The 9th Conference on Mathematics in Engineering: Theory and Applications
Novi Sad, May 31st–June 2nd 2024

METODI NJUTNOVOG TIPA [1]

Tijana Ostojić [2] i Manojlo Vuković [3]

https://doi.org/10.24867/META.2024.16 Review article

Sažetak. Njutnov metod je jedan od osnovnih i najpoznatijih alata u numeričkoj analizi, operacionim istraživanjima i optimizaciji. Pored dobrih teoretskih osobina, kao što je lokalna kvadratna konvergencija, ovaj iterativni metod ima brojne primene u različitim naukama, gde se pokazao kao efikasan za rešavanje problema numeričke optimizacije. Međutim, postoji značajan skup praktičnih problema u kojima implementacija Njutnovog metoda zahteva visoke računske troškove, što nas dovodi do metoda Njutnovog tipa, koji čine temu ovog preglednog rada. Cilj rada je da se da prikaz ključnh ideja metoda Njutnovog tipa u istorijskoj perspektivi, kao i savremena istraživanja uz odgovarajuće reference.

AMS klasifikacija (2020): 90C30, 90C53

Ključne reči: Njutnov metod, metodi Njutnovog tipa, kvazi-Njutnovi metodi, spektralni gradijentni metod, pravac pretrage

1. Uvod

Posmatramo problem optimizacije bez ograničenja gde je potrebno pronaći optimalno rešenje x^* tako da važi

$$f(x^*) = \min_{x\in\mathbb{R}^n} f(x). \tag{1.1}$$

Problem (1.1) se može rešiti iterativno tako što se odabere početna tačka x_0, a potom se generiše niz iterativnih tačaka $\{x_k\}_{k\in\mathbb{N}}$, pri čemu se naredna tačka određuje na osnovu prethodne. Cilj je da niz $\{x_k\}_{k\in\mathbb{N}}$ bude generisan tako da teži ka rešenju x^* posmatranog problema. U okviru ovog rada posmatraćemo metode linijskog pretraživanja [7], pa je u nastavku formalno navedeno pravilo po kom se ažurira iteracija u ovom metodu

$$x_{k+1} = x_k + \alpha_k p_k, \tag{1.2}$$

gde $\alpha_k > 0$ predstavlja dužinu koraka i p_k pravac pretraživanja. Osnovna ideja ovog metoda leži u tome da se pronađe nova iteracija x_{k+1} sa manjom vrednosti

[1] Ovo istraživanje je podržano od strane Fakulteta tehničkih nauka Univerziteta u Novom Sadu kroz projekat ”Unapređenje nastavnog procesa na engleskom jeziku u opštim disciplinama”.

[2] Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: tijana.ostojic@uns.ac.rs

[3] Departman za opšte discipline u tehnici, Fakultet tehničkih nauka, Univerzitet u Novom Sadu, e-mail: manojlo.vukovic@uns.ac.rs

funkcije f. Većina metoda linijskog pretraživanja zahteva da p_k bude opadajući pravac kako bi se ostvarilo poboljšanje.

Definicija 1.1. Za datu tačku $x_k \in \mathbb{R}^n$ pravac $p_k \in \mathbb{R}^n$ je opadajući, ako postoji $\overline{\alpha}$ tako da važi

$$f(x_k + \alpha p_k) < f(x_k),\ \forall \alpha \in (0, \overline{\alpha}).$$

Ako je funkcija f neprekidno diferencijabilna, može se pokazati da je pravac p_k opadajući u x_k ako je sledeći uslov zadovoljen

$$\nabla f(x_k)^T p_k < 0. \tag{1.3}$$

Ova osobina, koja garantuje da se vrednost funkcije f može smanjiti duž pravca p_k, je direktna posledica razvoja prvog reda funkcije f Tejlorovom formulom u okolini tačke x_k

$$f(x_k + \alpha p_k) = f(x_k) + \alpha p_k^T \nabla f(x_k) + O(\alpha^2).$$

Primetimo, za dovoljno mali korak α sledeća nejednakost će uvek biti zadovoljena

$$f(x_k + \alpha p_k) < f(x_k),$$

što je zapravo i glavni cilj u metodama optimizacije-da se u svakoj iteraciji dobije tačka koja je bolja od trenutne, tj. tačka sa manjom vrednosti funkcije cilja.

2. Metodi Njutnovog tipa

Optimizacioni metod koji za pravac pretraživanja koristi

$$p_k = -B_k^{-1} \nabla f(x_k),$$

gde je B_k simetrična nesingularna matrica, zove se *metod Njutnovog tipa.*

Lema 2.1. *(Opadajući pravac): Ako je matrica B_k pozitivno definitna, tj. ako važi $B_k \succ 0$, tada je $p_k = -B_k^{-1}\nabla f(x_k)$ opadajući pravac.*

Primetimo da prethodna lema važi jer za pravac $p_k = -B_k^{-1}\nabla f(x_k)$ važi da je

$$\nabla f(x_k)^T p_k = -\nabla f(x_k)^T B_k^{-1} \nabla f(x_k) < 0,$$

odnosno uslov (1.3) je zadovoljen.

Najjednostavniji metod, *gradijentni metod*, dobija se za izbor $B_k = I$, tj. opadajući pravac definisan je sa $p_k = -\nabla f(x_k)$. Ovaj metod se još naziva i *metod najbržeg pada* zato što se duž tog pravca vrednost funkcije najbrže smanjuje. Ukoliko se pretpostavi da je dužina koraka fiksna za sve iteracije, važi sledeći rezultat.

Teorema 2.2. *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ neprekidno diferencijabilna konveksna funkcija i neka je ∇f $L-$Lipšic neprekidno. Neka je $\{x_k\}$ niz generisan sa* (1.2) *za pravac pretrage $p_k = -\nabla f(x_k)$ i dužinu koraka $\alpha_k = \alpha$ za svako k. Ako je $\alpha \leq 1/L$, tada x_k konvergira linearno ka rešenju problema* (1.1).

Gradijentni metod sa fiksnom dužinom koraka je veoma primenjiv i široko korišćen zbog svoje jednostavnosti i niskih računskih troškova, pošto zahteva samo izvode prvog reda i nema dodatnih izračunavanja za izbor dužine koraka. Međutim, glavna mana ovog metoda je stopa konvergencije koja je najviše linearna. Prema tome, gradijentni metod može biti veoma spor i može zahtevati mnogo iteracija kako bi se pronašlo rešenje sa dobrom tačnošću.

Ukoliko pretpostavimo da je funkcija dva puta neprekidno diferencijabilna možemo dobiti sofisticiranije izbore za pravac pretrage p_k. Funkcija cilja $f \in C^2(\mathbb{R}^n)$ može biti aproksimirana u okolini trenutne iteracije x_k korišćenjem drugog reda Tejlorovog razvoja:

$$f(x_k + p) \approx f(x_k) + p^T \nabla f(x_k) + \frac{1}{2} p^T \nabla^2 f(x_k) p. \tag{2.1}$$

Ako označimo desnu stranu (2.1) sa $m_k(p)$, cilj je izračunati pravac p u iteraciji k minimizirajući kvadratnu funkciju m_k. Ako pretpostavimo da je $\nabla^2 f(x_k) \succ 0$, tada funkcija m_k ima jedinstveni minimizator oblika

$$p_k = -\left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k).$$

Drugim rečima, za $B_k = \nabla^2 f(x_k)$ dobija se *Njutnov metod*. Ovaj metod postiže lokalnu kvadratnu konvergenciju pod odgovarajućim pretpostavkama o funkciji.

Teorema 2.3. *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ dva puta neprekidno diferencijabilna funkcija i neka je Hesijan $\nabla^2 f(x)$ Lipšic neprekidan u okolini rešenja x^* gde su zadovoljeni dovoljni uslovi optimalnosti. Neka je $\{x_k\}$ niz generisan Njutnovim metodom pri čemu je u svakoj iteraciji linijskog pretraživanja (1.2) $\alpha = 1$. Tada važi:*

i) Ako je početna tačka x_0 dovoljno blizu x^*, niz iteracija $\{x_k\}$ konvergira ka x^*;

ii) Ako metod konvergira, stopa konvergencije niza $\{x_k\}$ je kvadratna;

iii) Ako metod konvergira, niz normi gradijenata $||\nabla f(x_k)||$ kvadratno konvergira ka nuli.

Kao što je već pomenuto, da bi se obezbedilo da je p_k opadajući pravac, uslov iz Leme 2.1 treba da bude zadovoljen. Tačnije, matrica $\nabla^2 f(x_k)$ mora biti pozitivno definitna u svakoj iteraciji.

Važno je istaći da uprkos dobrim teoretskim osobinama, u praksi se pojavljuju i izvesni nedostaci. Svaka iteracija Njutnovog metoda zahteva određivanje matrice Hesijana, što podrazumeva izračunavanje izvoda drugog reda u svakoj iteraciji, što samo po sebi može biti izuzetno skupo. Metod, takođe, čak i ako je $\nabla^2 f(x_k) \succ 0$, može postati nestabilan u slučaju loše uslovljenog $\nabla^2 f(x_k)$ u nekoj iteraciji.

Zbog ovih nedostataka, pojavile su se brojne varijante i modifikacije Njutnovog metoda koji imitiraju Njutnovu ideju i u čijoj osnovi je aproksimacija matrice Hesijana. U ovim metodama, nazvanim *metode Njutnovog tipa* ili *kvazi-Njutnove metode (QN)*, matrica Hesijana je zamenjena matricom $B_k \in \mathbb{R}^{n\times n}$, takvom da je

$$B_k \approx \nabla^2 f(x_k)$$

dobra aproksimacija prave matrice Hesijana sa nižim troškovima izračunavanja jer je B_k bazirana samo na informacijama prvog reda. Red konvergencije u ovim metodama je najviše superlinearan, pa je konvergencija sporija u poređenju sa Njutnovim metodom, ali s druge strane, trošak je značajno manji. Ideja na kojoj se zasnivaju QN metode je da dve uzastopne iteracije x_k i x_{k+1} zajedno sa gradijentima $\nabla f_k := \nabla f(x_k)$ i $\nabla f_{k+1} := \nabla f(x_{k+1})$ sadrže informacije o krivini (tj. Hesijanu).

U literaturi su predložene različite strategije za izračunavanje matrice B_k zasnovane na uslovima koje bi B_{k+1} trebalo da zadovolji. Glavni uslov poznat je kao jednačina sečice

$$B_{k+1}s_k = y_k, \tag{2.2}$$

gde je razlika između dve iteracije i između gradijenata u dve susedne iteracije data sa

$$s_k = x_{k+1} - x_k \text{ and } y_k = \nabla f_{k+1} - \nabla f_k.$$

Međutim, jedinstveno rešenje za B_{k+1} nije obezbeđeno iz uslova (2.2). Zato se nameću dodatni zahtevi na B_{k+1}, kao što su simetrčnost i ograničenje da razlika između uzastopnih aproksimacija B_k i B_{k+1} ima niski rang. Drugim rečima, B_{k+1} je rešenje sledećeg problema

$$\min ||B - B_k||_* \tag{2.3}$$

$$\text{tako da } B^T = B, Bs_k = y_k.$$

Različite formule za ažuriranje matrice B_{k+1} dobijaju se rešavanjem problema (2.3) u zavisnosti od matrične norme $||\cdot||_*$, [7].

Jedna od najčešće korišćenih formula za ažuriranje ovog tipa je DFP formula predložena od strane Dejvidsona, Flečera i Pauela. Ona se dobija korišćenjem ponderisane Frobeniusove norme i definisana je sa

$$B_{k+1} = (I - \frac{y_k s_k^T}{y_k^T s_k})B_k(I - \frac{s_k y_k^T}{y_k^T s_k}) + \frac{y_k y_k^T}{y_k^T s_k}.$$

Zatim, Brojden, Flečer, Goldfarb i Šenoa predložili su BFGS formulu oblika

$$H_{k+1} = (I - \frac{s_k y_k^T}{y_k^T s_k})H_k(I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k},$$

gde H_{k+1} predstavlja aproksimaciju inverzne matrice Hesijana, tj. $H_k = B_k^{-1}$ $(H_{k+1}y_k = s_k)$. Početna aproksimacija H_0 se bira od strane korisnika i često je definisana kao $H_0 = \gamma I, \gamma > 0$. Bitno je istaći da BFGS formula očuvava pozitivnu definitnost, tačnije, ako je H_k pozitivno definitna i $y_k^T s_k > 0$, tada je i H_{k+1} takođe pozitivno definitna.

2.1. Spektralni gradijentni metod

Sada ćemo predstaviti modifikovanu verziju QN metoda - *Spektralni gradijentni (SG) metod*. Ovaj metod je poznat po svojoj efikasnosti i jednostavnosti u rešavanju optimizacionih problema [1, 4, 5, 6, 8]. Originalno je predložen od strane Barzilaija i Borweina [2], pa se često naziva BB metodom. Strategija odabira dužine koraka u SG metodu je ključna za bržu konvergenciju u poređenju sa klasičnim gradijentnim metodama, jer inkorporira informacije drugog reda vezane za spektar matrice Hesijana. Tačnije, oslanja se na jednostavnu aproksimaciju drugih izvoda, koja ima oblik dijagonalne matrice pomnožene sa takozvanim spektralnim koeficijentom. Grubo govoreći, ovaj koeficijent aproksimira karakteristični koren matrice Hesijana i pruža barem neku vrstu informacije drugog reda, što je ključno za brzu konvergenciju. Iako teorijski rezultati nisu tako snažni kao kod Njutnovih metoda koji se oslanjaju na prave druge izvode ili bolje aproksimacije matrice Hesijana, spektralni gradijentni metodi su jeftini, laki za implementaciju i pružaju vrlo dobre numeričke rezultate, što ih čini popularnim u praksi.

Spektralni koeficijent je konstruisan tako da najbolje odgovara jednačini sečice, gde je jedan od ključnih delova razlika između dva uzastopna gradijenta- označena sa y_k. Stoga, cilj je pronaći dijagonalnu matricu D_k posebnog oblika

$$D_k = \gamma_k I, \quad \gamma_k \in \mathbb{R},$$

koja najbolje odgovara jednačini $H_{k+1} y_k = s_k$, pri čemu matrica $D_k \approx H_{k+1}$ predstavlja aproksimaciju inverzne matrice Hesijana $(\nabla^2 f(x_k))^{-1}$. Dakle, pravac pretrage je paralelan pravcu negativnog gradijenta, tj. važi

$$p_k = -\gamma_k I \nabla f(x_k) = -\gamma_k \nabla f(x_k).$$

Spektralni koeficijent γ_k je definisan uslovom iz jednačine sečice na sledeći način

$$\tilde{\gamma}_k = \arg\min_{\gamma \in \mathbb{R}} ||y_{k-1} - \gamma s_{k-1}||^2 \tag{2.4}$$

ili

$$\gamma_k = \arg\min_{\gamma \in \mathbb{R}} ||\gamma y_{k-1} - s_{k-1}||^2, \tag{2.5}$$

gde je $\gamma_k = \tilde{\gamma}_k^{-1}$. Iz (2.4) i (2.5) dobijamo spektralne koeficijente oblika

$$\gamma_k^{BB1} = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \quad \text{i} \quad \gamma_k^{BB2} = \frac{s_{k-1}^T y_{k-1}}{y_{k-1}^T y_{k-1}}. \tag{2.6}$$

Pored ova dva pravila za izračunavanje γ_k, u literaturi su predloženi brojni spektralni gradijentni metodi koji uopštavaju BB metode, kao na primer Adaptivni Barzilai–Borwein (ABB) [9] i njegova modifikacija ABBmin [3], koji se oslanjaju na adaptivne kriterijume korišćene za prelazak između γ_k^{BB1} i γ_k^{BB2}. Dužine koraka se definišu sledećim pravilima

$$\gamma_k^{ABB} := \begin{cases} \gamma_k^{BB2}, & \frac{\gamma_k^{BB2}}{\gamma_k^{BB1}} < \tau, \\ \gamma_k^{BB1}, & \text{inače,} \end{cases}$$

i

$$\gamma_k^{ABBmin} := \begin{cases} \min\{\gamma_j^{BB2} : j = \max\{1, k - m_a\}, ..., k\}, & \frac{\gamma_k^{BB2}}{\gamma_k^{BB1}} < \tau, \\ \gamma_k^{BB1}, & \text{inače}, \end{cases}$$

gde je $m_a \geq 0$ i $\tau \in (0, 1)$.

U slučaju kada uslov krivine $s_{k-1}^T y_{k-1} > 0$ nije zadovoljen, γ_k može biti negativno, tako da pravac pretraživanja nije opadajući pravac. Ova mana može biti prevaziđena korišćenjem safeguard-a [10] na sledeći način

$$\bar{\gamma}_k = \min\{\gamma_{max}, \max\{\gamma_k, \gamma_{min}\}\},$$

gde $0 < \gamma_{min} << 1 << \gamma_{max} < \infty$. Stavljajući da je $p_k = -\bar{\gamma}_k \nabla f(x_k)$ obezbeđeno je da je pravac opadajući i numerička stabilnost se može kontrolisati.

Literatura

[1] S. Bellavia, N. Krklec Jerinkić and G. Malaspina, "Subsampled nonmonotone spectral gradient methods", *Communications in Applied and Industrial Mathematics*, 11(1), pp. 19-34, 2020.

[2] J. Barzilai and J. M. Borwein, "Two-point step size gradient method", *IMA J. Numerical Analysis*, 8(1), pp. 141–148, 1988.

[3] G. Frassoldati , L. Zanni , G. Zanghirati, "New adaptive stepsize selections in gradient methods", *J. Ind. Manag. Optim.* 4 (2), pp. 299–312, 2008.

[4] N. Krejić and N. Krklec Jerinkić, "Spectral projected gradient method for stochastic optimization", *Journal of Global Optimization*, 73, pp. 59–81, 2018.

[5] N. Krejić and N. Krklec Jerinkić, T. Ostojić, "Spectral projected subgradient method for nonsmooth convex optimization problems", *Numerical Algorithms*, pp. 1-19, 2022.

[6] N. Krklec Jerinkić and T. Ostojić, "AN-SPS: Adaptive Sample Size Nonmonotone Line Search Spectral Projected Subgradient Method for Convex Constrained Optimization Problems", *Optimization Methods and Software*, 2024, DOI - 10.1080/10556788.2024.2324920.

[7] J. Nocedal and S. J. Wright, *Numerical Optimization, 2nd ed.*, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2006.

[8] C. Tan, S. Ma, Y. H. Dai and Y. Qian, "Barzilai-borwein step size for stochastic gradient descent", *Advances in neural information processing systems*, 29, 2016.

[9] B. Zhou, L. Gao and Y.H. Dai, "Gradient methods with adaptive stepsizes", *Comput. Optim. Appl.* 35 (1), pp.69–86, 2006.

[10] R. Tavakoli and H. Zhang, "A nonmonotone spectral projected gradient method for large-scale topology optimization problems", *Numerical Algebra, Control and Optimization* vol. 2, No. 2, pp. 395-412, 2012.