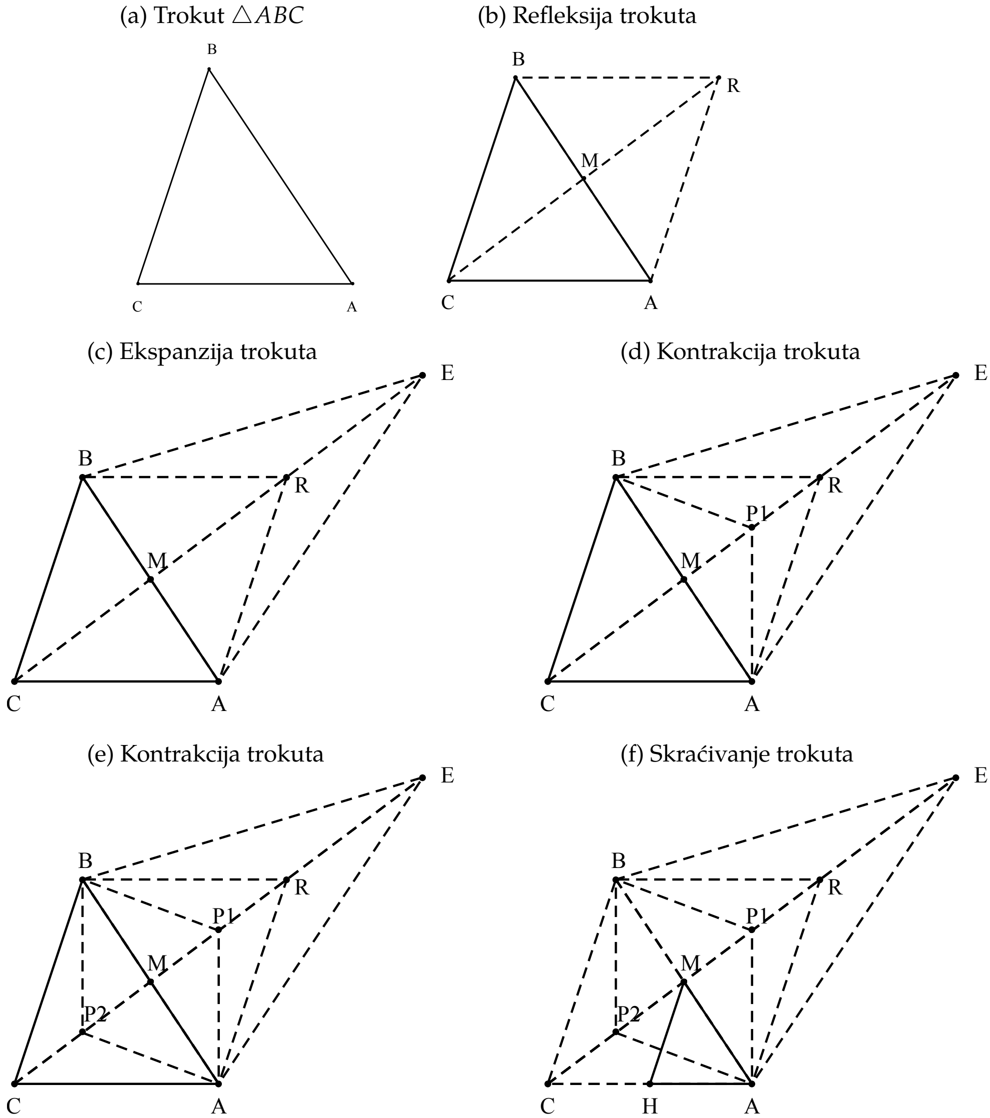
[1]Dijagram toka jedne iteracije Nelder-Meadove metode u $\mathbb{R}^2$ izradio je Branimir Šajinović

Algorithm 1 Jedna iteracija Nelder-Meadove metode u ravnini

Input: $f, \quad A = (x_1, y_1),\ B = (x_2, y_2),\ C = (x_3, y_3)$

poredaj A, B, C tako da je $f(A) \leq f(B) \leq f(C)$
$M = \frac{1}{2}(A + B)$ *//polovište dobre strane*
$R = M + (M - C) = 2M - C$ *//refleksija*
if $f(R) < f(A)$ **then**
  $E = R + (R - M) = 2R - M$ *//ekspanzija*
  if $f(E) < f(R)$ **then**
    $C = E$
  else
    $C = R$
  end if
else
  if $f(R) < f(B)$ **then**
    $C = R$
  else
    if $f(R) < f(C)$ **then**
      $P = M + \frac{1}{2}(R - M)$ *//kontrakcija*
    else
      $P = M - \frac{1}{2}(M - C)$ *//kontrakcija*
    end if
  end if
  if $f(P) < f(C)$ **then**
    $C = P$
  else
    $B = \frac{1}{2}(A + B)$ *//skraćivanje*
    $C = \frac{1}{2}(A + C)$ *//skraćivanje*
  end if
end if

Output: A, B, C



Slika 2: Transformacije trokuta

cije Nelder-Meadove metode nećemo govoriti, jer izlaze izvan okvira ovog članka. Zainteresiranog čitatelja upućujemo na [4].

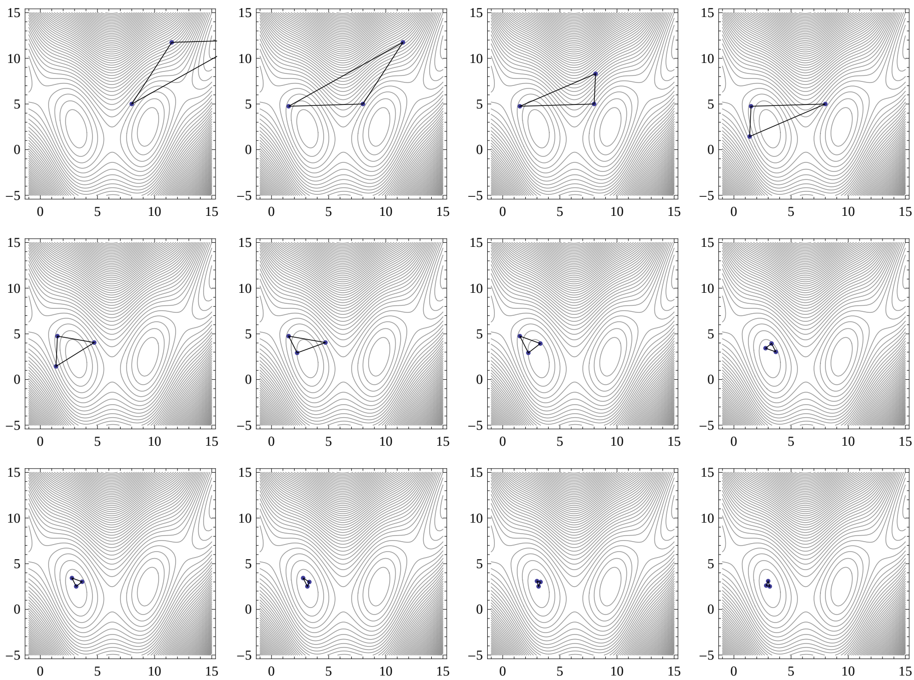
3 Numerički primjeri

U ovom odjeljku dat ćemo nekoliko primjera kojima ćemo ilustrirati mogućnosti i nedostatke Nelder-Meadove metode.

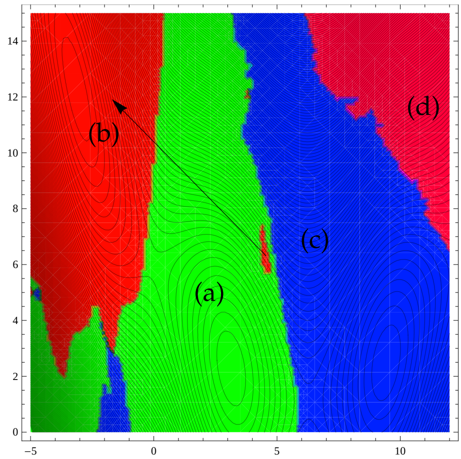
Primjer 1. Promatramo tzv. Braninovu funkciju $f : \mathbb{R}^2 \to \mathbb{R}$ (vidi [3]):

$$f(x_1, x_2) = \left(x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x_1 + 10$$

Ova funkcija postiže globalni minimum $f_{min} \approx 0.398$ u beskonačno mnogo točaka na $\mathbb{R}^2$. Na skupu $[-5, 10] \times [0, 15]$ su to točke $x_1^* = (\pi, 2.275)$, $x_2^\star = (-\pi, 12.275)$, $x_3^\star = (3\pi, 2.475)$. Spomenimo da skup $[-5, 10] \times [0, 15]$ ne sadrži niti jednu točku lokalnog minimuma Braninove funkcije.



Slika 3: Iterativni proces Nelder-Meadove metode kroz 12 iteraciju za Braninovu funkciju.



Slika 4: Područja početnih aproksimacija

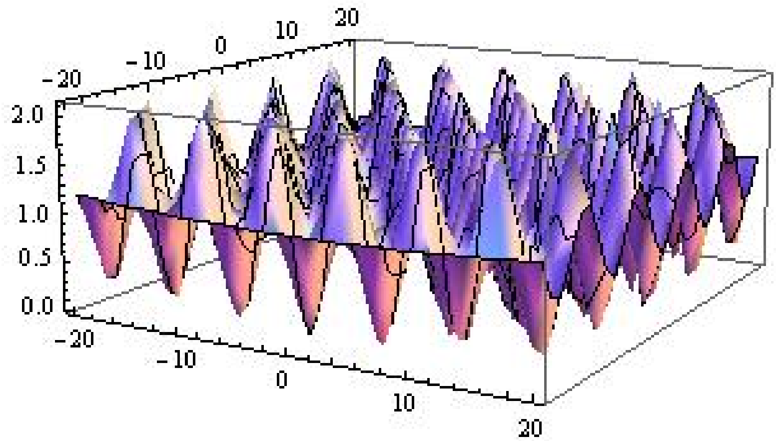
Na Slici 3 prikazan je cijeli iterativni proces Nelder-Meadove metode kroz 12 iteracija, pri čemu su vrhovi početnog trokuta $A = (8, 15), B = (10, 12), C = (10, 15)$. Vidimo da je iterativni proces konvergirao točki $x_1^* = (\pi, 2.275)$.

Smisleno je postaviti pitanje zašto je Nelder-Meadova metoda konvergirala upravo točki $x_1^* = (\pi, 2.275)$, a ne nekoj od točaka $x_2^\star$ ili $x_3^\star$. Odgovor na ovo pitanje krije se u izboru početne aproksimacije. U svrhu analize izbora početne aproksimacije, napravit ćemo jedan numerički eksperiment u kome ćemo odrediti skup onih početnih aproksimacija koje vode k istoj točki globalnog minimuma. Kako Nelder-Meadova metoda za početnu aproksimaciju zahtijeva tri vrha početnog trokuta $\triangle ABC$, problem ćemo malo pojednostaviti. Umjesto tri vrha, zadajemo jednu točku A dok ostale dvije određujemo tako da su $B = A + (1, 0)$ te $C = A + (0, 1)$. Na taj način možemo smatrati da Nelder-Meadovu metodu pokrećemo s jednom početnom aproksimacijom. Na Slici 4 istom su bojom prikazana područja početnih aproksimacija na skupu $[-5, 10] \times [0, 15]$ koje konvergiraju k istoj točki globalnog minimuma. Uočavamo četiri takva područja: (a), (b), (c) i (d). U područjima (a), (b), (c) nalaze se sve one početne aproksimacije koje konvergiraju redom točkama $x_1^*, x_2^\star$, te $x_3^\star$. Isto tako području (d) pripadaju one početne aproksimacije koje konvergiraju točki $x_4^\star = (5\pi, 12.875)$. U svrhu izrade Slike 4 definirana je funkcija $\Psi : [-5, 10] \times [0, 15] \to \mathbb{R}$, koja svakoj početnoj aproksimaciji $A = (x, y)$ na skupu $[-5, 10] \times [0, 15]$ pridružuje apscisu točke kojoj je Nelder-Meadova metoda konvergirala. Slika 4 rezultat je *Mathematica* naredbe `ContourPlot[Psi[x,y] , {x, -5, 12}, {y,0, 15}, ColorFunction -> Hue]`.

Primjer 2. Funkcija $f : \mathbb{R}^n \to \mathbb{R}$ zadana formulom

$$f(x_1, \dots, x_n) = 1 + \frac{1}{4000} \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos \frac{x_i}{\sqrt{i}}$$

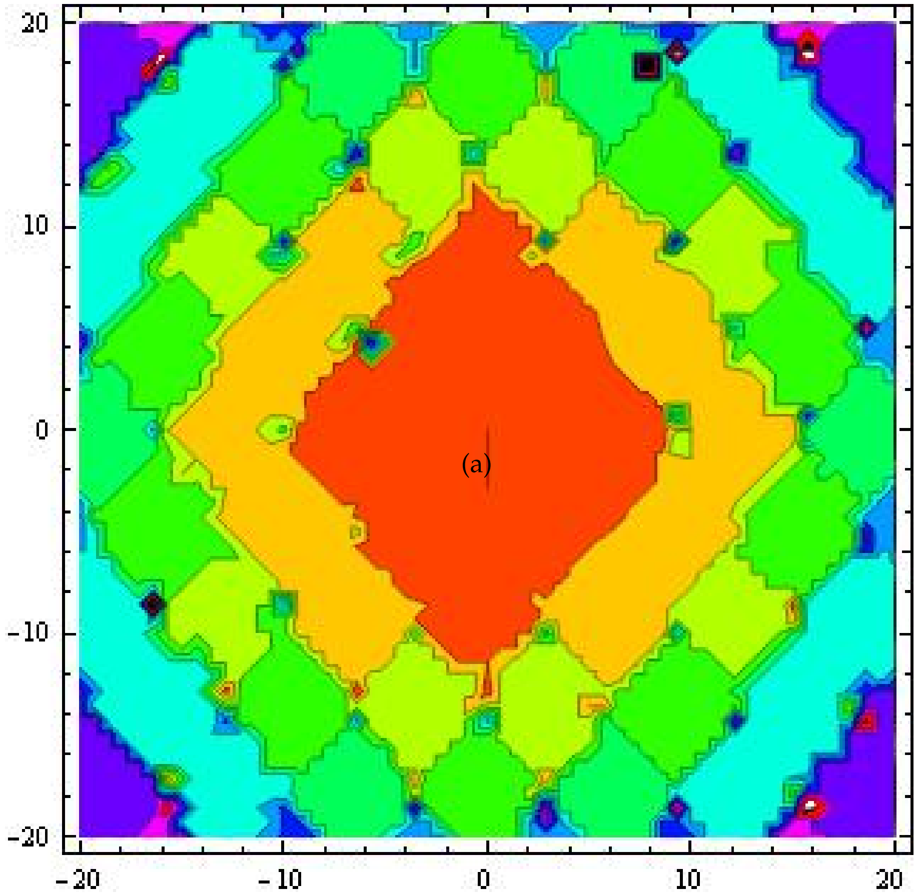
poznata je kao Griewankova funkcija ([2]) i često se koristi za testiranje metoda globalne optimizacije. Griewankova funkcija ima jedinstvenu točku globalnog minimuma $x^\star = (0, \dots, 0)$ te ima beskonačno mnogo točaka lokalnih minimuma na $\mathbb{R}^n$. Graf Griewankove funkcije za $n = 2$ prikazan je na Slici 5 . Na Slici 5 istom bojom obojane su one početne aproksimacije koje konvergiraju k istoj točki lokalnog minimuma. Nažalost, samo točke u području (a) konvergiraju točki globalnog minimuma, dok svi ostali izbori početne aproksimacije konvergiraju prema nekoj točki lokalnog minimuma, što upućuje na to da Nelder-Meadova metoda općenito nije prikladna za traženje točke globalnog minimuma.



Slika 5: Graf Griewankove funkcije na $[-20, 20] \times [-20, 20]$

Literatura

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



Slika 6: Područja početnih aproksimacija koje konvegiraju ka istoj točki lokalnog minimuma

[2] A. O. Griewank, *Generalized Decent for Global Optimization*, JOTA **34**(1981), 11–39

[3] D. R. Jones, C. D. Perttunen, B. E. Stuckman, *Lipschitzian optimization without the Lipschitz constant*, JOTA **79**(1993), 157–181

[4] J. C. Lagarias, J. A. Reeds, M. H. Wright, M.H., P. Wright, *Convergence properties of the Nelder-Mead simplex algorithm in low dimensions*, SIAM J. Optim. **9**(1998), 112–147

[5] J. A. Nelder, R. Mead, *A simplex method for function minimization*, Comput. J. **7**(1965), 308–313

[6] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evoluti on Programs*, Springer, New York, 1996

[7] W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, (second ed.), Cambridge University Press, Cambridge, New York, 1992

[8] K. V. Price, R. M. Storn, J. A. Lampinen, *Differential Evolution. A Practical Approach to Global Optimization*, Springer-Verlag Berlin Heidelberg, 2005

[9] R. Scitovski, N. Truhar, Z. Tomljanović, *Metode optimizacije*, Sveučilište Josipa Jurja Strossmayera u Osijeku, Odjel za matematiku., Osijek, 2014.

[10] http://www.scholarpedia.org/article/MATLAB

[11] S. Singer, J. Nelder, *Nelder-Mead algorithm*, Scholarpedia, **4**(2009), 7, 2928