UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU
I INFORMATIKU

Snežana S. Đorđević

Izbor parametara kod gradijentnih metoda za probleme optimizacije bez ograničenja

\- Doktorska disertacija –

Novi Sad, 2014

Izbor parametara kod gradijentnih metoda za probleme optimizacije bez ograničenja

Snežana S. Đorđević

December 24, 2014

Veliku zahvalnost dugujem svojoj porodici za neprekidnu i neizmernu podršku i ogromno strpljenje.

Posebnu zahvalnost dugujem mom mentoru prof. dr Nataši Krejić za veliko strpljenje i pomoć u izradi ove doktorske disertacije.

Leskovac, Decembar 2014. godine

Snežana S. Đorđević

Sadržaj

Glava 1

Uvod

1.1 Motivacija

Istraživanje u ovoj doktorskoj disertaciji motivisano je potrebom za metodama koje će brzo konvergirati, bez prisustva matrice hesijana, što je naročito aktuelno za sisteme velikih dimenzija i za probleme optimizacije u prisustvu šuma, kada ne raspolažemo ni tačnom vrednošću funkcije cilja, ni tačnom vrednošću gradijenta. Deo motivacije za istraživanjem ovde leži i u postojanju problema kod kojih je funkcija cilja rezultat simulacija.

Optimizacija predstavlja postupak nalaženja najboljeg rešenja nekog problema u određenom smislu i pod određenim uslovima.

Ona je važan alat u teoriji odlučivanja i analizi fizičkih sistema.

Teorija optimizacije je veoma razvijena oblast. Ona ima široku primenu u nauci, inženjerstvu, poslovnom upravljanju, vojnoj i kosmičkoj tehnologiji. Iako optimizacija datira još od prvih problema pronalaženja ekstremuma, postaje samostalna oblast matematike tek od 1940. godine, kada je G.B. Dantzig predstavio dobro poznat simpleks algoritam za linearno programiranje.

Nakon 1950. godine, kada su metod konjugovanih gradijenata i kvazi-Njutnov metod predstavljeni, nelinearno programiranje se ubrzano razvija. Danas se različiti moderni metodi optimizacije mogu primeniti za rešavanje širokog spektra problema optimizacije i predstavljaju neophodno sredstvo za rešavanje problema u različitim oblastima.

Najpre je potrebno definisati funkciju cilja, koja može biti, na primer, tehnički rashod, dobit ili čistoća materijala, profit, vreme, potencijalna energija. Funkcija cilja zavisi od izvesnih karakteristika sistema, koje se zovu promenljive ili nepoznate. Cilj je naći vrednosti tih promenljivih, za koje funkcija cilja dostiže svoju najbolju vrednost, koju zovemo ekstremum ili optimum.

Može se desiti da se pomenute promenljive biraju tako da zadovoljavaju izvesne uslove, odnosno ograničenja.

Proces identifikovanja funkcije cilja, promenljivih i ograničenja za dati problem zove se modelovanje.

Prvi i najvažniji korak u procesu optimizacije je konstrukcija odgovarajućeg modela. To može biti problem sam po sebi. Ukoliko se formira previše uprošćen model, on ne može biti veran odraz praktičnog problema. Na drugoj strani, ukoliko je konstruisan previše složen model, rešavanje problema će takođe biti previše složeno.

Nakon konstrukcije odgovarajućeg modela, potrebno je primeniti odgovarajući algoritam za rešavanje problema. Pri tom, ne postoji univerzalni algoritam za rešavanje postavljenog problema.

U primenama, skup ulaznih parametara je ograničen, to jest, ulazni parametri se menjaju unutar dozvoljenog prostora ulaznih parametara D_x; možemo pisati

$$x \in D_x. \tag{1.1.1}$$

Osim (1.1.1), mogu se nametnuti i uslovi oblika:

$$\varphi_l(x_1, \ldots, x_n) = \varphi_{0l},\ l = 1, \ldots, m_1 < n, \tag{1.1.2}$$

$$\psi_j(x_1, \ldots, x_n) \leq \psi_{0j},\ j = 1, \ldots, m_2. \tag{1.1.3}$$

Zadatak optimizacije jeste naći minimum (maksimum) funkcije cilja $f(x) = f(x_1, \ldots, x_n)$, pod uslovima (1.1.1), (1.1.2), (1.1.3).

Ako je funkcija cilja linearna i ako su funkcije $\varphi_l(x_1, \ldots, x_n)$, $l = 1, \ldots, m_1$, $\psi_j(x_1, \ldots, x_n)$, $j = 1, \ldots, m_2$, linearne, tada se radi o problemu linearnog programiranja; ako je bar jedna od pomenutih funkcija nelinearna, radi se o problemu nelinearnog programiranja.

Ovde će se izučavati problemi optimizacije koji ne sadrže ograničenja. Osim toga, posmatraće se problemi sa nelinearnom funkcijom cilja, što pripada oblasti nelinearnog programiranja.

Opšti oblik problema optimizacije bez ograničenja je

$$\min_{x \in \mathbb{R}^n} f(x), \tag{1.1.4}$$

gde je $x \in \mathbb{R}^n$ ulazni parametar, $f(x)$ funkcija cilja.

Algoritam za nalaženje minimuma može se iskoristiti za nalaženje maksimuma:

$$\min f(x) = -\max(-f(x)).$$

Važi i obratno:

$$\max f(x) = -\min(-f(x)).$$

Ova doktorska disertacija ima za cilj sistematizaciju poznatih rezultata, kao i teorijsku i numeričku analizu mogućnosti uvođenja parametra u gradijentne metode. Izbor parametra vršimo na slučajan način iz intervala $[\mu, 1)$, za $0 < \mu < 1$.

Takođe, formiramo jedan novi hibridni metod konjugovanog gradijenta, kod koga je parametar konjugovanog gradijenta β_k dobijen kao konveksna kombinacija poznatih parametara β_k^{LS} i β_k^{CD}.

Struktura ove disertacije je kao što sledi.

Najpre je dat pregled oznaka, teorema i definicija, korišćenih u disertaciji. Zatim je opisan osnovni problem nelinearne optimizacije bez ograničenja i predstavljeno je linijsko pretraživanje, kao i opis različitih linijskih pretraživanja.

U Glavi 2 opisani su gradijentni metodi kako prvog, tako i drugog reda.

U Glavi 3 dati su originalni rezultati. U ovoj glavi predstavljena je modifikacija metoda multiplikativnih parametara, datog u radu [113]. Modifikacija pomenutog metoda multiplikativnih parametara se vrši korišćenjem parametra koji se bira na slučajan način unutar intervala $[\mu, 1)$, za $0 < \mu < 1$.

Glava 4 sadrži pregled osnovnih pojmova i rezultata vezanih za metod konjugovanih gradijenata.

Glava 5 sadrži originalne rezultate koji se odnose na metod konjugovanih gradijenata. U ovoj glavi predstavljen je novi hibridni metod konjugovanih gradijenata, dobijen kombinovanjem dva poznata metoda konjugovanih gradijenata.

U Glavi 6 dati su rezultati numeričkih eksperimenata koji se odnose na metode iz Glave 3 i Glave 5.

1.2 Pregled oznaka, teorema i definicija

Uvodimo osnovne oznake koje su korišćene u ovoj disertaciji.

Skup prirodnih brojeva označava se sa $\mathbb{N}$, $\mathbb{R}$ označava skup realnih brojeva, $\mathbb{R}^n$ označava n-dimenzionalan prostor realnih brojeva a $\mathbb{R}^{m\times n}$ označava prostor realnih matrica sa m vrsta i n kolona. Vektor $x \in \mathbb{R}^n$ predstavlja se kao $x = (x_1, x_2, \dots, x_n)^T$
.

Ako je $x \in \mathbb{R}^n$, x^T označava transponovani vektor vektora x.

Potprostor prostora $\mathbb{R}^n$, definisan vektorima $v_1, v_2, \dots, v_k$, označavaćemo
$span\{v_1, v_2, \dots, v_k\}$.

Definicija 1.2.1. *Preslikavanje* $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ *zove se norma vektora ako i samo ako zadovoljava sledeće osobine:*

(i) $\|x\| \geq 0,\ x \in \mathbb{R}^n$;

(ii) $\|x\| = 0 \Leftrightarrow x = 0$;

(ii) $\|cx\| = |c|\|x\|,\ c \in \mathbb{R},\ x \in \mathbb{R}^n$;

(iii) $\|x + y\| \leq \|x\| + \|y\|,\ x, y \in \mathbb{R}^n$.

Definicija 1.2.2. l_p*-norma definiše se kao*

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}.$$

Specijalni slučajevi l_p*-norme su:*

$$\begin{aligned} \|x\|_\infty &= \max_{1\leq i\leq n} |x_i|, && (l_\infty\text{-}norma),\\ \|x\|_1 &= \sum_{i=1}^{n} |x_i|, && (l_1\text{-}norma),\\ \|x\|_2 &= \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}}, && (l_2\text{-}norma). \end{aligned}$$

Definicija 1.2.3. *Skalarni proizvod vektora* a *i* b *definiše se kao*
$< a, b >= a^T b = \|a\| \cdot \|b\| \cdot \cos\sphericalangle(a, b)$.

Definicija 1.2.4. *Sve pravce koji zadovoljavaju*

$$\nabla f(x_k)^T d_k < 0, \tag{1.2.1}$$

zovemo opadajući pravci za funkciju f *u tački* x_k.

Definicija 1.2.5. *Ako postoji konstanta $c > 0$, takva da važi*

$$\nabla f(x_k)^T d_k \leq -c\|g_k\|^2, \tag{1.2.2}$$

tada kažemo da pravac traženja d_k zadovoljava uslov dovoljnog pada.

Definicija 1.2.6. *Niz vektora $\{x_k\} \subset \mathbb{R}^n$ konvergira ka x^*, ako je*

$$\lim_{k\to\infty} \|x_k - x^*\| = 0.$$

Definicija 1.2.7. *Niz $\{x_k\} \subset \mathbb{R}^n$ zove se Košijev niz, ako je*

$$\lim_{m,l\to\infty} \|x_m - x_l\| = 0,$$

to jest, za dato $\epsilon > 0$, postoji ceo broj N, takav da važi $\|x_m - x_l\| < \epsilon$ za svako $m, l > N$.

Definicija 1.2.8. *Otvoren interval, u oznaci (a, b), definiše se kao*
$(a, b) = \{x | a < x < b\}$.

Definicija 1.2.9. *Zatvoren interval ili segment, u oznaci $[a, b]$, definiše se kao $[a, b] = \{x | a \leq x \leq b\}$.*

Definicija 1.2.10. *Neka iterativni niz $\{x_k\}$ konvergira ka x^* u nekoj normi.*

Ako postoji pozitivna konstanta $\alpha \geq 1$ i pozitivna konstanta β nezavisna od broja k, tako da važi

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

tada kažemo da $\{x_k\}$ ima red konvergencije jednak α.

Specijalno:

1. *kada je $\alpha = 1$ i $\beta \in (0, 1)$, kaže se da niz $\{x_k\}$ konvergira linearno;*
2. *kada je $\alpha = 1$ i $\beta = 0$, ili $1 < \alpha < 2$ i $\beta > 0$, tada se kaže da niz $\{x_k\}$ konvergira superlinearno;*
3. *kada je $\alpha = 2$, kaže se da niz $\{x_k\}$ konvergira kvadratno.*

Definicija 1.2.11. *Matrica čije su kolone jednake odgovarajućim vrstama matrice A, naziva se transponovana matrica matrice A i označava se A^T.*

Definicija 1.2.12. *Kvadratna matrica A naziva se simetrična matrica ako je $A^T = A$.*

Definicija 1.2.13. *Simetrična matrica A reda n je pozitivno definitna ako je $x^T Ax > 0$ za svaki nenula vektor x.*

Simetrična matrica A reda n je pozitivno semidefinitna ako je $x^T Ax \geq 0$ za svaki nenula vektor x.

Definicija 1.2.14. *Simetrična matrica A reda n je negativno definitna ako je $x^T Ax < 0$ za svaki nenula vektor x.*

Simetrična matrica A reda n je negativno semidefinitna ako je $x^T Ax \leq 0$ za svaki nenula vektor x.

Definicija 1.2.15. *Preslikavanje $\|\cdot\| : \mathbb{R}^{m\times n} \to \mathbb{R}$ zove se matrična norma ako i samo ako zadovoljava sledeće osobine:*

$$\begin{aligned}
&(i)\ \|A\| \geq 0,\ \ A \in \mathbb{R}^{m\times n},\\
&(ii)\ \|A\| = 0 \Leftrightarrow A = 0,\\
&(iii)\ \|\alpha A\| = |\alpha|\|A\|,\ \ \alpha \in \mathbb{R},\ \ A \in \mathbb{R}^{m\times n},\\
&(iv)\ \|A + B\| \leq \|A\| + \|B\|,\ \ A, B \in \mathbb{R}^{m\times n}.
\end{aligned}$$

Specijalni slučajevi matrične norme su:

$$\begin{aligned}
\|A\|_p &= \sup_{x\neq 0} \frac{\|Ax\|_p}{\|x\|_p},\\
\|A\|_1 &= \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|.\\
\|A\|_\infty &= \max_{1\leq i\leq m} \sum_{i=1}^{m} |a_{ij}|.
\end{aligned}$$

Definicija 1.2.16. *Kvadratna matrica A zove se regularna ako je $detA \neq 0$. Ako je $detA = 0$, kvadratna matrica A zove se singularna.*

Definicija 1.2.17. *Niz matrica $\{A_k\}$ konvergira ka matrici A, ako je*

$$\lim_{k\to\infty} \|A_k - A\| = 0.$$

Definicija 1.2.18. *Ako važi $Ax = \lambda x$, $A \in \mathbb{R}^{n\times n}$, $x \neq 0$, $x \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$, tada je x sopstveni vektor matrice A koji odgovara sopstvenoj vrednosti λ.*

Definicija 1.2.19. *Uslovni broj matrice A definiše se kao $\kappa(A) = \|A\|\|A^{-1}\|$.*

Definicija 1.2.20. *Spektralni radijus matrice A, čije su sopstvene vrednosti λ_i, $i = 1, \ldots, n$, definiše se kao*

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|.$$

Specijalno, ako je matrica $A \in \mathbb{R}^{n \times n}$ simetrična, sa sopstvenim vrednostima $\lambda_1, \ldots, \lambda_n$, tada je

$$\|A\|_2 = \max_{1 \leq i \leq n} |\lambda_i|.$$

Ako je matrica A simetrična, tada je uslovni broj matrice A

$$\kappa(A) = \frac{\max_i |\lambda_i|}{\min_i |\lambda_i|}.$$

Definicija 1.2.21. *Neka je G $n \times n$ simetrična i pozitivno definitna matrica, d_1, d_2, $\ldots$ $d_m \in \mathbb{R}^n$ nenula vektori, $m \leq n$. Ako važi*

$$d_i^T G d_j = 0, \ \textit{za svako } i, j, \ i \neq j,$$

tada se vektori d_1, d_2, $\ldots$ d_m zovu G -konjugovani vektori.

Definicija 1.2.22. *Neka je X proizvoljan neprazan skup i $d : X \times X \to \mathbb{R}$ preslikavanje takvo da za svako $x, y, z \in X$ važi*

$$\begin{aligned} &(1)\ d(x,y) \geq 0;\\ &(2)\ d(x,y) = 0 \Leftrightarrow x = y;\\ &(3)\ d(x,y) = d(y,x);\\ &(4)\ d(x,y) \leq d(x,z) + d(z,y). \end{aligned}$$

Tada je d metrika, ili rastojanje na skupu X, a uređen par (X, d) jeste metrički prostor.

Specijalni slučajevi metričkih prostora su:

$$\begin{aligned} &(i)\ (\mathbb{R}, d), \ \textit{gde je } d(x,y) = |x - y|;\\ &(ii)\ (\mathbb{R}^2, d), \ \textit{gde je } d(x,y) = \|x - y\|_2\\ &(iii)\ (\mathbb{R}^n, d), \ \textit{gde je } d(x,y) = \|x - y\|_2. \end{aligned}$$

Definicija 1.2.23. *Neka je (X, d) metrički prostor i $A \subset X$. Dijametar skupa A, u oznaci $diam(A)$, definiše se kao:*

$$diam(A) = \sup\{d(P, P') | P, P' \in A\}.$$

Definicija 1.2.24. *Skup $D \subset \mathbb{R}^n$ je kompakt ako je ograničen i zatvoren skup.*

Svaki niz $\{x_k\} \subset D$ ima konvergentan podniz čija je granična vrednost element skupa D.

Definicija 1.2.25. *Neka je skup $S \subset \mathbb{R}^n$. Ako za ma koje $x_1, x_2 \in S$ važi*

$$\alpha x_1 + (1-\alpha)x_2 \in S, \text{ za svako } \alpha \in [0,1],$$

tada je S konveksan skup.

Definicija 1.2.26. *Funkcija $f : \mathbb{R}^n \to \mathbb{R}$ je neprekidna u tački $\bar{x} \in \mathbb{R}^n$ ako, za ma koje dato $\epsilon > 0$, postoji $\delta > 0$ tako da $\|x - \bar{x}\| < \delta$ implicira $|f(x) - f(\bar{x})| < \epsilon$.*

Ako je f neprekidna u svakoj tački otvorenog skupa $D \subset \mathbb{R}^n$, tada se kaže da je f neprekidna na skupu D.

Definicija 1.2.27. *Funkcija $f : \mathbb{R}^n \to \mathbb{R}$ je neprekidno diferencijabilna u tački $x \in \mathbb{R}^n$ ako postoje parcijalni izvodi $\dfrac{\partial f}{\partial x_i}(x)$ i ako su neprekidni, za $i = 1, \ldots, n$.*

Ako je f neprekidno diferencijabilna u svakoj tački otvorenog skupa $D \subset \mathbb{R}^n$, tada je f neprekidno diferencijabilna na skupu D, u oznaci $f \in C^1(D)$.

Definicija 1.2.28. *Gradijent funkcije f u tački x definiše se kao*

$$\nabla f(x) = \left[\frac{\partial f}{\partial x_i}(x), \ldots, \frac{\partial f}{\partial x_i}(x)\right]^T.$$

Uobičajene oznake su: $g(x) = \nabla f(x)$, $g_k = g(x_k)$.

Definicija 1.2.29. *Tačka x^* je stacionarna tačka funkcije f ako je $\nabla f(x^*) = 0$.*

Definicija 1.2.30. *Funkcija $f : \mathbb{R}^n \to \mathbb{R}$ je dva puta neprekidno diferencijabilna u tački $x \in \mathbb{R}^n$ ako postoje $\dfrac{\partial^2 f}{\partial x_i \partial x_j}(x)$ i ako su neprekidni, $i = 1, \ldots, n$, $j = 1, \ldots, n$.*

Ako je funkcija f dva puta neprekidno diferencijabilna u svakoj tački otvorenog skupa $D \subset \mathbb{R}^n$, tada se kaže da je f dva puta neprekidno diferencijabilna na D i označava se $f \in C^2(D)$.

Definicija 1.2.31. *Hesijan funkcije f u tački x definiše se kao $n \times n$ matrica sa elementima $[\nabla^2 f(x)]_{ij} = \dfrac{\partial^2 f}{\partial x_i \partial x_j}(x)$, $1 \le i, j \le n$. Uobičajene oznake su:*

$G(x) = \nabla^2 f(x)$, $G_k = G(x_k)$.

Definicija 1.2.32. *Funkcija $f : \mathbb{R}^n \to \mathbb{R}$ je glatka ako postoje parcijalni izvodi proizvoljnog reda funkcije f i pri tom su neprekidni.*

Definicija 1.2.33. *Neka je skup $S \subset \mathbb{R}^n$ neprazan konveksan skup. Neka je $f : S \subset \mathbb{R}^n \to \mathbb{R}$.*

Ako za ma koje $x_1, x_2 \in S$ i svako $\alpha \in [0, 1]$ važi

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2),$$

tada je f konveksna funkcija na skupu S.

Ako za ma koje $x_1, x_2 \in S$, $x_1 \neq x_2$ i svako $\alpha \in [0, 1]$ važi

$$f(\alpha x_1 + (1-\alpha)x_2) < \alpha f(x_1) + (1-\alpha)f(x_2),$$

tada je f striktno konveksna funkcija na skupu S.

Ako postoji konstanta $c > 0$, takva da za ma koje x_1, $x_2 \in S$,

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2) - \frac{1}{2}c\alpha(1-\alpha)\|x_1 - x_2\|^2,$$

tada je f uniformno (ili jako) konveksna funkcija na S.

Funkcija $f : S \subset \mathbb{R}^n \to \mathbb{R}$ naziva se konkavna (striktno konkavna, uniformno konkavna) funkcija na S ako je $-f$ konveksna (striktno konveksna, uniformno konveksna) funkcija na S.

Definicija 1.2.34. *Funkcija $f : \mathbb{R}^n \to \mathbb{R}$ je Lipšic neprekidna funkcija ako postoji konstanta $L < \infty$ takva da je*

$$\|f(x) - f(y)\| \leq L\|x - y\|, \text{ za svako } x,\ y \in \mathbb{R}^n.$$

Teorema 1.2.1. *Niz $\{x_k\} \subseteq \mathbb{R}^n$ konvergira ako i samo ako je $\{x_k\}$ Košijev niz.*

Teorema 1.2.2. *(Lagranžova teorema o srednjoj vrednosti) Neka je funkcija $f(x)$:*
(1) *neprekidna na segmentu $[a, b]$;*
(2) *ima izvod u intervalu (a, b).*
Tada postoji $\xi \in (a, b)$, takva da važi $f'(\xi) = \dfrac{f(b) - f(a)}{(b-a)}$.

Teorema 1.2.3. *(Koši-Švarcova nejednakost) Neka su x, $y \in \mathbb{R}^n$. Tada važi*

$$|x^T y| \leq \|x\|_2 \|y\|_2. \tag{1.2.3}$$

U izrazu (1.2.3) jednakost važi ako i samo ako su x i y linearno zavisni.

Teorema 1.2.4. *Neka je funkcija* $f : \mathbb{R}^n \to \mathbb{R}$ *glatka na intervalu* $(x_0 - a, x_0 + a)$. *Ako je* $x \in (x_0 - a, x_0 + a)$, *tada važi Tejlorova formula:*

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_n(x),$$

gde je

$$R_n(x) = \frac{f^{(n)}(x_0 + \theta(x - x_0))}{n!}(x - x_0)^n, \ 0 < \theta < 1.$$

Ako je $\lim_{n\to\infty} R_n(x) = 0$ *za svako* $x \in (x_0 - a, x_0 + a)$, *tada je*

$$f(x) = \lim_{n\to\infty} \sum_{k=0}^{n-1} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k. \tag{1.2.4}$$

Red (1.2.4) je Tejlorov red funkcije $f(x)$ *na intervalu* $(x_0 - a, x_0 + a)$.

Teorema 1.2.5. *(Potreban uslov optimalnosti prvog reda) Neka je funkcija* $f(x)$ *diferencijabilna u tački* x^*. *Ako je* x^* *lokalni minimum, važi* $g(x^*) = 0$.

Teorema 1.2.6. *(Potreban uslov optimalnosti drugog reda) Neka je funkcija* $f(x)$ *dva puta neprekidno diferencijabilna u tački* x^*. *Ako je* x^* *lokalni minimum, tada je* $g(x^*) = 0$ *i* $G(x^*)$ *pozitivno semidefinitna matrica.*

Teorema 1.2.7. *(Dovoljan uslov optimalnosti) Neka je funkcija* $f(x)$ *dva puta neprekidno diferencijabilna u tački* x^*. *Ako je* $g(x^*) = 0$ *i* $G(x^*)$ *pozitivno definitna matrica, tada je* x^* *lokalni minimum. Ako je* $g(x^*) = 0$ *i* $G(x^*)$ *negativno definitna matrica, tada je* x^* *lokalni maksimum.*

Teorema 1.2.8. *(Sherman, Morrison) [116] Neka je* $A \in \mathbb{R}^{n\times n}$ *regularna matrica,* $u, v \in \mathbb{R}^n$ *proizvoljni vektori. Ako važi*

$$1 + v^T A^{-1} u \neq 0,$$

onda je ažuriranje ranga 1, $A + uv^T$, *matrice* A *regularna matrica i važi*

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1}uv^T A^{-1}}{1 + v^T A^{-1} u}. \tag{1.2.5}$$

Pretpostavka P_1 (Lipšicova pretpostavka): U nekoj okolini $\mathcal{N}$ nivo skupa

$$\mathcal{L} = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\},$$

gradijent $g(x)$ je Lipšic neprekidna funkcija.

Pretpostavka P_2 (Pretpostavka o ograničenosti): Nivo skup $\mathcal{L}$ je ograničen. To znači, postoji konstanta $B < \infty$, takva da je

$$\|x\| \leq B, \text{ za svako } x \in \mathcal{L}.$$

1.3 Linijsko pretraživanje

Razmatramo problem nelinearne optimizacije bez ograničenja

$$\min\{f(x) : x \in \mathbb{R}^n\}, \tag{1.3.1}$$

gde je $f : \mathbb{R}^n \to \mathbb{R}$ neprekidno diferencijabilna funkcija, ograničena sa donje strane. Postoji veliki broj različitih metoda za rešavanje problema (1.3.1).

U metodama optimizacije zasnovanim na linijskom pretraživanju, za dato x_k, opšta iterativna šema je data izrazom:

$$x_{k+1} = x_k + t_k d_k, \tag{1.3.2}$$

gde x_{k+1} predstavlja aproksimaciju tačnog rešenja u novoj iteraciji, d_k je pravac pretraživanja, a t_k je dužina koraka u pravcu d_k.

Najpre razmatramo monotono linijsko traženje.

Sledi opšta iterativna šema metoda monotonog linijskog traženja.

Algoritam 1.3.1. *Ulazni parametri:* $\epsilon > 0$, x_0, $k := 0$.

Korak 1. Ako je $\|g_k\| \leq \epsilon$, *STOP.*

Korak 2. Naći vektor d_k, *koji je pravac pada funkcije cilja.*

Korak 3. Naći veličinu koraka t_k *tako da važi* $f(x_k + t_k d_k) < f(x_k)$.

Korak 4. Postaviti $x_{k+1} = x_k + t_k d_k$.

Korak 5. $k := k + 1$; *ići na Korak 1.*

Uvodimo oznaku:

$$\Phi(t) = f(x_k + td_k).$$

Rešavamo li problem minimizacije funkcije, tada tragamo za veličinom koraka t_k, u pravcu d_k, tako da važi:

$$\Phi(t_k) < \Phi(0).$$

Upravo taj postupak zovemo monotono linijsko traženje.

Dužinu koraka t_k možemo tražiti tako da važi:

$$f(x_k + t_k d_k) = \min_{t>0} f(x_k + td_k), \tag{1.3.3}$$

to jest

$$\Phi(t_k) = \min_{t>0} \Phi(t), \tag{1.3.4}$$

ili možemo koristiti sledeću formulu:

$$t_k = \min\{t | g(x_k + td_k)^T d_k = 0, \ t \geq 0\}. \tag{1.3.5}$$

U ovom slučaju radi se o *tačnom* ili *optimalnom* linijskom traženju (*exact line search*), pri čemu je parametar t_k, dobijen rešavanjem jednodimenzionalnog optimizacionog problema (1.3.4), optimalna veličina koraka.

Umesto da koristimo relaciju (1.3.3) ili relaciju (1.3.5), možemo se zadovoljiti traganjem za onim t_k koje će biti prihvatljivo ukoliko nam odgovara relacija:

$$f(x_k) - f(x_k + t_k d_k) > \delta_k > 0.$$

Tada se radi o *netačnom*, ili *približnom*, ili *prihvatljivom* linijskom traženju (*inexact line search*). U prilog tome da se veoma mnogo koristi približno linijsko traženje ide činjenica da je cena tačnog linijskog traženja visoka. Dalje, kada je iteracija daleko od rešenja problema, tačno linijsko traženje nije efikasno. Takođe, u praksi, red konvergencije mnogih metoda optimizacije, među kojima su, na primer, Njutnov ili kvazi-Njutnov metod, ne zavisi od tačnog linijskog traženja. Zbog toga, sve dok se koristi prihvatljivo pravilo za određivanje dužine koraka, koje obezbeđuje dovoljan pad funkcije cilja, tačno linijsko traženje se može izbeći.

Dalje opisujemo jedno od netačnih linijskih traženja, linijsko traženje unazad (*backtracking*).

Algoritam 1.3.2. *Ulazni parametri:* x_k, *opadajući pravac* d_k, $0 < \delta < \frac{1}{2}$, $\beta \in (0, 1)$.

Korak 1. $t := 1$.

Korak 2. Dok važi $f(x_k + td_k) > f(x_k) + \delta t g_k^T d_k$, $t := t\beta$.

Korak 3. Postaviti $t_k := t$.

Sledi Armižovo (Armijo) pravilo.

Teorema 1.3.1. *[14] Neka je* $f \in C^1(\mathbb{R}^n)$ *i neka je* d_k *opadajući pravac. Tada postoji nenegativan ceo broj* m_k*, takav da važi*

$$f(x_k + \beta^{m_k} d_k) \leq f(x_k) + c_1 \beta^{m_k} g_k^T d_k, \text{ za neku konstantu } c_1 \in (0, 1), \text{ pri čemu je}$$

$\beta \in (0, 1)$.

Slede još neka pravila za određivanje dužine koraka približnim linijskim traženjem, pri čemu pretpostavljamo da je d_k opadajući pravac.

Goldštajnovo (Goldstein) pravilo [65]

Bira se t tako da važi:

$$f(x_k + td_k) \leq f(x_k) + \delta t g_k^T d_k, \tag{1.3.6}$$

$$f(x_k + td_k) \geq f(x_k) + (1 - \delta) t g_k^T d_k, \tag{1.3.7}$$

pri čemu je $0 < \delta < \frac{1}{2}$.

Volfova (Wolfe) pravila [120], [121]

Standardna Volfova pravila su

$$f(x_k + t_k d_k) - f(x_k) \leq \delta t_k g_k^T d_k, \tag{1.3.8}$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k, \tag{1.3.9}$$

gde je d_k opadajući pravac, a $0 < \delta \leq \sigma < 1$. Ova efikasna strategija sastoji se zapravo u prihvatanju pozitivne dužine koraka t_k, ako su zadovoljeni uslovi (1.3.8) i (1.3.9).

Jaki Volfovi uslovi sastoje se od (1.3.8) i sledeće jače verzije uslova (1.3.9):

$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k. \tag{1.3.10}$$

U generalisanim Volfovim uslovima [34], apsolutna vrednost u (1.3.10) zamenjena je parom nejednakosti

$$\sigma_1 g_k^T d_k \leq g_{k+1}^T d_k \leq -\sigma_2 g_k^T d_k, \tag{1.3.11}$$

pri čemu je $0 < \delta < \sigma_1 < 1$ i $\sigma_2 \geq 0$.

Specijalan slučaj $\sigma_1 = \sigma_2 = \sigma$ odgovara jakim Volfovim uslovima.

Aproksimativni Volfovi uslovi [73] su

$$\sigma g_k^T d_k \leq g_{k+1}^T d_k \leq (2\delta - 1) g_k^T d_k, \tag{1.3.12}$$

gde je $0 < \delta < \dfrac{1}{2}$ i $\delta < \sigma < 1$.

Prva nejednakost u (1.3.12) ista je kao u (1.3.9).

Druga nejednakost u (1.3.12) ekvivalentna je sa (1.3.8) kada je f kvadratna funkcija. U opštem slučaju, kada se $\Phi(t) = f(x_k + td_k)$ zameni kvadratnim interpolantom $q(\cdot)$, koji se poklapa sa $\Phi(t)$ za $t = 0$ i $\Phi'(t)$ za $t = 0$ i $t = t_k$, (1.3.8) se redukuje na drugu nejednakost u (1.3.12). Primetimo da aproksimativni Volfovi uslovi imaju isti oblik kao generalisani Volfov uslov (1.3.11), ali sa specijalnim izborom vrednosti σ_2.

Važi sledeća lema.

Lema 1.3.1. *[96] Neka je $f \in C(\mathbb{R}^n)$. Neka je d_k opadajući pravac u tački x_k, i pretpostavimo da je funkcija f ograničena sa donje strane duž pravca $\{x_k + td_k | t > 0\}$. Tada ako je $0 < \delta < \sigma < 1$, postoje intervali unutar kojih dužina koraka zadovoljava Volfove uslove (1.3.8) i (1.3.9) i jake Volfove uslove (1.3.8)-(1.3.10).*

Uvođenje nemonotonog linijskog traženja motivisano je postojanjem problema kod kojih pravac traženja ne mora biti opadajući pravac. To se može dogoditi, na primer, kada izvod funkcije nije dostupan, recimo, to se može desiti u stohastičkoj optimizaciji.

Takođe, neki efikasni kvazi-Njutnovi metodi, kao što je SR1 ažuriranje, ne proizvode opadajući pravac u svakoj iteraciji [96]. Dalje, poznato je da neki vrlo efikasni metodi, poput spektralnog metoda, uopšte ne poseduju monotonost.

Inače, numerički rezultati, dati u [28], [74], [105], [117], pokazuju da su nemonotone tehnike bolje od monotonih kada se radi o nalaženju globalnih optimalnih vrednosti funkcije cilja.

Algoritmi nemonotonog linijskog traženja ne insistiraju na padu funkcije u svakom koraku. Ipak, i ovi algoritmi zahtevaju umanjenje funkcije f nakon nekog unapred određenog broja iteracija.

Prva tehnika nemonotonog linijskog traženja predstavljena je u [67]. Naime, u [67] traga se za veličinom koraka koja zadovoljava

$$f(x_k + t_k d_k) \le \max_{0 \le j \le m(k)} f(x_{k-j}) + \delta t_k g_k^T d_k, \tag{1.3.13}$$

pri čemu je $m(0) = 0$, $0 \le m(k) \le \min\{m(k-1) + 1, M\}$, za $k \ge 1$, $\delta \in (0,1)$, i pri čemu je M nenegativan ceo broj. Opisana strategija zapravo je generalizacija Armižovog pravila. U [67] autori pretpostavljaju da pravci traženja zadovoljavaju sledeće uslove za neke pozitivne konstante b_1 i b_2:

$$g_k^T d_k \le -b_1 \|g_k\|^2$$
$$\|d_k\| \le b_2 \|g_k\|.$$

Sledeće nemonotono linijsko traženje opisano je u [74].

Neka je data početna tačka x_0, parametri $0 \le \eta_{min} \le \eta_{max} \le 1$,
$0 < \delta < \sigma < 1 < \rho$, $\mu > 0$. Neka je $C_0 = f(x_0)$, $Q_0 = 1$.
Veličina koraka t_k treba da zadovoljava sledeće uslove:

$$f(x_k + t_k d_k) \le C_k + \delta t_k g_k d_k,$$
$$g(x_k + t_k d_k) d_k \ge \sigma g_k d_k,$$

Bira se $\eta_k \in [\eta_{min}, \eta_{max}]$ i zatim izračunava $Q_{k+1} = \eta_k Q_k + 1$,

$$C_{k+1} = \frac{\eta_k Q_k C_k + f(x_{k+1})}{Q_{k+1}}.$$

Nemonotona pravila koja sadrže niz nenegativnih parametara $\{\epsilon_k\}$ prvi put su korišćena u [59], a uspešno su korišćena u mnogim drugim algoritmima, na primer, u [19]. Sledeća osobina parametara ϵ_k je pretpostavljena:

$$\epsilon_k > 0, \; \sum_k \epsilon_k = \epsilon < \infty,$$

a odgovarajuće pravilo glasi

$$f(x_k + t_k d_k) \le f(x_k) + c_1 t_k d_k^T g_k + \epsilon_k.$$

Glava 2

Gradijentni metodi

2.1 Gradijentni metodi prvog reda

Metodi optimizacije bez ograničenja, koji se baziraju na izvodima funkcije cilja, zovu se gradijentni metodi. Razlikujemo dve klase ovih metoda. Jednoj klasi pripadaju metodi koji koriste samo prvi izvod funkcije cilja i zovu se gradijentni metodi prvog reda. Najpoznatiji među gradijentnim metodima prvog reda jeste Košijev metod najbržeg pada.

2.1.1 Košijev metod najbržeg pada

Metod najbržeg pada je jedan od osnovnih metoda za rešavanje problema minimizacije bez ograničenja. Pošto koristi pravac negativnog gradijenta kao opadajući pravac, zove se takođe gradijentni metod.

Pretpostavimo da je funkcija $f(x)$ neprekidno diferencijabilna u okolini x_k, a gradijent $g_k \neq 0$. Iz Tejlorovog razvoja dobijamo:

$$f(x) = f(x_k) + (x - x_k)^T g_k + o(\|x - x_k\|). \tag{2.1.1}$$

Uvedimo oznaku $x - x_k = td_k$. Podsetimo da se pravac d_k, takav da važi $d_k^T g_k < 0$, zove opadajući pravac. Iz relacije (2.1.1) sledi da u tom slučaju važi $f(x) < f(x_k)$.

Ako fiksiramo t, možemo zaključiti da što je manja vrednost izraza $d_k^T g_k$ (t.j., što je veća vrednost $|d_k^T g_k|$), brže opada vrednost funkcije.

Iz Koši-Švarcove nejednakosti

$$|d_k^T g_k| \leq \|d_k\| \|g_k\|,$$

sledi da je vrednost $d_k^T g_k$ najmanja ako i samo ako je $d_k = -g_k$. Zbog toga je $-g_k$ pravac najbržeg pada.

Iterativna šema metoda najbržeg pada je

$$x_{k+1} = x_k - t_k g_k. \tag{2.1.2}$$

Sledi odgovarajući algoritam [116].

Algoritam 2.1.1. *Ulazni parametri:* $\epsilon > 0$, x_0, $k := 0$.

Korak 1. Ako je $\|g_k\| \leq \epsilon$, *STOP, inače, neka je* $d_k = -g_k$.

Korak 2. Naći dužinu koraka t_k *nekim pravilom linijskog traženja.*

Korak 3. Izračunati $x_{k+1} = x_k + t_k d_k$.

Korak 4. $k := k + 1$, *ići na Korak 1.*

U *Koraku 2.* može se koristiti tačno ili približno linijsko traženje; zavisno od toga, razlikujemo algoritam najbržeg pada sa tačnim, odnosno približnim linijskim traženjem.

2.1.2 Konvergencija metoda najbržeg pada

Metod najbržeg pada je važan u oblasti optimizacije sa teoretske tačke gledišta.

Sada razmatramo red globalne i red lokalne konvergencije metoda najbržeg pada.

Teorema 2.1.1. *[116] Neka je* $f \in C^1(\mathbb{R}^n)$. *Onda svaka tačka nagomilavanja iterativnog niza* $\{x_k\}$, *generisana algoritmom najbržeg pada sa tačnim linijskim traženjem, jeste stacionarna tačka.*

Ako se radi o metodu najbržeg pada sa približnim linijskim traženjem, globalna konvergencija data je sledećom teoremom.

Teorema 2.1.2. *[116] Neka je* $f \in C^1(\mathbb{R}^n)$. *Razmotrimo metod najbržeg pada sa približnim linijskim traženjem. Onda je svaka tačka nagomilavanja niza* $\{x_k\}$ *stacionarna tačka.*

Globalna konvergencija ne garantuje da je metod najbržeg pada efikasan metod. Za mnoge probleme, metod najbržeg pada je veoma spor. Naime, ovaj metod obično se dobro ponaša u početnim iteracijama, ali u blizini stacionarne tačke postaje veoma spor zbog takozvanog *cik-cak* fenomena.

To se može objasniti sledećim činjenicama.

Iz tačnog linijskog traženja, važi

$$g_{k+1}^T d_k = 0,$$

to jest

$$g_{k+1}^T g_k = d_{k+1}^T d_k = 0.$$

Ovim je dokazano da su gradijenti, izračunati u susednim iteracijama, ortogonalni jedan na drugi, te su susedni pravci takođe ortogonalni, što vodi ka pojavi pomenutog fenomena.

Uporedo sa približavanjem stacionarnoj tački, vrednost $\|g_k\|$ postaje veoma mala. Iz sledećeg izraza

$$f(x_k + td) = f(x_k) + tg_k^T d + o(\|td\|),$$

jasno je da je vrednost izraza prvog reda $tg_k^T d = -t\|g_k\|^2$ za $d = -g_k$ veoma mala, te je, takođe, pad funkcije f mali.

Dalje razmatramo red konvergencije metoda najbržeg pada.

Teorema 2.1.3. *[116] Razmotrimo problem optimizacije*

$$\min_{x \in R^n} f(x) = \frac{1}{2} x^T G x - b^T x + c, \tag{2.1.3}$$

gde je $G \in \mathbb{R}^{n\times n}$ simetrična i pozitivno definitna matrica. Neka su λ_1 i λ_n redom najveća i najmanja sopstvena vrednost matrice G. Neka je x^* rešenje problema (2.1.3). Onda niz $\{x_k\}$, generisan metodom najbržeg pada, konvergira ka x^*, konvergencija je najmanje linearna, i važe sledeće nejednakosti:

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} \leq \frac{(v-1)^2}{(v+1)^2} = \frac{(\lambda_1 - \lambda_n)^2}{(\lambda_1 + \lambda_n)^2}, \tag{2.1.4}$$

$$\frac{\|x_{k+1} - x^*\|_G}{\|x_k - x^*\|_G} \leq \frac{v-1}{v+1} = \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}, \tag{2.1.5}$$

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \sqrt{v} \cdot \frac{v-1}{v+1} = \sqrt{\frac{\lambda_1}{\lambda_n}} \cdot \frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}, \tag{2.1.6}$$

gde je $v = \dfrac{\lambda_1}{\lambda_n}$, $\|x\|_G = \|Gx\|$.

Teorema 2.1.4. *[116] Neka je $f \in C^2(D)$, pri čemu je D okolina minimuma x^* funkcije $f(x)$ i neka postoje $\epsilon > 0$ i $M > m > 0$, tako da važi, za $\|x - x^*\| < \epsilon$,*

$$m\|y\|^2 \leq y^T G(x) y \leq M\|y\|^2, \text{ za svako } y \in \mathbb{R}^n. \tag{2.1.7}$$

Ako niz $\{x_k\}$, generisan metodom najbržeg pada, konvergira ka x^*, tada je konvergencija najmanje linearna.

Teorema 2.1.5. *[116] Neka je $f \in C^2(D)$, pri čemu je D okolina minimuma x^* funkcije $f(x)$ i neka je $g(x^*) = 0$ i $G(x^*)$ pozitivno definitna funkcija. Neka niz $\{x_k\}$, generisan metodom najbržeg pada, konvergira ka x^*. Neka je*

$$\frac{f(x_{k+1}) - f(x^*)}{f(x_k) - f(x^*)} = \beta_k.$$

Tada je $\beta_k < 1$ za svako k i

$$\limsup_{k\to\infty} \beta_k \leq \frac{M-m}{M} < 1,$$

pri čemu M i m zadovoljavaju

$$0 < m \leq \lambda_n \leq \lambda_1 \leq M,$$

a λ_n i λ_1 su redom najmanja i najveća sopstvena vrednost matrice $G(x)$.

Primetimo da M i m iz Teoreme 2.1.5 zadovoljavaju i relaciju (2.1.7).

2.2 Gradijentni metodi drugog reda

Ovoj klasi metoda pripadaju oni metodi koji koriste i prvi i drugi izvod funkcije cilja, ili neke njihove aproksimacije. Najpoznatiji među ovim metodima je Njutnov metod.

2.2.1 Njutnov metod

Osnovna ideja Njutnovog metoda za optimizaciju bez ograničenja je iterativna primena kvadratne aproksimacije $q^{(k)}$ funkcije cilja f u tekućoj iteraciji x_k i minimizacija aproksimacije $q^{(k)}$.

Neka je $f \in C^2(\mathbb{R}^n)$, $x_k \in \mathbb{R}^n$, a hesijan G_k pozitivno definitna matrica. Aproksimiramo funkciju f u tački x_k kvadratnom aproksimacijom $q^{(k)}$:

$$f(x_k + s) \approx q^{(k)}(s) = f(x_k) + g_k^T s + \frac{1}{2} s^T G_k s,$$

gde je $s = x - x_k$.

Minimiziranjem $q^{(k)}(s)$ dobija se

$$x_{k+1} = x_k - G_k^{-1} g_k, \tag{2.2.1}$$

gde je $s_k = x_{k+1} - x_k = -G_k^{-1} g_k$ Njutnov pravac.

Jasno je da je Njutnov pravac opadajući pravac, ako je G_k pozitivno definitna matrica, pošto važi:

$$g_k^T s_k = -g_k^T G_k^{-1} g_k < 0.$$

Sledi odgovarajući algoritam.

Algoritam 2.2.1. *Ulazni parametri: $\epsilon > 0$, x_0, $k := 0$.*

Korak 1. Ako je $\|g_k\| \leq \epsilon$, STOP.

Korak 2. Rešiti $G_k d_k = -g_k$ po d_k.

Korak 3. Postaviti $x_{k+1} = x_k + d_k$.

Korak 4. $k := k + 1$, ići na Korak 1.

Za kvadratnu funkciju sa pozitivno definitnim hesijanom, minimum se može dostići samo jednom iteracijom Njutnovog metoda. U opštem slučaju, važi sledeća teorema.

Teorema 2.2.1. *[116] Neka je $f \in C^2(\mathbb{R}^n)$ i x_k dovoljno blizu rešenja x^* problema (1.3.1), gde je $g(x^*) = 0$. Ako je hesijan $G(x^*)$ pozitivno definitna matrica, a $G(x)$ zadovoljava Lipšicov uslov, to jest:*

$$|G_{ij}(x) - G_{ij}(y)| \leq \beta\|x - y\|,$$

tada je, za svako k, Njutnova iteracija (2.2.1) dobro definisana; generisani niz $\{x_k\}$ konvergira ka x^* sa kvadratnim redom konvergencije.

Primetimo da je Njutnov metod lokalan metod. Pošto je linijsko traženje globalna strategija, možemo koristiti Njutnov metod sa linijskim traženjem da obezbedimo globalnu konvergenciju. Sledi odgovarajući algoritam.

Algoritam 2.2.2. *Ulazni parametri: $\epsilon > 0$, x_0, $k := 0$.*

Korak 1. Ako je $\|g_k\| \leq \epsilon$, STOP.

Korak 2. Rešiti $G_k d = -g_k$ po d i postaviti $d_k = d$.

Korak 3. Naći t_k korišćenjem nekog pravila linijskog traženja.

Korak 4. Postaviti $x_{k+1} = x_k + t_k d_k$.

Korak 5. $k := k + 1$, ići na Korak 1.

Inače, primetimo da samo u slučaju kada niz veličina koraka $\{t_k\}$ teži 1, konvergencija Njutnovog metoda, datog Algoritmom 2.2.2, jeste kvadratna [116].

U *Koraku 3.* može se koristiti tačno ili približno linijsko traženje; zavisno od toga, razlikujemo Njutnov metod sa tačnim, odnosno približnim linijskim traženjem.

Sledi teorema koja tvrdi da je Algoritam 2.2.2, u kome se u *Koraku 3.* koristi tačno linijsko traženje, globalno konvergentan.

Teorema 2.2.2. *[116] Neka je $f \in C^2(D)$, gde je $D \subseteq \mathbb{R}^n$ otvoren i konveksan skup. Pretpostavimo da za ma koje $x_0 \in D$ postoji konstanta $m > 0$, takva da $f(x)$ zadovoljava*

$$u^T G(x) u \geq m\|u\|^2, \text{ za svako } u \in \mathbb{R}^n,\ x \in \mathcal{L}(x_0), \tag{2.2.2}$$

gde je $\mathcal{L}(x_0) = \{x | f(x) \leq f(x_0)\}$ odgovarajući nivo skup. Tada je niz $\{x_k\}$, generisan Algoritmom 2.2.2 sa tačnim linijskim traženjem, ili konačan, sa osobinom da je $g_k = 0$ za poslednje k, ili beskonačan i konvergira ka jedinstvenom minimumu x^* funkcije f.

Lema 2.2.1. *[116] Neka je $f : D \subset R^n \to R$ neprekidno diferencijabilna funkcija i neka $g(x)$ zadovoljava pretpostavku P_1. Ako je $f(x_k + td_k)$ ograničena sa donje strane, $t > 0$, onda za svako $t_k > 0$ koje zadovoljava (1.3.8) i (1.3.9), važi*

$$f(x_k) - f(x_k + t_k d_k) \geq \overline{\eta}\|g_k\|^2 \cos^2 \angle(d_k, -g_k), \tag{2.2.3}$$

gde je $\overline{\eta} > 0$ neka konstanta, nezavisna od k.

Teorema 2.2.3. *[116] Neka je $f \in C^2(D)$, gde je $D \subset \mathbb{R}^n$ otvoren i konveksan skup. Pretpostavimo da za ma koje $x_0 \in \mathbb{R}^n$ postoji $m > 0$ tako da $f(x)$ zadovoljava (2.2.2) na nivo skupu $\mathcal{L}(x_0)$. Ako linijsko traženje zadovoljava (2.2.3), tada za niz $\{x_k\}$, generisan Njutnovim algoritmom, važi*

$$\lim_{k\to\infty} \|g_k\| = 0,$$

a $\{x_k\}$ konvergira ka jedinstvenom minimumu funkcije $f(x)$.

2.2.2 Kvazi-Njutnovi metodi

Za različite praktične probleme, izračunavanje hesijana, koje je neophodno u Njutnovom metodu, može biti veoma skupo, ili se može desiti da hesijan ne može da se izračuna. Zbog toga je formirana klasa metoda koji koriste samo vrednosti funkcije i gradijenta funkcije, a bliski su Njutnovom metodu. Ovde se ne izračunava hesijan, ali se generiše niz aproksimacija hesijana, a pri tom se zadržava brza konvergencija.

Umesto izračunavanja hesijana G_k, kod ovih metoda konstruiše se aproksimacija hesijana, B_k. Pri tom, cilj je da niz $\{B_k\}$ poseduje pozitivnu definitnost, uz zadržavanje dobrih osobina Njutnovog metoda.

Najpre nalazimo uslove koje treba da zadovolji niz $\{B_k\}$.

Neka je $f : \mathbb{R}^n \to \mathbb{R}$ dva puta neprekidno diferencijabilna na otvorenom skupu $D \subset \mathbb{R}^n$. Neka je kvadratna aproksimacija funkcije f u tački x_{k+1} data na sledeći način:

$$f(x) \approx f(x_{k+1}) + g_{k+1}^T(x - x_{k+1}) + \frac{1}{2}(x - x_{k+1})^T G_{k+1}(x - x_{k+1}).$$

Nalaženjem izvoda, dobijamo

$$g(x) \approx g_{k+1} + G_{k+1}(x - x_{k+1}). \tag{2.2.4}$$

Korišćenjem oznaka $x = x_k,\ s_k = x_{k+1} - x_k,\ y_k = g_{k+1} - g_k$, dobija se

$$G_{k+1}^{-1} y_k \approx s_k. \tag{2.2.5}$$

U izrazu (2.2.5) važi znak jednakosti, ukoliko je u pitanju kvadratna funkcija f, sa hesijanom G, to jest,

$$s_k = G^{-1} y_k, \text{ ili } y_k = G s_k. \tag{2.2.6}$$

Sada ispitujemo mogućnost da aproksimacije inverza hesijana, u oznaci H_{k+1}, u kvazi-Njutnovom metodu zadovoljavaju relaciju (2.2.6), odnosno, da važi

$$H_{k+1} y_k = s_k, \tag{2.2.7}$$

koja je poznata kao kvazi-Njutnova jednačina, ili kvazi-Njutnov uslov. Njoj ekvivalentna jednačina

$$B_{k+1} s_k = y_k, \tag{2.2.8}$$

gde je $B_{k+1} = H_{k+1}^{-1}$, poznata je kao jednačina sečice.

Potrebno je izračunati H_{k+1} ili B_{k+1} primenom nekih pogodnih metoda, takvih da važi kvazi-Njutnova jednačina (2.2.7) ili (2.2.8).

Sledi opšti algoritam kvazi-Njutnovog metoda.

Algoritam 2.2.3. *Ulazni parametri:* $\epsilon > 0$, x_0, $k := 0$, H_0.

Korak 1. *Ako je* $\|g_k\| \le \epsilon$, *STOP.*

Korak 2. *Izračunati* $d_k = -H_k g_k$.

Korak 3. *Naći veličinu koraka* t_k *korišćenjem linijskog traženja.*
Postaviti
$x_{k+1} = x_k + t_k d_k$.

Korak 4. *Ažurirati* H_k *u* H_{k+1} *tako da važi kvazi-Njutnova jednačina (2.2.7).*

Korak 5. $k := k + 1$; *ići na Korak 1.*

U Algoritmu 2.2.3 jedna od mogućnosti je $H_0 = I$, i u tom slučaju prva iteracija je upravo iteracija najbržeg pada.

Druga varijanta Algoritma 2.2.3 umesto *Koraka 2.* i *Koraka 4.*, koristi sledeće korake:

Korak 2.* *Rešiti* $B_k d = -g_k$ *po* d_k,

Korak 4.* *Ažurirati* B_k *u* B_{k+1} *tako da važi kvazi-Njutnova jednačina (2.2.8).*

Razmotrimo sada neke od poznatih postupaka za ažuriranje kvazi-Njutnove matrice.

2.2.3 DFP ažuriranje

DFP ažuriranje je najpre predloženo u [48], a kasnije je razmatrano i u [55].

Neka je $H_0 \in R^{n\times n}$ simetrična i pozitivno definitna matrica.

Ovde se matrica H_{k+1} formira tako što se matrici H_k dodaju dve simetrične matrice, od kojih je svaka prvog ranga:

$$H_{k+1} = H_k + auu^T + bvv^T,$$

gde su $u,\ v \in \mathbb{R}^n$, a $a,\ b$ su skalari koje treba odrediti.

Iz kvazi-Njutnove jednačine (2.2.7) sledi

$$H_k y_k + auu^T y_k + bvv^T y_k = s_k. \tag{2.2.9}$$

Pri tom, u i v nisu jedinstveno određeni. Mogući izbori su

$$u = s_k,\ v = H_k y_k.\ [116]$$

Iz (2.2.9) dobija se

$$a = \frac{1}{u^T y_k} = \frac{1}{s_k^T y_k},\ b = -\frac{1}{v^T y_k} = -\frac{1}{y_k^T H_k y_k}.$$

Sada važi

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}. \tag{2.2.10}$$

Formula (2.2.10) je prvo kvazi-Njutnovo ažuriranje. Otuda potiče naziv DFP ažuriranje.

Teorema 2.2.4. *DFP ažuriranje (2.2.10) je pozitivno definitna matrica ako i samo ako važi* $s_k^T y_k > 0$.

Pretpostavka 2.2.1. (*a*) $f \in C^2(D)$, *gde je* $D \subset \mathbb{R}^n$ *otvoren konveksan skup.*

(*b*) *Funkcija* f *ima lokalni minimum* $x^* \in D$, *pri čemu je* $G(x^*)$ *simetrična i pozitivno definitna matrica.*

(*c*) *Postoji okolina* $N(x^*, \epsilon)$ *tačke* x^* *tako da važi*

$$\|G(\bar{x}) - G(x)\| \leq L\|\bar{x} - x\|,$$

za svako x, $\bar{x} \in N(x^*, \epsilon)$, *pri čemu je* L *Lipšicova konstanta.*

Teorema 2.2.5. *[116] Neka funkcija* f *zadovoljava Pretpostavku 2.2.1. Takođe, neka važi*

$$L \cdot \|\nabla^2 f(x^*)^{-1}\| \cdot \max\{\|x_k - x^*\|, \|x_{k+1} - x^*\|\} \leq \frac{1}{3}$$

u okolini tačke x^*. *Tada je DFP metod superlinearno konvergentan.*

2.2.4 BFGS ažuriranje

Analogno relaciji (2.2.10), može se dobiti relacija

$$B_{k+1}^{(BFGS)} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{2.2.11}$$

koja se zove BFGS ažuriranje, a koja je data u [20], [56] [64], [109].

Formula (2.2.11) dobija se pomoću formule (1.2.5).

Može se dobiti sledeća formula

$$H_{k+1}^{BFGS} = \left(I - \frac{s_k y_k^T}{s_k^T y_k}\right) H_k \left(I - \frac{y_k s_k^T}{s_k^T y_k}\right) + \frac{s_k s_k^T}{s_k^T y_k}. \tag{2.2.12}$$

Ako u (2.2.12) izvršimo zamene $H_k \leftrightarrow B_k$, $s_k \leftrightarrow y_k$, dobijamo sledeću formulu vezanu za DFP ažuriranje:

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

Trenutno BFGS ažuriranje predstavlja najbolje od svih kvazi-Njutnovih ažuriranja [116]. Ovo ažuriranje ima sve dobre osobine DFP ažuriranja.

Teorema 2.2.6. *[116] Neka je $f \in C^2(\mathbb{R}^n)$ i neka je matrica hesijana G Lipšic neprekidna u tački x^*. Pretpostavimo da niz generisan BFGS algoritmom konvergira ka minimumu x^* i da važi*

$$\sum_{k=1}^{\infty} \|x_k - x^*\| < \infty.$$

Tada $\{x_k\}$ konvergira ka x^* superlinearno.

Numerički rezultati dobijeni za BFGS metod bolji su od onih koji su dobijeni za DFP metod [116].

BFGS se takođe dobro ponaša ako se kombinuje sa linijskim traženjem čak i manje preciznosti [116].

Glava 3

Modifikacija metoda multiplikativnih parametara

U [106] autori razmatraju sledeći problem

$$\min f(x),\ \ x \in \mathbb{R}^n, \tag{3.0.1}$$

gde je $f(x) = \dfrac{1}{2}x^T Qx - b^T x$, $Q \in R^{n\times n}$ simetrična pozitivno definitna matrica i $b \in \mathbb{R}^n$.

Odgovarajuća iterativna šema je

$$x_{k+1} = x_k - \theta_k t_k g_k, \tag{3.0.2}$$

gde je $\theta_k \in (0, 2)$.

U istom radu, dokazana je sledeća teorema.

Teorema 3.0.7. *[106] Ako niz θ_k ima tačku nagomilavanja $\bar{\theta} \in (0, 2)$, onda niz x_k, generisan pomoću (3.0.2), konvergira ka x^*.*

Autori rada [106] smatraju da je ovakav postupak dobar način da se ubrza Košijev metod.

U ovom radu koristimo sličnu ideju da modifikujemo i ubrzamo metod iz rada [113]. Analiziramo osnovni problem optimizacije (1.1.4), pri čemu je $f \in C^2(\mathbb{R}^n)$ jako konveksna funkcija, ograničena sa donje strane.

Odgovarajuća iterativna šema je data relacijom

$$x_{k+1} = x_k + t_k d_k, \tag{3.0.3}$$

gde x_{k+1} predstavlja aproksimaciju tačnog rešenja u novoj iteraciji, d_k je pravac pretraživanja, a t_k je dužina koraka u pravcu d_k.

Ključni problem ovde je naći d_k i t_k. Potrebno je da pravac traženja zadovoljava uslov pada (1.2.1). Za izbor pravca traženja postoje razne procedure. U Njutnovom metodu sa linijskim traženjem, pravac traženja d_k dobija se rešavanjem sistema $G_k d = -g_k$. U metodu najbržeg pada, pravac traženja d_k definisan je kao $d_k = -g_k$.

Ovde koristimo proceduru linijskog traženja unazad da odredimo veličinu koraka, u oznaci t_k.

Podsetimo da je opšta iterativna šema metoda linijskog traženja data Algoritmom 1.3.1, a procedura linijskog traženja unazad data je Algoritmom 1.3.2.

3.1 Motivacija

Modifikovani Njutnovi metodi ne izračunavaju matricu hesijana. Oni generišu aproksimacije inverzne matrice matrice hesijana na neki jevtiniji način, uz zadržavanje brze konvergencije.

Opšta iterativna šema modifikovanih Njutnovih metoda je

$$x_{k+1} = x_k - t_k S_k g_k, \tag{3.1.1}$$

gde je S_k aproksimacija inverzne matrice matrice hesijana.

Ovde koristimo sledeću skalarnu aproksimaciju inverzne matrice matrice hesijana

$$S_k = \gamma_k^{-1} I,\ \ \gamma_k \in \mathbb{R} \setminus \{0\}, \tag{3.1.2}$$

koja je takođe korišćena u [113].

Primenom aproksimacije (3.1.2), modifikovani Njutnov metod (3.1.1) se redukuje na iterativnu šemu

$$x_{k+1} = x_k - t_k \gamma_k^{-1} g_k. \tag{3.1.3}$$

Inače, ideja da se hesijan aproksimira skalarnom matricom potiče iz [16]. U [16] je predložen algoritam, koji je nazvan BB algoritam. Iterativna šema koja je korišćena u [16] je $x_{k+1} = x_k - S_k^{-1} g_k$, gde je $S_k = \gamma_k^{BB} I$. Pri tom je γ_k^{BB} vrednost koja minimizira izraz

$$\|S_k s_{k-1} - y_{k-1}\|,$$

te je odgovarajuća iterativna šema BB algoritma

$$x_{k+1} = x_k - \frac{1}{\gamma_k^{BB}} g_k,\ \ \gamma_k^{BB} = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}.$$

Sa druge strane, u $(k+1)-$oj iteraciji iterativne šeme (3.1.3), primenom Tejlorovog razvoja dobija se

$$f(x_{k+1}) = f(x_k) - t_k g_k^T \gamma_k^{-1} g_k + \frac{1}{2} t_k^2 (\gamma_k^{-1} g_k)^T G(\xi) \gamma_k^{-1} g_k, \tag{3.1.4}$$

pri čemu je $\xi \in [x_k, x_{k+1}]$ dato relacijom

$$\xi = x_k + a(x_{k+1} - x_k) = x_k - a t_k \gamma_k^{-1} g_k,\ \ 0 \leq a \leq 1. \tag{3.1.5}$$

Ako su x_k i x_{k+1} dovoljno blizu, može se uzeti $a = 1$, te se tako dobija aproksimacija

$$\xi = x_{k+1}, \tag{3.1.6}$$

pa je

$$G(\xi) = \gamma_{k+1} I. \tag{3.1.7}$$

Dakle, polazeći od $\gamma_0 = 1$, korišćenjem Tejlorovog razvoja funkcije f u okolini tačke x_{k+1}, vrši se aproksimacija hesijana G_k dijagonalnom matricom $B_k = \gamma_k I$, te se dobija

$$\gamma_0 = 1, \tag{3.1.8}$$

$$\gamma_{k+1} = 2\gamma_k \frac{\gamma_k[f(x_{k+1}) - f(x_k)] + t_k\|g_k\|^2}{t_k^2\|g_k\|^2}. \tag{3.1.9}$$

Pošto je funkcija f konveksna, hesijan funkcije f je pozitivan, te mora važiti

$$\gamma_k > 0, \tag{3.1.10}$$

za svako k. Ukoliko relacija (3.1.10) nije zadovoljena, definiše se

$$\gamma_k = 1. \tag{3.1.11}$$

Motivacija za izbor parametra γ_k kao u (3.1.11) leži u sledećem: kada G_k nije pozitivno definitna matrica, koristi se pravac negativnog gradijenta $-g_k$, uz korišćenje $t_k = 1$.

Koristimo vrednost parametra θ_k koji je izabran na slučajan način iz intervala $[\mu, 1)$, za neko $\mu > 0$. Ovaj parametar treba da bude parametar ubrzanja.

Pokazaćemo da uvođenje parametra θ_k ne smanjuje red konvergencije, a numerički rezultati, dati u Glavi 6, pokazuju da uvođenje takvog parametra može biti korisno.

3.2 Modifikacija korišćenjem slučajno izabranog parametra

Polazimo od iterativne šeme

$$x_{k+1} = x_k - \theta_k t_k G_k^{-1} g_k.$$

Dalje koristimo aproksimaciju hesijana G_k matricom $B_k = \gamma_k I$, pri čemu je γ_k dato relacijama (3.1.8)-(3.1.9).

Dakle, važi

$$x_{k+1} = x_k - \theta_k t_k \gamma_k^{-1} g_k, \tag{3.2.1}$$

gde je $\theta_k \in [\mu, 1)$ slučajno izabrani parametar, $k \in \mathbb{N}$, $x_0 \in \mathbb{R}^n$.

Tako dobijamo tzv. relaksirani kvazi-Njutnov metod, u oznaci RGDQN, koji je izvesna relaksacija metoda GDQN datog u radu [113].

Sledi odgovarajući algoritam.

Algoritam 3.2.1. *Ulazni parametri:*
$\epsilon > 0$, x_0, $k := 0$, $\theta_0 = 1$, $c_1 \in (0,1)$, $\mu > 0$.

Korak 1. Ako je $\|g_k\| \leq \epsilon$, *STOP.*

Korak 2. Odrediti γ_k *na osnovu (3.1.8)-(3.1.9) i definisati* $d_k = -\gamma_k^{-1} g_k$.

Korak 3. Odrediti t_k *na osnovu Algoritma linijskog pretraživanja unazad, tako da važi*

$$f(x_k + t_k d_k) \leq f(x_k) + c_1 t_k g_k^T d_k.$$

Korak 4. Izabrati $\theta_k \in [\mu, 1)$ *i definisati* $x_{k+1} = x_k + \theta_k t_k d_k$.

Korak 5. Postaviti $k = k + 1$ *i ići na Korak 1.*

U [108] je dokazano da, ukoliko važi $f \in C^2(\mathbb{R}^n)$ i ukoliko je funkcija f uniformno konveksna na $\mathbb{R}^n$, tada važi:

funkcija $f(x)$ ima jedinstven minimum u tački x^*, i postoje realni brojevi m, M, takvi da važi $0 < m \leq M$ i takvi da važe sledeće nejednakosti:

$$m\|y\|^2 \leq y^T \nabla^2 f(x) y \leq M\|y\|^2, \text{ za svako } x, y \in \mathbb{R}^n; \tag{3.2.2}$$

$$\frac{1}{2} m\|x - x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2} M\|x - x^*\|^2, \text{ za svako } x \in \mathbb{R}^n; \tag{3.2.3}$$

$$m\|x - y\|^2 \leq (g(x) - g(y))^T (x - y) \leq M\|x - y\|^2, \text{ za svako } x, y \in \mathbb{R}^n. \tag{3.2.4}$$

Uzimanjem $y = x^*$ u nejednakosti (3.2.4), zatim primenom Teoreme o srednjoj vrednosti i Koši-Švarcove nejednakosti, dobijamo

$$m\|x - x^*\| \leq \|g(x)\| \leq M\|x - x^*\|, \tag{3.2.5}$$

za svako $x \in \mathbb{R}^n$.

Teorema 3.2.1. *Neka je* $f \in C^2(\mathbb{R}^n)$ *uniformno konveksna funkcija. Niz* $\{x_k\}$, *generisan Algoritmom 3.2.1, konvergira ka jedinstvenom minimumu funkcije* f *u tački* x^*.

Dokaz. Za proizvoljno x_k posmatramo funkciju $h(t) = f(x_k - t\gamma_k^{-1} g_k)$.

Kako je f konveksna funkcija na $\mathbb{R}^n$, tada je $h(t)$ konveksna funkcija na $\mathbb{R}$ za $t \geq 0$. Takođe je $h(t)$ ograničena funkcija sa donje strane, na osnovu ograničenosti funkcije f. Posmatrajmo linearnu funkciju

$$l(t) = f(x_k) - c_1 t \gamma_k^{-1} g_k^T g_k.$$

Očigledno je

$$\lim_{t \to \infty} l(t) = -\infty,$$

jer je $c_1 > 0$.

Znamo da je $l(0) = h(0) = f(x_k)$.

Takođe znamo da je $h(t) = f(x_k - td_k) = f(x_k) - tg(y)^T g_k$, gde je $d_k = \gamma_k^{-1} g_k$, $y = x_k - \tau t d_k$, $\tau \in (0,1)$. Za dovoljno malo τ važi $g(y)^T g_k > 0$, te postoji $\bar{t}$ tako da za $t \in (0, \bar{t})$ važi $h(t) < l(t)$, jer je $c_1 < 1$. Sa druge strane, $l(t)$ je neograničena, a $h(t)$ je ograničena funkcija, pa postoji t^* u kojoj funkcija h seče pravu l, to jest, važi

$$
\begin{aligned}
& h(t^*) = f(x_k - t^* d_k) = f(x_k) - c_1 t^* d_k^T g_k, \text{ i} \\
& h(t) \leq l(t), \; t \in [0, t^*], \\
& h(t) > l(t), \; t > t^*.
\end{aligned}
$$

Konveksnost funkcije f obezbeđuje da je tačka t^* jedina tačka sa osobinama koje smo naveli, odnosno,

$$
f(x_k - t\gamma_k^{-1} g_k) > f(x_k) - c_1 t \gamma_k^{-1} g_k^T g_k, \text{ za } t > t^*, \text{ i} \tag{3.2.6}
$$

$$
f(x_k - t\gamma_k^{-1} g_k) \leq f(x_k) - c_1 t \gamma_k^{-1} g_k^T g_k, \text{ za } t \in (0, t^*). \tag{3.2.7}
$$

Na osnovu (3.2.7) zaključujemo da za svako $k \in \mathbb{N}$ postoji j_k tako da za $t_k = \beta^{j_k}$ važi

$$
f(x_k + t_k d_k) \leq f(x_k) + c_1 t_k g_k^T d_k, \text{ i} \tag{3.2.8}
$$

$$
f(x_k + \beta^{j_k - 1} d_k) > f(x_k) + c_1 \beta^{j_k - 1} g_k^T d_k. \tag{3.2.9}
$$

Pokazaćemo da postoji $t_* > 0$ tako da je $t_k = \beta^{j_k - 1} \geq t_* > 0$. Pretpostavimo suprotno, da postoji podniz $\{t_k\}_{k \in K_0}$ tako da je

$$
\lim_{k \in K_0} t_k = 0.
$$

Tada je

$$
\lim_{k \in K_0} \beta^{j_k - 1} = 0,
$$

odnosno

$$
\lim_{k \in K_0} (j_k - 1) = \infty.
$$

Sada važi

$$
\lim_{k \in K_0} \frac{f(x_k + \beta^{j_k - 1} d_k) - f(x_k)}{\beta^{j_k - 1}} \geq c_1 g_k^T d_k,
$$

to jest,

$$
g_k^T d_k \geq c_1 g_k^T d_k. \tag{3.2.10}
$$

Međutim, $0 < c_1 < 1$, $g_k^T d_k < 0$, odakle sledi da je (3.2.10) ispunjeno samo ako je $g_k^T d_k = 0$.

Sa druge strane, kako je

$$
g_k^T d_k = -\gamma_k^{-1} g_k^T g_k = -\gamma_k^{-1} \|g_k\|^2,
$$

zaključujemo da je $\|g_k\| = 0$.

Dakle, važi $t_k \geq t_*$. Kako je $\theta_k \in [\mu, 1]$, onda je $0 < t_*\mu \leq \theta_k t_k \leq t^*$, te važi

$$f(x_k - \theta_k t_k \gamma_k^{-1} g_k) \leq f(x_k) - c_1 \theta_k t_k \gamma_k^{-1} g_k^T g_k,$$

odnosno

$$f(x_{k+1}) \leq f(x_k). \tag{3.2.11}$$

Iz nejednakosti (3.2.11) i toga što je f ograničena sa donje strane na skupu $\mathcal{L}(x_0)$, zaključujemo da je

$$\lim_{k\to\infty} (f(x_{k+1}) - f(x_k)) = 0. \tag{3.2.12}$$

Funkcija $f(x)$ je uniformno konveksna, te je nivo skup $\mathcal{L}(x_0)$ zatvoren i ograničen. Niz $\{x_k\} \subseteq \mathcal{L}(x_0)$ je ograničen, pa niz $\{x_k\}$ ima konvergentan podniz. Neka je $K \subseteq \mathbb{N}$ tako da je

$$\lim_{k\in K} x_k = \bar{x}.$$

Kako je $\theta_k \in [\mu, 1) \subset [\mu, 1]$ za $k \in K$, onda i niz $\{\theta_k\}_{k\in K}$ ima konvergentan podniz, označimo ga sa $\{\theta_k\}_{k\in K_1}$.

Dakle, za $k \in K_1$ važi

$$\lim_{k\in K_1} x_k = \bar{x}, \ \lim_{k\in K_1} \theta_k = \bar{\theta}.$$

Takođe važi

$$f(x_{k+1}) \leq f(x_k) - c_1 t_k \theta_k \gamma_k^{-1} g_k^T g_k, \ k \in K_1. \tag{3.2.13}$$

Sumiranjem nejednakosti (3.2.13) za $k \in K_1$ dobijamo

$$f(x_{k+1}) \leq f(x_0) - c_1 \sum_{i=1}^{k} t_i \theta_i \gamma_i^{-1} g_i^T g_i. \tag{3.2.14}$$

Ograničenost f sa donje strane sada implicira da je

$$\sum_{i=1}^{k} t_i \theta_i \gamma_i^{-1} g_i^T g_i \leq \infty,$$

za proizvoljno $k \in K_1$, te važi

$$\lim_{k\in K_1} t_k \theta_k \gamma_k^{-1} g_k^T g_k = 0.$$

Takođe važi

$$\begin{aligned} 0 = \lim_{k\in K_1} t_k \theta_k \gamma_k^{-1} g_k^T g_k &= \bar{\theta} \lim_{k\in K_1} t_k \gamma_k^{-1} g_k^T g_k \\ &\geq \bar{\theta} t_* \mu \lim_{k\in K_1} \gamma_k^{-1} g_k^T g_k, \end{aligned}$$

i $\gamma_k^{-1} g_k^T g_k \geq 0$, pa zaključujemo da je

$$\lim_{k\in K_1} \gamma_k^{-1} g_k^T g_k = 0. \tag{3.2.15}$$

Niz $\{\gamma_k\}_{k\in K_1}$ je definisan sa (3.1.8)-(3.1.9).

Dalje, imamo u vidu formule

$$\gamma_0 = 1,$$
$$\gamma_{k+1} = 2\gamma_k \frac{\gamma_k(f(x_{k+1}) - f(x_k)) + t_k\|g_k\|^2}{t_k^2\|g_k\|^2}.$$

Znamo da važi

$$x_{k+1} = x_k - t_k\theta_k\gamma_k^{-1}g_k.$$

Dalje je

$$f(x_{k+1}) = f(x_k) - t_k\theta_k\gamma_k^{-1}\|g_k\|^2 + \frac{1}{2}t_k^2\theta_k^2\gamma_k^{-2}g_k^T\nabla^2 f(\xi)g_k,\ \xi \in [x_k, x_{k+1}],$$

Množenjem sa γ_k izraza $f(x_{k+1}) - f(x_k)$, dobijamo

$$\gamma_k(f(x_{k+1}) - f(x_k)) = -t_k\theta_k\|g_k\|^2 + \frac{1}{2}t_k^2\theta_k^2\gamma_k^{-1}g_k^T\nabla^2 f(\xi)g_k.$$

Sada važi

$$2\gamma_k \frac{\gamma_k(f(x_{k+1}) - f(x_k)) + t_k\|g_k\|^2}{t_k^2\|g_k\|^2} = \frac{2\gamma_k(1-\theta_k)t_k\|g_k\|^2}{t_k^2\|g_k\|^2} + \frac{t_k^2\theta_k^2 g_k^T\nabla^2 f(\xi)g_k}{t_k^2\|g_k\|^2}.$$

Kako je f jako (uniformno) konveksna, važi (3.2.2), odakle dobijamo

$$m\|g_k\|^2 \leq g_k^T\nabla^2 f(\xi)g_k \leq M\|g_k\|^2.$$

Posmatrajmo relaciju

$$\gamma_{k+1} = \frac{2(1-\theta_k)\gamma_k}{t_k} + \frac{t_k^2\theta_k^2 g_k^T\nabla^2 f(\xi)g_k}{t_k^2\|g_k\|^2}.$$

Pošto je $\theta_k \in [\mu, 1]$, važi $1 - \theta_k > 0$. Takođe, pokazali smo da važi $t_k \geq t_* > 0$. Zaključujemo da, ako je $\gamma_k > 0$, važi

$$\gamma_{k+1} > \frac{t_k^2\theta_k^2 g_k^T\nabla^2 f(\xi)g_k}{t_k^2\|g_k\|^2},$$

odnosno,

$$\gamma_{k+1} > m\theta_k^2 > m\mu^2. \tag{3.2.16}$$

Kako je $\gamma_0 > 0$, (3.2.16) važi za sve vrednosti k.

Koristimo relaciju (3.2.16), pa važi:

$$0 = \lim_{k\in K_1} \gamma_k^{-1}g_k^T g_k \leq \frac{1}{m\mu^2}\lim_{k\in K_1}\|g_k\|^2,$$

što implicira

$$\lim_{k\in K_1}\|g_k\|^2 = 0,$$

odnosno,

$$g(\bar{x}) = 0.$$

Znamo da je

$$\lim_{k \in K_1} x_k = \bar{x},$$

te sledi da je

$$\lim_{k \in K_1} f(x_k) = f(\bar{x}).$$

No, niz $\{f(x_k)\}$ zadovoljava (3.2.12), pa kako ima konvergentan podniz, zaključujemo da je ceo niz konvergentan, odnosno važi

$$\lim_{k \to \infty} f(x_k) = f(\bar{x}).$$

Kako je funkcija f konveksna, to je x^* jedinstveni minimum, pa $g(\bar{x}) = 0$ implicira $\bar{x} = x^*$.

Pored toga, kako je

$$\lim_{k \to \infty} f(x_k) = f(x^*),$$

primenom relacije (3.2.3), dobijamo

$$\lim_{k \to \infty} x_k = x^*,$$

čime je teorema dokazana. □

U daljem tekstu uvodimo pretpostavku da gradijent funkcije f zadovoljava Lipšicov uslov.

Lema 3.2.1. *Neka je $f \in C^2(\mathbb{R}^n)$ uniformno konveksna funkcija, i neka μ zadovoljava uslove*

$$\mu \in \left(\frac{1}{2}, 1\right) \quad i \quad \frac{2(1-\mu)}{t_*} < 1. \tag{3.2.17}$$

Za niz $\{x_k\}$, generisan Algoritmom 3.2.1, pri čemu je $c_1 < \frac{1}{2}$, važi

$$f(x_k) - f(x_{k+1}) \geq \eta \|g_k\|^2, \tag{3.2.18}$$

gde je

$$\eta = \min \left\{ \frac{c_1}{2\left(1 + \dfrac{M}{1-c}\right)}, \frac{c_1 \mu \left(1 - c_1\right) \beta}{L} \right\}, \tag{3.2.19}$$

a L je Lipšicova konstanta.

Dokaz. Uočimo skupove $T_1 = \{k | t_k = 1\}$, $T_2 = \{k | t_k < 1\}$ iz Algoritma 3.2.1. Koristeći proceduru linijskog traženja dobijamo da važi:

$$f(x_k) - f(x_{k+1}) \geq -c_1 \theta_k g_k^T d_k, \ k \in T_1, \tag{3.2.20}$$

$$f(x_k) - f(x_{k+1}) \geq -c_1 \theta_k t_k g_k^T d_k, \ k \in T_2. \tag{3.2.21}$$

Kako je $t_k < 1$, onda za $t'_k = \dfrac{t_k}{\beta}$, $k \in T_2$, važi

$$f(x_k) - f\left(x_k + \theta_k \frac{t_k}{\beta} d_k\right) < -c_1 \theta_k \frac{t_k}{\beta} g_k^T d_k. \tag{3.2.22}$$

Sada, primenom Teoreme o srednjoj vrednosti na levoj strani nejednakosti (3.2.22), postoji $\xi \in (0,1)$, tako da važi

$$-\frac{t_k}{\beta}\theta_k g\left(x_k + \xi\theta_k \frac{t_k}{\beta} d_k\right)^T d_k < -c_1 \theta_k \frac{t_k}{\beta} g_k^T d_k,$$

to jest

$$g\left(x_k + \xi\theta_k \frac{t_k}{\beta} d_k\right)^T d_k > c_1 g_k^T d_k. \tag{3.2.23}$$

Sada, oduzimanjem izraza $g_k^T d_k$ od leve i desne strane nejednakosti (3.2.23), dobijamo

$$\begin{aligned}
-(1-c_1) g_k^T d_k &< g(x_k + \theta_k \xi \frac{t_k}{\beta} d_k)^T d_k - g_k^T d_k && (3.2.24)\\
&= (g(x_k + \xi\theta_k \frac{t_k}{\beta} d_k) - g_k)^T d_k \le L \xi \theta_k \frac{t_k}{\beta} d_k^T d_k && (3.2.25)\\
&\le L \frac{t_k}{\beta} \|d_k\|^2, && (3.2.26)
\end{aligned}$$

jer je $\xi\theta_k < 1$.

Pretpostavili smo da važi $\mu > \dfrac{1}{2}$, $c_1 < \dfrac{1}{2}$.

Kako je $d_k = -\gamma_k^{-1} g_k$, to je

$$(1-c_1)\,\gamma_k^{-1}\|g_k\|^2 \le L\frac{t_k}{\beta}\gamma_k^{-2}\|g_k\|^2.$$

Dalje je

$$\begin{aligned}
t_k &\ge \frac{\beta\gamma_k}{L}(1-c_1) && (3.2.27)\\
&> \frac{m\mu^2\beta}{L}(1-c_1)\,,\ \ k \in T_2. && (3.2.28)
\end{aligned}$$

Dokazaćemo sada da pod uslovom (3.2.17), važi $\gamma_{k+1} \le 1 + \dfrac{M}{1-c}$.

Dakle, važi

$$\gamma_{k+1} = \frac{2(1-\theta_k)\gamma_k}{t_k} + \theta_k^2 \frac{t_k^2 g_k^T \nabla^2 f(x_k) g_k}{t_k^2 \|g_k\|^2},$$

i takođe,

$$\begin{aligned}
\gamma_{k+1} &\le \frac{2(1-\theta_k)\gamma_k}{t_k} + \theta_k^2 M\\
&\le \frac{2(1-\theta_k)}{t_k}\left(\frac{2(1-\theta_{k-1})}{t_{k-1}}\gamma_{k-1} + \theta_{k-1}^2 M\right) + \theta_k^2 M\\
&= a_k(a_{k-1}\gamma_{k-1} + b_{k-1}) + b_k,
\end{aligned}$$

pri čemu je $a_k = \dfrac{2(1-\theta_k)}{t_k}$, $b_k = \theta_k^2 M$.

Dakle, sada važi

$$\begin{aligned}
\gamma_{k+1} &\leq a_k(a_{k-1}\gamma_{k-1} + b_{k-1}) + b_k \\
&\leq a_k(a_{k-1}(a_{k-2}\gamma_{k-2} + b_{k-2}) + b_{k-1}) + b_k \\
&= a_k a_{k-1} a_{k-2}\gamma_{k-2} + a_k a_{k-1} b_{k-2} + a_k b_{k-1} + b_k.
\end{aligned}$$

Važi

$$a_k = \frac{2(1-\theta_k)}{t_k} \leq \frac{2(1-\mu)}{t_*} < 1,$$
$$b_k = \theta_k^2 M < M,$$

i

$$\gamma_{k+1} \leq \gamma_0 \prod_{j=0}^{k} a_j + b_0 \prod_{j=1}^{k} a_j + b_1 \prod_{j=2}^{k} a_j + \cdots + b_k.$$

Takođe važi

$$\prod_{j=0}^{k} a_j \leq \left(\frac{2(1-\mu)}{t_*}\right)^{k+1},$$

a na osnovu (3.2.17)

$$c = \frac{2(1-\mu)}{t_*} < 1.$$

Sada imamo

$$\gamma_{k+1} \leq \gamma_0 c^{k+1} + b_0 c^k + b_1 c^{k-1} + \cdots + b_{k-1} c + b_k.$$

Kako važi da je $b_k \leq M$ za $k = 0, 1, \ldots,$ to je

$$\begin{aligned}
\gamma_{k+1} &\leq \gamma_0 c^{k+1} + M(c^k + c^{k-1} + \cdots + c + 1) \\
&= \gamma_0 c^{k+1} + M\frac{1-c^{k+1}}{1-c}.
\end{aligned}$$

Dakle,

$$\gamma_{k+1} \leq 1 + \frac{M}{1-c}. \tag{3.2.29}$$

Sada koristimo (3.2.27) i činjenicu da je $\theta_k \geq \mu$, te dobijamo

$$f(x_k) - f(x_{k+1}) \geq c_1 \theta_k t_k \gamma_k^{-1} \|g_k\|^2 \tag{3.2.30}$$
$$\geq c_1 \mu\,(1-c_1)\,\frac{\beta}{L}\|g_k\|^2,\ \ k \in T_2. \tag{3.2.31}$$

Sa druge strane, za $k \in T_1$ važi

$$f(x_k) - f(x_{k+1}) \geq -c_1 \theta_k g_k^T d_k = c_1 \theta_k \gamma_k^{-1} \|g_k\|^2 \tag{3.2.32}$$
$$\geq \mu c_1 \gamma_k^{-1} \|g_k\|^2 \geq \frac{c_1}{2\left(1 + \dfrac{M}{1-c}\right)}\|g_k\|^2, \tag{3.2.33}$$

pošto je $\mu > \frac{1}{2}$, $\gamma_k \leq 1 + \frac{M}{1-c}$.

Sada iz relacija (3.2.31) i (3.2.32), zaključujemo da važi (3.2.18), pri čemu važi (3.2.19).□

Teorema 3.2.2. *Neka je $f \in C^2(\mathbb{R}^n)$ uniformno konveksna funkcija, neka je $c_1 < \frac{1}{2}$, i neka važi (3.2.17). Tada niz $\{x_k\}$, generisan Algoritmom 3.2.1, konvergira ka jedinstvenom minimumu x^* funkcije f najmanje linearno.*

Dokaz. Počev od nejednakosti (3.2.18), nakon primene (3.2.5) i (3.2.3), dobijamo

$$f(x_k) - f(x_{k+1}) \geq \eta \|g(x_k)\|^2 \geq \eta m^2 \|x_k - x^*\|^2 \geq 2\eta \frac{m^2}{M}(f(x_k) - f(x^*)).$$

Ostaje da dokažemo da važi

$$2\eta \frac{m^2}{M} < 1.$$

Neka je najpre $\eta = \dfrac{c_1}{2\left(1 + \dfrac{M}{1-c}\right)}$.

Tada važi

$$2\eta \frac{m^2}{M} = \frac{2c_1 m^2}{2M\left(1 + \dfrac{M}{1-c}\right)} = \frac{c_1 m^2}{M\left(1 + \dfrac{M}{1-c}\right)} < \frac{m}{1 + \dfrac{M}{1-c}} < 1,$$

pošto je $1 + \dfrac{M}{1-c} > 1 + M > m$.

Neka je sada $\eta = \dfrac{c_1(1-c_1)\beta\mu}{L}$.

Dakle, ispitujemo da li važi

$$2\eta \frac{m^2}{M} = \frac{2c_1(1-c_1)\beta\mu}{L} \cdot \frac{m^2}{M} < 1.$$

Važi $\dfrac{m}{M} < 1$, a iz (3.2.4) važi $\dfrac{m}{L} < 1$.

Dalje važi

$$2c_1(1-c_1)\beta\mu < 1,$$

čime je dokaz završen. □

Izvesna modifikacija korišćenjem slučajno izabranog parametra razmatrana je takođe u radu [51].

Glava 4

Metod konjugovanih gradijenata

4.1 Generalna forma metoda konjugovanih gradijenata

Za potrebe ove sekcije podsetimo se Volfovih uslova.

Standardna Volfova pravila su

$$f(x_k + t_k d_k) - f(x_k) \le \delta t_k g_k^T d_k, \tag{4.1.1}$$

$$g_{k+1}^T d_k \ge \sigma g_k^T d_k, \tag{4.1.2}$$

gde je d_k opadajući pravac, a $0 < \delta \le \sigma < 1$. Ova efikasna strategija sastoji se zapravo u prihvatanju pozitivne dužine koraka t_k, ako su zadovoljeni uslovi (4.1.1) i (4.1.2).

Jaki Volfovi uslovi sastoje se od (4.1.1) i sledeće jače verzije uslova (4.1.2):

$$|g_{k+1}^T d_k| \le -\sigma g_k^T d_k. \tag{4.1.3}$$

U generalisanim Volfovim uslovima [34], apsolutna vrednost u (4.1.3) zamenjena je parom nejednakosti

$$\sigma_1 g_k^T d_k \le g_{k+1}^T d_k \le -\sigma_2 g_k^T d_k, \tag{4.1.4}$$

pri čemu je $0 < \delta < \sigma_1 < 1$ i $\sigma_2 \ge 0$.

Specijalan slučaj $\sigma_1 = \sigma_2 = \sigma$ odgovara jakim Volfovim uslovima.

Jedan od najjednostavnijih i verovatno najelegantnijih metoda za rešavanje problema (1.3.1) jeste metod konjugovanih gradijenata.

Ovaj metod deformiše pravac najbržeg pada dodajući ovom pravcu pozitivan umnožak pravca korišćenog u poslednjem koraku.

Metod konjugovanih gradijenata prevazilazi problem spore konvergencije koji je karakteristika metoda najbržeg pada. Bitne osobine ovog metoda su mali zahtevi za memorijom i lokalna i globalna konvergencija. On zahteva samo izvode prvog reda, te nema potrebe za izračunavanjem i skladištenjem izvoda drugog reda što karakteriše Njutnov metod. Upravo zbog toga što ne zahteva ni hesijan niti aproksimaciju hesijana, danas se ovaj metod koristi za rešavanje različitih mnogobrojnih problema optimizacije.

Za rešavanje problema (1.3.1), počev od inicijalne tačke $x_0 \in \mathbb{R}^n$, metod konjugovanih gradijenata generiše niz $\{x_k\}$ kao

$$x_{k+1} = x_k + t_k d_k, \tag{4.1.5}$$

gde je $t_k > 0$ veličina koraka, dobijena linijskim traženjem, a pravci d_k generisani su kao

$$d_0 = -g_0,\ d_{k+1} = -g_{k+1} + \beta_k d_k. \tag{4.1.6}$$

U izrazu (4.1.6), β_k je parametar konjugovanog gradijenta.

Zaključak sledeće teoreme, poznat kao Zoutendijkov uslov, često se koristi, između ostalog, da se dokaže globalna konvergencija nelinearnih metoda konjugovanih gradijenata; dali su je Zoutendijk [128] i Wolfe [120], [121].

Teorema 4.1.1. *Razmotrimo ma koji iterativni metod oblika (4.1.5), pri čemu d_k zadovoljava uslov pada (1.2.1), a t_k zadovoljava standardne Volfove uslove (4.1.1) i (4.1.2). Ako važi pretpostavka P1, tada*

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{4.1.7}$$

Jedna od glavnih osobina metoda konjugovanih gradijenata jeste ta što su njegovi pravci konjugovani. Zbog toga, najpre razmatramo konjugovane pravce i metode konjugovanih pravaca.

4.1.1 Metodi konjugovanih pravaca

Očigledno, ako su vektori d_1, d_2, $\ldots$ d_m G-konjugovani vektori, oni su linearno nezavisni. Ako je $G = I$, tada je konjugovanost vektora ekvivalentna ortogonalnosti vektora.

Sledi algoritam opšteg metoda konjugovanih pravaca.

Algoritam 4.1.1. *Ulazni parametri: $\epsilon > 0$, x_0, $k := 0$, d_0, g_0, $d_0^T g_0 < 0$.*

Korak 1. *Ako je $\|g_k\| \le \epsilon$, STOP.*

Korak 2. *Izračunati t_k nekim metodom linijskog traženja.*
Postaviti $x_{k+1} = x_k + t_k d_k$.

Korak 3. *Izračunati d_{k+1}, tako da važi*

$$d_{k+1}^T G d_j = 0,\ j = 0, 1, \ldots, k.$$

Korak 4. *Postaviti $k := k + 1$, ići na Korak 1.*

Metod konjugovanih pravaca je važna klasa metoda optimizacije.

Sledeća teorema pokazuje da, ukoliko se radi o tačnom linijskom traženju, metodi konjugovanih pravaca imaju osobinu kvadratnog završetka, što znači da metod nalazi rešenje u najviše n koraka, ako se primeni na kvadratnu funkciju sa pozitivno definitnim hesijanom.

Neka je data kvadratna funkcija

$$f(x) = \frac{1}{2}x^T Gx + b^T x + c, \tag{4.1.8}$$

gde je G $n \times n$ simetrična pozitivno definitna matrica, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$.

Teorema 4.1.2. *[116] Za kvadratnu funkciju (4.1.8), sa pozitivno definitnim hesijanom, metod konjugovanih pravaca nalazi rešenje u najviše n tačnih linijskih traženja. Svako x_{i+1} je minimum funkcije cilja u potprostoru generisanom vektorima x_0, $d_0, d_1, \dots, d_i$.*

Teorema 4.1.2 je jednostavna, ali je veoma važna. Svi metodi konjugovanih pravaca se oslanjaju na ovu teoremu. Uz primenu tačnog linijskog traženja, svi metodi konjugovanih pravaca zadovoljavaju

$$g_{i+1}^T d_j = 0, \ j = 0, 1, \dots, i$$

i imaju osobinu kvadratnog završetka. Dakle, konjugovanost i tačno linijsko traženje zajedno impliciraju kvadratni završetak.

4.2 Različiti izbori parametra kod metode konjugovanih gradijenata

Među metodima konjugovanih pravaca, metod konjugovanih gradijenata je od posebnog značaja.

Jedan od poznatih metoda konjugovanih gradijenata je FR metod, sa parametrom

$$\beta_{k-1} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \tag{4.2.1}$$

koji je u literaturi poznat kao FR [58]. Metod konjugovanih gradijenata ima globalnu konvergenciju i lokalnu kvadratnu konvergenciju u n koraka.

Sledeća teorema daje glavne osobine metoda konjugovanih gradijenata.

Teorema 4.2.1. *[116] Za pozitivno definitnu kvadratnu funkciju (4.1.8), metod konjugovanih gradijenata (4.1.5), (4.1.6), (4.2.1) sa tačnim linijskim traženjem generiše tačno rešenje posle $m \le n$ koraka, a sledeće osobine važe za svako i, $0 \le i \le m$:*

$$d_i^T G d_j = 0,\ j = 0, 1, \ldots, i-1, \tag{4.2.2}$$

$$g_i^T g_j = 0,\ j = 0, 1, \ldots, i-1, \tag{4.2.3}$$

$$d_i^T g_i = -g_i^T g_i, \tag{4.2.4}$$

$$span\{g_0, g_1, \ldots, g_i\} = span\{g_0, Gg_0, \ldots, G^i g_0\}, \tag{4.2.5}$$

$$span\{d_0, d_1, \ldots, d_i\} = span\{g_0, Gg_0, \ldots, G^i g_0\}, \tag{4.2.6}$$

gde je m broj različitih sopstvenih vrednosti matrice G.

Prethodni uslovi (4.2.2), (4.2.3) i (4.2.4) predstavljaju redom konjugovanost pravaca, ortogonalnost gradijenata i pravac opadanja.

Potprostor $span\{g_0, Gg_0, \ldots, G^i g_0\}$ zove se potprostor Krilova (Krylov).

Teorema 4.2.2. *Neka je $f(x)$ kvadratna funkcija. Tada je*

$$\frac{\|x_k - x^*\|_G}{\|x_0 - x^*\|_G} \leq \left(\frac{\sqrt{\kappa(G)} - 1}{\sqrt{\kappa(G)} + 1} \right)^k,$$

gde je $\kappa(G)$ uslovni broj matrice G.

Teorema 4.2.3. *Neka je $f(x)$ kvadratna funkcija. Počev od x_1, iteracija x_{k+2} metoda konjugovanih gradijenata nakon $k+1$ iteracija zadovoljava*

$$E(x_{k+2}) \leq \left(\frac{\lambda_{k+1} - \lambda_n}{\lambda_{k+1} + \lambda_n} \right)^2 E(x_1) = \left(\frac{1 - \lambda_n/\lambda_{k+1}}{1 + \lambda_n/\lambda_{k+1}} \right)^2 E(x_1), \tag{4.2.7}$$

gde je $E(x)$ definisano kao

$$E(x) = \frac{1}{2}(x - x^*)^T G (x - x^*),$$

a sopstvene vrednosti λ_i matrice G zadovoljavaju

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq \lambda_{k+1} \geq \ldots \geq \lambda_n > 0.$$

Formula (4.2.7) ukazuje da će, nakon svake dodatne iteracije metoda konjugovanih gradijenata, uticaj veće sopstvene vrednosti biti uklonjen.

Sledi algoritam metoda konjugovanih gradijenata u opštem slučaju.

Algoritam 4.2.1. *Ulazni parametri: $\epsilon > 0$, x_0, $k := 0$, $t_0 = 0$, $d_{-1} = 0$, $d_0 = -g_0$, $\beta_{-1} = 0$, $\beta_0 = 0$.*

Korak 1. Ako je $\|g_k\| \leq \epsilon$, STOP.

Korak 2. Izračunati veličinu koraka t_k nekim metodom linijskog traženja.
Izračunati β_k nekim metodom konjugovanih gradijenata, pri čemu važi $\beta_{-1} = 0$, $\beta_0 = 0$.
Izračunati $d_k = -g_k + \beta_{k-1} d_{k-1}$.
Postaviti $x_{k+1} = x_k + t_k d_k$.

Korak 3. Postaviti $k := k + 1$ i ići na Korak 1.

Različiti metodi konjugovanih gradijenata odgovaraju različitim izborima parametra konjugovanih gradijenata β_k.

Neka je $\|\cdot\|$ Euklidska norma. Podsetimo se oznake $y_k = g_{k+1} - g_k$.

Neki izbori parametra β_k su:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \ [84], \tag{4.2.8}$$

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \ [58], \tag{4.2.9}$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \ [98], [99], \tag{4.2.10}$$

$$\beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-g_k^T s_k}, \ [57], [12], \tag{4.2.11}$$

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-g_k^T s_k}, \ [88], [12], \tag{4.2.12}$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}, \ [39], \tag{4.2.13}$$

$$\beta_k^{N} = \left(y_k - 2 d_k \frac{\|y_k\|^2}{d_k^T y_k} \right)^T \frac{g_{k+1}}{d_k^T y_k}, \ [71]. \tag{4.2.14}$$

U radu [84] prvi put je predstavljen algoritam konjugovanih gradijenata, koji je formiran za rešavanje sistema linearnih jednačina i zove se linearan metod konjugovanih gradijenata. To je jedan iterativni metod, namenjen rešavanju sistema sa pozitivno definitnim matricama sistema. Međutim, rešavanje linearnog sistema ekvivalentno je rešavanju problema minimizacije kvadratne funkcije (4.1.8), što je motivisalo proširenje HS metoda konjugovanih gradijenata za nelinearnu optimizaciju. Ako je funkcija $f(x)$ kvadratna, koristeći tačno linijsko traženje i konjugovanost pravaca, dobija se (4.2.9). Ali, proizvoljnu funkciju f, koja je dovoljno glatka, možemo aproksimirati kvadratnim modelom. Tu leži ideja da se formula (4.2.9) koristi i u minimizaciji proizvoljne, dovoljno glatke funkcije.

Metod sa parametrom (4.2.11) odlikuje se osobinom da uvek proizvodi pravac pada ukoliko su zadovoljeni jaki Volfovi uslovi (4.1.1)-(4.1.3).

Cilj je naći metod konjugovanog gradijenta koji generiše opadajuće pravce ukoliko su standardni Volfovi uslovi zadovoljeni.

Motivacija za formiranje DY metoda sa parametrom (4.2.13) opisana je u radu [39]. Najpre se pretpostavlja da važi $g_k^T d_k < 0$, to jest, da je d_k opadajući pravac. Potrebno je naći β_k, tako da je d_{k+1} opadajući pravac. Dakle, množenjem relacije (4.1.6) izrazom g_{k+1}^T sa leve strane, a imajući u vidu da treba da važi $g_{k+1}^T d_{k+1} < 0$, dobija se

$$-\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k < 0.$$

Uz pretpostavku da je $\beta_k > 0$ i uvođenjem oznake $\tau_k = d_k^T y_k$, dobija se (4.2.13).

CG metod sa parametrom β_k^N dobija se modifikovanjem HS metoda na sledeći način:

$$\beta_k^\theta = \beta_k^{HS} - \theta_k \left(\frac{\|y_k\|^2 g_{k+1}^T d_k}{(d_k^T y_k)^2} \right). \tag{4.2.15}$$

Ako se radi o minimizaciji jako konveksne kvadratne funkcije i tačnom linijskom traženju, onda su svi navedeni izbori parametra β_k ekvivalentni.

Međutim, za funkcije cilja koje nisu kvadratne, različiti izbori parametra β_k vode ka različitim performansama.

Inače, teorija konvergencije za metode sa brojiocem $\|g_{k+1}\|^2$ mnogo je razvijenija nego za metode sa brojiocem $g_{k+1}^T y_k$. Ali, praktično ponašanje metoda sa brojiocem $g_{k+1}^T y_k$ često je bolje od metoda sa brojiocem $\|g_{k+1}\|^2$.

Analiza metoda sa izborom β_k^N, koji je nazvan CG-DESCENT u [72] dovodi nas do zaključka da se ovaj izbor razlikuje od drugih po tome što je d_{k+1} uvek opadajući pravac za ma koju veličinu koraka $t_k > 0$, ako važi $d_k^T y_k \neq 0$.

4.2.1 Početni pravac traženja

Uobičajeno je uzeti $d_0 = -g_0$ u metodima konjugovanih gradijenata. U [24] pokazano je da konvergencija ne mora biti bolja od linearne ako početni pravac traženja nije pravac najbržeg pada, čak i za jako konveksne kvadratne funkcije. Kasnije, u [102] dat je jedan značajan rezultat; naime, pokazano je da ako je funkcija cilja konveksna kvadratna funkcija i ako je pravac traženja proizvoljan opadajući pravac, tada se ili optimalnost postiže u najviše $n + 1$ koraka, ili je konvergencija samo linearna.

U cilju postizanja konvergencije za proizvoljan početni pravac traženja, u [92] je predložen CG algoritam, baziran na rekurentnoj relaciji od tri izraza:

$$d_{k+1} = -y_k + \frac{y_k^T y_k}{d_k^T y_k} d_k + \frac{y_{k-1}^T y_k}{d_{k-1}^T y_{k-1}} d_{k-1}, \tag{4.2.16}$$

gde je d_0 proizvoljan opadajući pravac a $d_{-1} = 0$.

Ako je f konveksna kvadratna funkcija, tada su za ma koju veličinu koraka t_k, pravci traženja dati pomoću (4.2.16), konjugovani u odnosu na hesijan funkcije f. Ta interesantna činjenica, međutim, nije našla značajnu primenu u praksi.

4.2.2 Metodi kod kojih je $\|g_{k+1}\|^2$ brojilac parametra β_k

Podsetimo se izraza: $\beta_k^{FR} = \dfrac{\|g_{k+1}\|^2}{\|g_k\|^2}$, $\beta_k^{CD} = \dfrac{\|g_{k+1}\|^2}{-g_k^T s_k}$, $\beta_k^{DY} = \dfrac{\|g_{k+1}\|^2}{d_k^T y_k}$.

Naredna teorema [44], [46] daje rezultat konvergencije primenljiv na ma koji metod, kod koga β_k može da se izrazi kao

$$\beta_k = \frac{\Phi_{k+1}}{\Phi_k}. \tag{4.2.17}$$

FR metod odgovara izboru $\Phi_k = \|g_k\|^2$.

Koristeći (4.1.6), β_k^{DY} može da se napiše kao

$$\beta_k^{DY} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}.$$

Tako, DY metod ima formu (4.2.17), uz $\Phi_k = g_k^T d_k$.

Teorema 4.2.4. *[44], [46] Razmotrimo ma koji iterativni metod oblika (4.1.5)-(4.1.6), gde β_k ima oblik (4.2.17), d_k zadovoljava uslov pada (1.2.1), a važi pretpostavka P1. Ako Zoutendijkov uslov važi i ako je*

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\Phi_k^2} = \infty, \ \textit{ili} \sum_{k=0}^{\infty} \frac{\|g_k\|^2}{\Phi_k^2} = \infty, \ \textit{ili} \sum_{k=1}^{\infty} \prod_{i=1}^{k} \beta_i^{-2} = \infty,$$

tada je algoritam (4.1.5)-(4.1.6)-(4.2.17) globalno konvergentan.

Kao posledica ovog rezultata, DY metod jeste globalno konvergentan kada se implementira sa standardnim Volfovim linijskim traženjem, pošto važi:

$$\sum_{k=0}^{N} \frac{(g_k^T d_k)^2}{\Phi_k^2} = N + 1,$$

kada je $\Phi_k = g_k^T d_k$. FR je globalno konvergentan kada se implementira sa jakim Volfovim linijskim traženjem, uz $\sigma \leq \dfrac{1}{2}$, pošto važi

$$\sum_{k=0}^{N} \frac{\|g_k\|^2}{\Phi_k^2} = N + 1,$$

kada je $\Phi_k = \|g_k\|^2$.

Pimetimo da proizvoljni CG metod može da se izrazi u obliku (4.2.17), uzimajući da je $\Phi_0 = 1$ i

$$\Phi_k = \prod_{j=1}^{k} \beta_k, \ \text{za } k > 0.$$

Dalje, konstatujemo da metodi FR, CD i DY imaju zajednički brojilac $\|g_{k+1}\|^2$. Jedna bitna razlika između ovih metoda s jedne strane i ostalih izbora parametra β_k jeste da teoreme globalne konvergencije zahtevaju samo pretpostavku P1, ne i pretpostavku P2.

Prvi rezultat globalne konvergencije FR metoda dao je Zoutendijk [128] 1970. godine. On je dokazao da FR metod konvergira globalno ako je t_k tačno rešenje problema

$$\min_{t\geq 0} f(x_k + td_k). \tag{4.2.18}$$

Drugim rečima, globalna konvergencija se dobija kada je linijsko traženje tačno. Kasnije, u [101] utvrđeno je da je FR metod, sa tačnim linijskim traženjem, osetljiv na fenomen zaglavljivanja. To znači da algoritam može proizvesti niz kratkih koraka bez značajnog progresa ka minimumu. Prvi rezultat globalne konvergencije FR metoda sa netačnim linijskim traženjem dat je u [1]. Pod jakim Volfovim uslovima, za $\sigma < \frac{1}{2}$, u radu [1] dokazano je da FR metod generiše pravce dovoljnog pada.

Preciznije, dokazano je da važi

$$\frac{1 - 2\sigma + \sigma^{k+1}}{1-\sigma} \leq \frac{-g_k^T d_k}{\|g_k\|^2} \leq \frac{1-\sigma^{k+1}}{1-\sigma},$$

za svako $k \geq 0$. Kao posledica toga, globalna konvergencija je utvrđena uz korišćenje Zoutendijkovog uslova.

Za $\sigma = \frac{1}{2}$, utvrđeno je u [1] da je d_k opadajući pravac; ipak, za $\sigma = \frac{1}{2}$, u [1] nije pokazano da važi dovoljan pad.

Dalje, u [87] dokaz globalne konvergencije koji je dat u [1] proširen je na slučaj $\sigma = \frac{1}{2}$. Daljom analizom u [34] pokazano je da bar jedna iteracija od dve uzastopne FR iteracije zadovoljava osobinu dovoljnog pada. Drugim rečima,

$$\max\left\{-\frac{g_k^T d_k}{\|g_k\|^2}, -\frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2}\right\} \geq \frac{1}{2}.$$

Dokazi globalne konvergencije metoda konjugovanih gradijenata često se zasnivaju na Zoutendijkovom uslovu, u kombinaciji sa analizom koja pokazuje da uslov dovoljnog pada (1.2.2) važi, i da postoji konstanta β takva da je

$$\|d_k\|^2 \leq \beta k. \tag{4.2.19}$$

Kombinovanjem (1.2.2), (4.1.7) i (4.2.19), dobijamo

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{4.2.20}$$

Globalna konvergencija metoda konjugovanih gradijenata zapravo znači da ili je $g_k = 0$ za neko k, ili važi (4.2.20).

Jedan rezultat koji se odnosi na Zoutendijkov uslov, a koji se može naći u [40], ili u [75], a sve uz pretpostavku da pravci traženja jesu pravci pada, dat je sledećom teoremom.

Teorema 4.2.5. *Razmotrimo ma koji iterativni metod oblika (4.1.5)-(4.1.6), gde d_k zadovoljava uslov pada (1.2.1) i t_k zadovoljava jake Volfove uslove. Ako pretpostavka P1 važi, tada je ili*

$$\liminf_{k\to\infty} \|g_k\| = 0,$$

ili

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty.$$

Pri implementaciji linijskog traženja, mora biti zadovoljena neka verzija Volfovih uslova, a takođe treba da obezbedimo da je novi pravac - opadajući pravac. U novijim verzijama CG algoritama [39], [42] pridruženim izboru β_k^{DY}, opadajući pravac važi automatski, čim se zadovolje standardni Volfovi uslovi. Teorema 4.2.5 takođe se može koristiti da se dobije rezultat globalne konvergencije FR metoda sa jakim Volfovim linijskim traženjem i $\sigma \le \frac{1}{2}$, pošto su pravci traženja uvek pravci pada.

U [41] je dokazano da, u FR šemi, jaki Volfovi uslovi ne garantuju opadajući pravac kada je $\sigma > \frac{1}{2}$, čak i za funkciju $f(x) = \lambda\|x\|^2$, gde je $\lambda > 0$ konstanta.

Zbog toga, ograničenje $\sigma \le \frac{1}{2}$ mora važiti da bi se obezbedio pad.

Na drugoj strani, u [35] je pokazano da kada je $\sigma > \frac{1}{2}$ i $g_k^T d_k > 0$, $-d_k$ može da se iskoristi kao pravac traženja. A ako je $g_k^T d_k = 0$, može se uzeti $x_{k+1} := x_k$. Ako postoji konstanta γ, tako da važi $\|g_k\| \le \gamma$, tada pod pretpostavkom $P1$, FR metod, sa standardnim Volfovim traženjem i uz navedene korekcije kada je $g_k^T d_k \ge 0$, jeste globalno konvergentan, što je, takođe, dokazano u [35].

U [34], pod pretpostavkom da se radi o generalisanom Volfovom traženju, globalna konvergencija je dobijena kada je $\sigma_1 + \sigma_2 \le 1$.

Metod dat u [57], koji koristi parametar β_k^{CD}, blisko je povezan sa FR metodom. Ako je linijsko traženje tačno, važi $\beta_k^{FR} = \beta_k^{CD}$. Jedna važna razlika između ovih metoda jeste da kod CD metoda dovoljan pad (1.2.2) važi za jako Volfovo linijsko traženje. Ograničenje $\sigma \le \frac{1}{2}$, koje važi za FR, nije potrebno za CD. Za linijsko traženje koje zadovoljava generalisane Volfove uslove, uz $\sigma_1 < 1$ i $\sigma_2 = 0$, pokazuje se da važi $0 \le \beta_k^{CD} \le \beta_k^{FR}$. Zbog toga, iz analize u [1], ili korišćenjem Teoreme 4.2.5, sledi globalna konvergencija CD metoda.

Na drugoj strani, ako je $\sigma_1 \ge 1$ ili $\sigma_2 > 0$, u [36] su konstruisani primeri gde $\|d_k\|^2$ raste eksponencijalno, a CD metod konvergira ka tački u kojoj gradijent ne teži nuli.

DY metod, razmatran najpre u [39], fundamentalno je različit od FR i CD metoda. Uz standardno Volfovo linijsko traženje, DY metod uvek generiše pravce pada. Štaviše, važi globalna konvergencija, kada pretpostavka P1 važi.

Sledeća teorema povezuje pravce pada, generisane pomoću DY, sa dovoljnim uslovom pada.

Teorema 4.2.6. *[43] Razmotrimo metod (4.1.5)-(4.1.6), gde je $\beta_k = \beta_k^{DY}$. Ako je DY metod implementiran sa ma kojim linijskim traženjem za koje su pravci traženja* *pravci pada, i ako postoje konstante γ_1 i γ_2, takve da važi $0 < \gamma_1 \leq \|g_k\| \leq \gamma_2$, za svako $k \geq 1$, onda za ma koje $p \in (0,1)$, postoji konstanta $c > 0$, takva da važi dovoljan pad:*

$$g_i^T d_i \leq -c\|g_i\|^2,$$

za najmanje $\lfloor p \cdot k \rfloor$ indeksa $i \in [1,k]$, gde $\lfloor r \rfloor$ označava najveći ceo broj manji ili jednak r.

4.2.3 Metodi kod kojih je $g_{k+1}^T y_k$ brojilac parametra β_k

Potrebno je da navedemo sledeću pretpostavku.

Pretpostavka P_3: Razmotrimo metod

$x_{k+1} = x_k + t_k d_k,\ d_0 = -g_0,\ d_{k+1} = -g_{k+1} + \beta_k d_k$.

Neka važi $0 \leq \gamma \leq \|g_k\| \leq \overline{\gamma}$, za svako $k \geq 0$, gde su γ i $\overline{\gamma}$ dve pozitivne konstante. Postoje konstante $b > 1$ i $\lambda > 0$, takve da za svako k važi:

$$|\beta_k| \leq b,\ \|s_k\| \leq \lambda \ \Rightarrow\ |\beta_k| \leq \frac{1}{2b}.\ [63]$$

Podsetimo se izraza: $\beta_k^{PRP} = \dfrac{g_{k+1}^T y_k}{\|g_k\|^2}$, $\beta_k^{HS} = \dfrac{g_{k+1}^T y_k}{d_k^T y_k}$ i $\beta_k^{LS} = \dfrac{g_{k+1}^T y_k}{-g_k^T s_k}$.

Za razliku od jake teorije konvergencije koja je razvijena za metode kod kojih je $\|g_{k+1}\|^2$ brojilac parametra β_k, svi metodi kod kojih je $g_{k+1}^T y_k$ brojilac parametra β_k su osetljivi na fenomen zaglavljivanja.

PRP, HS i LS metodi, koji imaju zajednički brojilac $g_{k+1}^T y_k$, poseduju ugrađen restart, koji se odnosi na problem zaglavljivanja: kada je korak $s_k = x_{k+1} - x_k$ mali, izraz $y_k = g_{k+1} - g_k$ u brojiocu parametra β_k teži nuli. Tako se umanjuje β_k, te novi pravac traženja d_{k+1} postaje pravac najbržeg pada. U stvari, metodi PRP, HS i LS automatski podešavaju β_k da izbegnu fenomen zaglavljivanja; u opštem slučaju, performanse ovih metoda su bolje od performansi metoda kod kojih je $\|g_{k+1}\|^2$ u brojiocu parametra β_k.

U [98] globalna konvergencija PRP metoda utvrđena je kada je funkcija f jako konveksna i linijsko traženje tačno. Pod pretpostavkom da je pravac traženja pravac opadanja, u [125] utvrđena je globalna konvergencija PRP metoda za jako konveksne funkcije cilja i Volfovo linijsko traženje. Za jako Volfovo linijsko traženje u [26] dat je primer da čak i kada je funkcija cilja jako konveksna i $\sigma \in (0,1)$ dovoljno malo, PRP metod može ipak da generiše pravac rasta funkcije.

Pretpostavimo sada da važi:

(A_1) $f : \mathbb{R}^n \to \mathbb{R}$ je tri puta neprekidno diferencijabilna;

(A_2) postoje konstante $M > m > 0$, takve da važi

$$m\|y\|^2 \leq y^T G(x) y \leq M\|y\|^2,\ \text{za svako } y \in \mathbb{R}^n,\ x \in \mathcal{L},$$

gde je $\mathcal{L}$ ograničen nivo skup.

Teorema 4.2.3.1. [116] *Neka važe pretpostavke* (A_1) *i* (A_2). *Tada niz* $\{x_k\}$, *generisan PRP metodom ili FR metodom konjugovanih gradijenata sa restartom, ima kvadratnu konvergenciju u* n *koraka, to jest, postoji konstanta* $c > 0$, *takva da važi*

$$\limsup_{k\to\infty} \frac{\|x_{kr+n} - x^*\|}{\|x_{kr} - x^*\|^2} \le c < \infty,$$

gde r *znači da se metod restartuje na svakih* r *iteracija.*

Dalje, u [107] je pokazano da je konvergencija superkvadratna u n koraka, to jest:

$$\|x_{k+n} - x^*\| = o(\|x_k - x^*\|^2).$$

Još neki rezultati o redu konvergencije metoda konjugovanih gradijenata mogu se naći u [114].

U [128] je pokazano da ovaj metod ne može biti neuspešan. Naime, dokazano je u [128] da FR metod sa tačnim linijskim traženjem jeste globalno konvergentan za proizvoljnu funkciju. U [1] ovaj rezultat proširen je na približno linijsko traženje.

Metodi HS i PRP veoma se slično ponašaju u praksi, i bolji su od FR metoda [63].

Podsetimo, međutim, da je u [100] pokazano da PRP metod sa tačnim linijskim traženjem može da divergira.

Isti rezultat važi za HS metod, pošto su ova dva metoda identična kada je $g_k^T d_{k-1} = 0$, to jest za tačno linijsko traženje.

Analiza PRP metoda koja je realizovana u radu [100], pokazala je da PRP metod može da divergira čak i u prisustvu tačnog linijskog traženja. Takođe, u radu [100], utvrđeno je da konvergencija PRP metoda nije izvesna za nekonveksne funkcije cilja. Zbog svega toga, u radu [100] autor je predložio sledeću modifikaciju PRP metoda:

$$\beta_k^{PRP+} = \max\{0, \beta_k^{PRP}\}. \tag{4.2.21}$$

U [63] je pokazana konvergencija PRP+ metoda. Naime, u [63] je pokazano da izbor β_k dat pomoću (4.2.21) zaista rezultira globalnom konvergencijom, kako za tačno, tako i za približno linijsko traženje. Takođe, analiza izvršena u [63] primenjuje se i na familiju metoda sa nenegativnim parametrom β_k, koji dele zajedničku osobinu sa PRP metodom, u oznaci P_3.

U [63] definisana je sledeća strategija:

pozitivna dužina koraka t_k se prihvata ako važi

$$f(x_k + t_k d_k) \le f(x_k + \overline{t_k} d_k), \tag{4.2.22}$$

gde je $\overline{t_k}$ najmanja pozitivna stacionarna tačka funkcije $\xi_k(t) = f(x_k + td_k)$. Ovde se u stvari radi o takozvanom idealnom linijskom traženju, a uslov (4.2.22) zove se uslov idealnog linijskog traženja. Ovo linijsko traženje zove se idealno upravo zbog toga što se kao njegov rezultat dobija $\bar{t_k}$ koje je najmanja stacionarna tačka funkcije $\xi_k(t) = f(x_k + td_k)$. Pretpostavke P1 i P2 garantuju da $\overline{t_k}$ postoji.

Teorema 4.2.7. *[63] Neka važe pretpostavke P1 i P2; razmotrimo ma koju iteraciju oblika (4.1.5), gde je d_k pravac traženja, a t_k zadovoljava jedan od sledećih uslova linijskog traženja:*

$$(i)\ \ \textit{Volfove uslove traženja}\ (4.1.1)\ \textit{i}\ (4.1.2), \tag{4.2.23}$$

$$(ii)\ \textit{uslov idealnog linijskog traženja}\ (4.2.22). \tag{4.2.24}$$

Tada važi:

$$\sum_{k\geq 1} \cos^2(\theta_k)\|g_k\|^2 < \infty. \tag{4.2.25}$$

(4.2.25) takođe je poznat kao *Zoutendijkov uslov*.

4.2.4 Metodi bliski PRP metodu sa nenegativnim parametrom β_k

Sada se bavimo metodima kod kojih je $\beta_k \geq 0$ za svako k. Motivacija za uzimanje ovog ograničenja potiče od primera datog u [100], gde se PRP metod ponaša ciklično, bez približavanja rešenju. Drugi razlog za uzimanje ograničenja $\beta_k \geq 0$ jeste taj što nam ono dozvoljava da lako primenimo osobinu pada u algoritmu, kao što ćemo sada diskutovati.

Razmotrimo metod (4.1.5)-(4.1.6), sa ma kojim $\beta_k \geq 0$. Zahtevaćemo da dovoljan pad (1.2.2) važi za neko $0 < c \leq 1$ i za svako $k \geq 1$.

Za razliku od FR metoda, sada jaki Volfovi uslovi (4.1.1) i (4.1.3) ne obezbeđuju dovoljan pad (1.2.2).

Iz (4.1.6) imamo

$$g_k^T d_k = -\|g_k\|^2 + \beta_k g_k^T d_{k-1}. \tag{4.2.26}$$

Zbog toga, da bi se obezbedio pad u svakoj iteraciji algoritma približnog linijskog traženja, mora se obezbediti da izraz $\beta_k g_k^T d_{k-1}$ nije previše veliki.

Za rezultate koji slede ne specificiramo posebno strategiju linijskog traženja.

Pretpostavljamo da je $g_k \neq 0$ za svako k.

Lako se dokazuje da ako važe pretpostavke P_1 i P_2, tada, kada su u pitanju PRP i HS, važi i pretpostavka P_3.

Teorema 4.2.8. *[63] Neka važe pretpostavke* $P1$, $P2$ *i* $P3$. *Razmotrimo metod (4.1.5)-(4.1.6) sa sledećim osobinama:*

1. $\beta_k \geq 0$ *za svako* k,

2. *linijsko traženje zadovoljava* $\{x_k\} \subset \mathcal{L}$, *Zoutendijkov uslov (4.2.25) i uslov dovoljnog pada (1.2.2).*

Tada je $\liminf \|g_k\| = 0$.

Pošto metodi PRP i HS zadovoljavaju pretpostavku P_3, prethodna teorema se primenjuje na njih, ukoliko β_k ograničimo da bude nenegativno. To, između ostalog, sugeriše formulu (4.2.21) i sledeću formulu:

$$\beta_k = |\beta_k^{PRP}|,$$

a takođe i odgovarajuće formule za HS metod. Od posebnog su interesa približna linijska traženja, kao što je Volfovo traženje.

Posledica 4.2.1. *[63] Neka važe pretpostavke P_1 i P_2. Razmotrimo metod (4.1.5)-(4.1.6), uz (4.2.21) i sa linijskim traženjem koje zadovoljava Volfove uslove i uslov dovoljnog pada (1.2.2). Tada je*

$$\liminf \|g_k\| = 0.$$

Primetimo da postoji veza između (4.2.21), koji se može razumeti kao automatsko restartovanje i restart kriterijuma koji je dat u [101]. Restart kriterijum, dat u [101], tvrdi da restart nije potreban sve dok važi

$$|g_k^T g_{k-1}| \leq \nu \|g_k\|^2, \tag{4.2.27}$$

pri čemu je ν mala pozitivna konstanta.

Iz izraza za β_k^{PRP} uslov $\beta_k^{PRP} \geq 0$ je ekvivalentan uslovu

$$g_k^T g_{k-1} \leq \|g_k\|^2.$$

Sledi da se rezultat globalne konvergencije Posledice 4.2.1 takođe primenjuje na PRP metod sa restartom (4.2.27), pod uslovom da je $\nu \leq 1$.

PRP+ metod predstavljen je da ispravi nedostatke konvergencije PRP metoda, kada se implementira sa Volfovim linijskim traženjem.

Drugi način da se ispravi nedostatak konvergencije PRP metoda, jeste da se zadrži PRP formula, ali da se modifikuje linijsko traženje. Specijalno, u [69] je predloženo novo linijsko traženje Armijovog tipa, sledećeg oblika

$$t_k = \max\left\{\lambda^j \frac{\tau |g_k^T d_k|}{\|d_k\|^2}\right\}, \tag{4.2.28}$$

gde je $j \geq 0$ najmanji ceo broj sa osobinom da važi

$$f(x_{k+1}) \leq f(x_k) - \delta t_k^2 \|d_k\|^2, \tag{4.2.29}$$

$$-c_1\|g_{k+1}\|^2 \leq g_{k+1}^T d_{k+1} \leq c_2\|g_{k+1}\|^2, \tag{4.2.30}$$

pri čemu su $0 < c_2 < 1 < c_1$, $0 < \lambda < 1$, $\tau > 0$ konstante.

Koristeći linijsko traženje, dato relacijama (4.2.28), (4.2.29), (4.2.30), u [69] je dokazana globalna konvergencija PRP metoda.

U skorijem radu [70], ovo linijsko traženje kombinuje se sa tehnikom oblasti poverenja.

U drugom istraživanju, pokazano je u [41] da je PRP metod globalno konvergentan kada se linijsko traženje primeni sa konstantnom veličinom koraka $t_k = \eta < \frac{1}{4}L$, gde je L Lipšicova konstanta za $g(x)$.

HS metod ima osobinu da uslov konjugacije

$$d_{k+1}^T y_k = 0 \tag{4.2.31}$$

uvek važi, nezavisno od linijskog traženja.

Za tačno linijsko traženje važi $\beta_k^{HS} = \beta_k^{PRP}$. Zbog toga bi osobine konvergencije HS metoda trebalo da budu slične osobinama konvergencije PRP metoda.

Specijalno, iz analize izvršene u [100], sledi da HS metod sa tačnim linijskim traženjem ne mora da konvergira za nelinearnu funkciju u opštem slučaju.

Ponovimo da ako pravci traženja zadovoljavaju uslov dovoljnog pada i ako se koristi standardno Volfovo traženje, tada HS metod zadovoljava pretpostavku $P3$. Slično PRP+ metodu, ako je

$$\beta_k^{HS+} = \max\{\beta_k^{HS}, 0\},$$

tada iz Teoreme 4.2.8, sledi da je HS+ metod globalno konvergentan.

4.3 Adaptivni izbor parametra

Hibridni algoritmi konjugovanih gradijenata dinamički podešavaju formulu za β_k sa svakom novom iteracijom.

Sledi pregled nekih hibridnih metoda, predstavljen u [3], [5]. Koristimo oznake iz [3], [5].

$$\beta_k^{hDY} = \max\{c\beta_k^{DY}, \min\{\beta_k^{HS}, \beta_k^{DY}\},\ \ c = \frac{1-\sigma}{1+\sigma} \quad [42],$$

$$\beta_k^{hDY_z} = \max\{0, \min\{\beta_k^{HS}, \beta_k^{DY}\}\} \quad [42],$$

$$\beta_k^{GN} = \max\{-\beta_k^{FR}, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\} \quad [63],$$

$$\beta_k^{HuS} = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\} \quad [85],$$

$$\beta_k^{TaS} = \begin{cases} \beta_k^{PRP}, & 0 \le \beta_k^{PRP} \le \beta_k^{FR} \\ \beta_k^{FR}, & \text{inače} \end{cases} \quad [118],$$

$$\beta_k^{LS-CD} = \max\{0, \min\{\beta_k^{LS}, \beta_k^{CD}\}\} \quad [88].$$

Kao što smo mogli primetiti, sa jedne strane, metodi FR, DY, CD imaju jake osobine konvergencije, ali se nisu najbolje pokazali u praksi, zahvaljujući fenomenu zaglavljivanja. S druge strane, metodi PRP, HS i LS ne moraju da konvergiraju u opštem slučaju, ali se u praksi često bolje ponašaju od FR, DY, CD. Prirodno se dolazi do ideje da se napravi kombinacija metoda jedne i metoda druge grupe, da bi se iskoristile dobre osobine metoda obeju grupa.

U [118] predložen je sledeći hibridni metod

$$\beta_k^{TaS} = \begin{cases} \beta_k^{PRP}, \text{ ako je } 0 \le \beta_k^{PRP} \le \beta_k^{FR}, \\ \beta_k^{FR}, \text{inače.} \end{cases} \tag{4.3.1}$$

Zaključujemo da se, pri pojavi zaglavljivanja, koristi PRP metod.

Slično, u [85] predložen je sledeći hibridni metod

$$\beta_k^{HuS} = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}. \tag{4.3.2}$$

U [63] je primećeno da β_k^{PRP} može imati negativnu vrednost čak i za jako konveksne funkcije. U naporu da se prošire dopušteni izbori parametra β_k^{PRP}, uz zadržavanje globalne konvergencije, u [63] predložen je sledeći izbor parametra:

$$\beta_k = \max\{-\beta_k^{FR}, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}.$$

U ovom hibridnom metodu, β_k može biti negativno, obzirom da je β_k^{FR} uvek nenegativno. Iz numeričkih rezultata [63], zaključuje se da performansa ovog metoda nije bolja nego performansa PRP+ metoda za testirani skup primera.

Rezultati konvergencije koji u sebi sadrže izvesno ograničenje parametrom β_k^{FR} mogu se naći u [63], [75], [85].

Specijalno, u [75] je dokazana sledeća teorema.

Teorema 4.3.1. *Razmotrimo ma koji CG metod oblika (4.1.5)-(4.1.6), koji koristi jako Volfovo traženje (4.1.1)-(4.1.3), uz* $\sigma \le \frac{1}{2}$. *Ako važi pretpostavka* $P1$ *i* $2\sigma|\beta_k| \le \beta_k^{FR}$, *tada je metod globalno konvergentan, a pravci traženja su uvek pravci pada.*

U [42] je proučavana mogućnost kombinacije DY metoda sa drugim CG metodima.

Za standardno Volfovo linijsko traženje i za $\beta_k \in [-\eta\beta_k^{DY}, \beta_k^{DY}]$, gde je $\eta = \dfrac{1-\sigma}{1+\sigma}$, u [42] je utvrđena globalna konvergencija ako važi pretpostavka $P1$.

U [42] su predložena sledeća dva hibridna metoda:

$$\beta_k = \max\left\{-\frac{1-\sigma}{1+\sigma}\beta_k^{DY}, \min\{\beta_k^{HS}, \beta_k^{DY}\}\right\}. \tag{4.3.3}$$

i

$$\beta_k = \max\{0, \min\{\beta_k^{HS}, \beta_k^{DY}\}\}. \tag{4.3.4}$$

Numerički rezultati dati u [45], pokazuju da metod (4.3.4) daje najbolje rezulate, ponašajući se bolje nego PRP+.

Drugi hibridni metod, koji je predložen u [27], koristi ili DY ili CD šemu:

$$\beta_k = \frac{\|g_{k+1}\|^2}{\max\{d_k^T y_k, -g_k^T d_k\}}.$$

U [27] je pokazano da ova hibridna šema generiše pravce pada nezavisno od linijskog traženja. Ova osobina pada jača je od same DY šeme, gde pad važi uz Volfovo linijsko traženje. U [27] je takođe pokazano da za ovu hibridnu šemu, važi $\beta_k \in [0, \beta_k^{DY}]$.

U [37], [46], predložena je jednoparametarska familija CG metoda, gde je

$$\beta_k = \frac{\|g_{k+1}\|^2}{\lambda_k \|g_k\|^2 + (1-\lambda_k) d_k^T y_k},$$

pri čemu je $\lambda_k \in (0,1)$ parametar. Izbor $\lambda_k = 1$ odgovara FR metodu, a izbor $\lambda_k = 0$ odgovara DY metodu. U [38], ova familija je proširena, uzimanjem da je $\lambda_k \in (-\infty, \infty)$; ako važi pretpostavka $P1$, onda važi globalna konvergencija za svakog člana familije ako se primeni generalisano Volfovo traženje, gde je $\sigma_1 - 1 \leq (\sigma_1 + \sigma_2)\lambda_k \leq 1$.

Razmatranjem konveksnih kombinacija brojioca i imenioca za β_k^{FR} i β_k^{HS}, u [93] je predložena dvoparametarska familija CG metoda:

$$\beta_k = \frac{\mu_k \|g_{k+1}\|^2 + (1-\mu_k) g_{k+1}^T y_k}{\lambda_k \|g_k\|^2 + (1-\lambda_k) d_k^T y_k},$$

gde su μ_k, $\lambda_k \in [0,1]$.

Ova dvoparametarska familija uključuje FR, DY, PRP, HS metode u krajnjim slučajevima. Primetivši da standardni CG metodi (njih šest) imaju dva brojioca i tri imenioca, autori u [44] razmatraju čak širu familiju CG metoda uključivanjem više parametara:

$$\beta_k = \frac{\mu_k \|g_{k+1}|\|^2 + (1-\mu_k) g_{k+1}^T y_k}{(1-\lambda_k-\omega_k)\|g_k\|^2 + \lambda_k d_k^T y_k - \omega_k d_k^T g_k},$$

gde su λ_k, $\mu_k \in [0,1]$, $\omega_k \in [0, 1-\lambda_k]$. Ova troparametarska familija uključuje šest standardnih CG metoda, prethodne jednoparametarske i dvoparametarske familije i mnoge hibridne metode kao specijalne slučajeve. Da bi se obezbedilo da pravci traženja, generisani ovom familijom budu pravci pada, koristi se restart kriterijum (4.2.27) [101].

U [30] autori su modifikovali brojilac parametra β_k^{HS} i dobili

$$\beta_k^{DL} = \frac{g_{k+1}^T (y_k - t s_k)}{d_k^T y_k} = \beta_k^{HS} - t \frac{g_{k+1}^T s_k}{d_k^T y_k},$$

gde je $t > 0$ neka konstanta. Za tačno linijsko traženje, g_{k+1} je ortogonalno na s_k. Tako se za tačno linijsko traženje DL metod redukuje na HS i PRP metod.

Ponovo zahvaljujući primeru datom u [100], možemo zaključiti da DL metod može da ne konvergira za tačno linijsko traženje. Slično PRP+ metodu, u [30] je takođe modifikovan metod sa parametrom β_k^{DL}, da bi se obezbedila konvergencija, korišćenjem

$$\beta_k^{DL+} = \max\left\{ \frac{g_{k+1}^T y_k,}{d_k^T y_k}, 0 \right\} - t \frac{g_{k+1}^T s_k}{d_k^T y_k}.$$

Ako važe pretpostavke $P1$ i $P2$ i ako uslov dovoljnog pada (1.2.2) važi, u [30] je dokazano da je DL+, implementiran sa jakim Volfovim traženjem, globalno konvergentan.

4.3.1 Iteracije ograničene FR metodom

Moguće je dobiti globalnu konvergenciju metoda konjugovanih gradijenata, čiji je parametar β_k pogodno ograničen. Razmatramo metod (4.1.5)-(4.1.6), gde je β_k ma koji skalar, takav da je

$$|\beta_k| \leq \beta_k^{FR}, \tag{4.3.5}$$

za svako $k \geq 2$, a dužina koraka zadovoljava jake Volfove uslove (4.1.1)-(4.1.3), uz uslov da je $\sigma < \dfrac{1}{2}$. Primetimo da Zoutendijkov rezultat, Teorema 4.2.5 , važi u ovom slučaju, pošto jaki Volfovi uslovi impliciraju Volfove uslove (4.1.1)- (4.1.2).

Sledeći rezultat zasnovan je na radu [1] za FR metod, i neznatno je jači od onog koji je dat u [118].

Teorema 4.3.2. *Neka važe pretpostavke $P1$ i $P2$. Razmotrimo ma koji metod (4.1.5)-(4.1.6), gde β_k zadovoljava (4.3.5), a dužina koraka zadovoljava jake Volfove uslove (4.1.1)-(4.1.3), uz $0 < \delta < \sigma < \dfrac{1}{2}$. Onda važi*

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

Ova teorema sugeriše sledeći metod, koji je modifikacija PRP metoda, a koji je globalno konvergentan: za svako $k \geq 2$ neka je

$$\beta_k = \begin{cases} -\beta_k^{FR}, & \text{ako je } \beta_k^{PRP} < -\beta_k^{FR}, \\ \beta_k^{PRP}, & \text{ako je } |\beta_k^{PRP}| \leq \beta_k^{FR}, \\ \beta_k^{FR}, & \text{ako je } \beta_k^{PRP} > \beta_k^{FR}. \end{cases} \tag{4.3.6}$$

Ovaj novi metod dozvoljava negativne vrednosti za β_k. Ova strategija izbegava jednu od glavnih mana FR metoda, kao što ćemo sada diskutovati.

Podsetimo da se u izvesnim numeričkim testovima primećuje da FR metod sa tačnim linijskim traženjem ponekad usporava daleko od rešenja: koraci postaju veoma mali i ovo ponašanje može da se nastavi za vrlo veliki broj iteracija, osim ako se metod ne restartuje. Pomenuli smo da je ovakvo ponašanje zapaženo u [101], a autor rada [101] obezbeđuje i objašnjenje, pod pretpostavkom da se radi o tačnom linijskom traženju. Njegov argument se može proširiti na približno linijsko traženje.

Navodimo objašnjenje dato u [101].

Pretpostavimo da je u k-toj iteraciji jedan "loš" pravac traženja generisan, takav da je $\cos\theta_k \approx 0$, a $x_{k+1} \approx x_k$. Tada je $\|g_{k+1}\| \approx \|g_k\|$, i

$$\beta_{k+1}^{FR} \approx 1. \tag{4.3.7}$$

Štaviše, važi

$$\|g_{k+1}\| \approx \|g_k\| \ll \|d_k\|.$$

Iz poslednje relacije, (4.3.7) i (4.1.6), važi $\|d_{k+1}\| \approx \|d_k\| \gg \|g_{k+1}\|$, odakle sledi da je $\cos\theta_{k+1} \approx 0$.

PRP metod bi se ponašao potpuno drugačije od FR metoda u ovoj situaciji;

ako je $g_{k+1} \approx g_k$, tada je $\beta_{k+1}^{PRP} \approx 0$, odakle se dobija $\cos\theta_{k+1} \gg \cos\theta_k$. Tako se PRP "izvlači" iz te situacije.

Razmotrimo sada ponašanje metoda (4.3.6) pod ovim okolnostima. Videli smo da $\beta_{k+1}^{FR} \approx 1$, a $\beta_{k+1}^{PRP} \approx 0$ u ovom slučaju. Metod (4.3.6) će tako postaviti $\beta_{k+1} = \beta_{k+1}^{PRP}$, kao što je poželjno. Opet se uveravamo da modifikacija (4.3.6) izbegava neefikasnost PRP metoda.

Prethodna diskusija rasvetljava osobinu PRP metoda koju nema FR metod: kada je korak mali, β_{k+1}^{PRP} biće malo. Ova osobina biće bitna za kasniju analizu, a podsetimo da se radi o Pretpostavci $P3$.

Prirodno se pitamo da li je moguće ograničenje $|\beta_k| \leq \beta_k^{FR}$ zameniti sledećim ograničenjem

$$|\beta_k| \leq c\beta_k^{FR}, \tag{4.3.8}$$

gde je $c > 1$ neka pogodno izabrana konstanta. U ovom slučaju globalna konvergencija nije se mogla dokazati, iako se, ipak, može pokazati da se osobina pada pravaca traženja može još uvek dobiti pod uslovom da je $\sigma < \dfrac{1}{2c}$.

Glava 5

Novi hibridni metod konjugovanih gradijenata

5.1 Uvod

Podsetimo da hibridni metodi konjugovanih gradijenata kombinuju različite metode konjugovanih gradijenata da poboljšaju ponašanje ovih metoda i da izbegnu fenomen zaglavljivanja. Ovi metodi mogu izbeći pojavu malih koraka koja karakteriše fenomen zaglavljivanja.

U radu [118] uveden je jedan od prvih hibridnih algoritama konjugovanih gradijenata, u kome se parametar β_k izračunava sledećom formulom:

$$\beta_k^{TaS} = \begin{cases} \beta_k^{PRP} = \dfrac{g_{k+1}^T y_k}{\|g_k\|^2}, \text{ ako } 0 \leq \beta_k^{PRP} \leq \beta_k^{FR}, \\ \beta_k^{FR} = \dfrac{\|g_{k+1}\|^2}{\|g_k\|^2}, \text{ inače.} \end{cases} \tag{5.1.1}$$

Kada dođe do zaglavljivanja iteracija, metod koji je dat u [118] koristi PRP metod, imajući u vidu da PRP metod ima osobinu automatskog restarta koja je ključna u rešavanju problema zaglavljivanja.

Naime, kada je korak s_k mali, onda veličina y_k u formuli (4.2.10) teži nuli, vrednost parametra β_k^{PRP} postaje sve manja, a istovremeno pravac d_{k+1} postaje vrlo blizak pravcu najbržeg pada $-g_{k+1}$.

U radu [85] predstavljen je jedan drugi hibridni metod konjugovanih gradijenata sa sledećim izborom parametra β_k:

$$\beta_k^{HuS} = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}. \tag{5.1.2}$$

Takođe, ako dođe do zaglavljivanja u metodu datom u radu [85], koristi se PRP metod.

Parametar β_k biramo tako da bude konveksna kombinacija parametara (4.2.11) i (4.2.12).

Koristićemo dobre osobine konvergencije metoda CD i u isto vreme, dobro ponašanje u praksi metoda LS.

Kombinacija metoda LS i CD može dovesti do sledećeg hibridnog metoda [5], [119]:

$$\beta_k^{LS-CD} = \max\{0, \min\{\beta_k^{LS}, \beta_k^{CD}\}\}. \tag{5.1.3}$$

U radovima [5], [73] primećeno je da za tačno linijsko traženje, LS metod jeste identičan PRP metodu, a CD metod je identičan FR metodu. Dalje je u radu [5], izveden zaključak da se hibridni LS-CD metod sa parametrom (5.1.3) i u prisustvu tačnog linijskog traženja, ponaša slično kao hibridni metod iz rada [85].

Koristimo približno linijsko traženje.

5.2 Konveksna kombinacija

Ovde definišemo sledeći parametar konjugovanog gradijenta

$$\beta_k^{hyb} = (1-\theta_k)\cdot\beta_k^{LS} + \theta_k\cdot\beta_k^{CD}. \tag{5.2.1}$$

Sada je pravac d_k dat sledećom relacijom:

$$d_0 = -g_0,\ \ d_{k+1} = -g_{k+1} + \beta_k^{hyb} s_k. \tag{5.2.2}$$

Parametar θ_k je skalarni parametar koji treba odrediti.

Imajući u vidu relacije (4.2.11) i (4.2.12), relacija (5.2.1) postaje

$$\beta_k^{hyb} = (1-\theta_k)\cdot\frac{g_{k+1}^T y_k}{-g_k^T s_k} + \theta_k\cdot\frac{g_{k+1}^T g_{k+1}}{-g_k^T s_k}, \tag{5.2.3}$$

te relacija (5.2.2) postaje

$$d_{k+1} = -g_{k+1} + (1-\theta_k)\frac{g_{k+1}^T y_k}{-g_k^T s_k}\cdot s_k + \theta_k\cdot\frac{g_{k+1}^T g_{k+1}}{-g_k^T s_k}\cdot s_k. \tag{5.2.4}$$

Naći ćemo vrednost parametra θ_k tako da važi uslov konjugacije

$$y_k^T d_{k+1} = 0. \tag{5.2.5}$$

Najpre množimo obe strane relacije (5.2.4) izrazom y_k^T sa leve strane:

$$y_k^T d_{k+1} = -y_k^T g_{k+1} + (1-\theta_k)\frac{g_{k+1}^T y_k}{-g_k^T s_k}\cdot y_k^T s_k + \theta_k\frac{\|g_{k+1}\|^2}{-g_k^T s_k}\cdot y_k^T s_k.$$

Koristeći (5.2.5), dobijamo

$$0 = -y_k^T g_{k+1} + (1-\theta_k)\cdot\frac{g_{k+1}^T y_k}{-g_k^T s_k}\cdot y_k^T s_k + \theta_k\frac{\|g_{k+1}\|^2}{-g_k^T s_k} y_k^T s_k.$$

Dalje važi

$$y_k^T g_{k+1} = \frac{(g_{k+1}^T y_k)(y_k^T s_k)}{-g_k^T s_k} - \theta_k \frac{(g_{k+1}^T y_k)(y_k^T s_k)}{-g_k^T s_k} + \theta_k \cdot \frac{(g_{k+1}^T g_{k+1})(y_k^T s_k)}{-g_k^T s_k},$$

te koristeći relaciju $y_k = g_{k+1} - g_k$, dobijamo

$$\begin{aligned} y_k^T g_{k+1} = {} & \frac{(g_{k+1}^T y_k)(y_k^T s_k)}{-g_k^T s_k} - \theta_k \cdot \frac{(g_{k+1}^T g_{k+1}) \cdot (y_k^T s_k)}{-g_k^T s_k} \\ & + \theta_k \cdot \frac{(g_{k+1}^T g_k) \cdot (y_k^T s_k)}{-g_k^T s_k} + \theta_k \frac{(g_{k+1}^T g_{k+1})(y_k^T s_k)}{-g_k^T s_k}. \end{aligned}$$

Sada važi

$$\theta_k \cdot \frac{(g_{k+1}^T g_k) \cdot (y_k^T s_k)}{-g_k^T s_k} + \frac{(g_{k+1}^T y_k)(y_k^T s_k)}{-g_k^T s_k} = y_k^T g_{k+1}.$$

Treba da nađemo θ_k, te računamo:

$$\theta_k \cdot \frac{(g_{k+1}^T g_k) \cdot (y_k^T s_k)}{-g_k^T s_k} = y_k^T g_{k+1} - \frac{(g_{k+1}^T y_k)(y_k^T s_k)}{-g_k^T s_k},$$

t.j.

$$\theta_k \cdot \frac{(g_{k+1}^T g_k) \cdot (y_k^T s_k)}{-g_k^T s_k} = \frac{(y_k^T g_{k+1})(-g_k^T s_k) - (g_{k+1}^T y_k)(y_k^T s_k)}{-g_k^T s_k}.$$

Konačno dobijamo

$$\theta_k = \frac{(y_k^T g_{k+1})(-g_k^T s_k) - (g_{k+1}^T y_k)(y_k^T s_k)}{(g_{k+1}^T g_k)(y_k^T s_k)},$$

odnosno

$$\theta_k = \frac{(g_{k+1}^T y_k)(-g_k^T s_k - y_k^T s_k)}{(g_{k+1}^T g_k)(y_k^T s_k)}.$$

Koristeći ponovo relaciju $y_k = g_{k+1} - g_k$, dobijamo:

$$\theta_k = \frac{(g_{k+1}^T y_k) \cdot (-g_k^T s_k - g_{k+1}^T s_k + g_k^T s_k)}{(g_{k+1}^T g_k)(y_k^T s_k)},$$

odakle zaključujemo da je

$$\theta_k = \frac{-(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{(g_{k+1}^T g_k)(y_k^T s_k)}. \tag{5.2.6}$$

Konačno, imajući u vidu relaciju (5.2.6), definišemo:

$$\theta_k = \begin{cases} \dfrac{-(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{(g_{k+1}^T g_k)(y_k^T s_k)}, & 0 \le \dfrac{-(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{(g_{k+1}^T g_k)(y_k^T s_k)} \le 1; \\ 0, & \dfrac{-(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{(g_{k+1}^T g_k)(y_k^T s_k)} < 0; \\ 1, & \dfrac{-(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{(g_{k+1}^T g_k)(y_k^T s_k)} > 1. \end{cases} \tag{5.2.7}$$

Osobina pada je veoma važna da bi jedan iterativni metod bio globalno konvergentan, naročito ako je u pitanju metod konjugovanog gradijenta [47].

U [101] je dokazano da za nelinearnu funkciju u opštem slučaju, ukoliko se radi o tačnom linijskom traženju i ako važi:
(a) korak s_k teži nuli,
(b) linijsko traženje je tačno,
(c) pretpostavka P1 važi,
onda je PRP metod globalno konvergentan.

Na drugoj strani, u [100] je pokazano kasnije, korišćenjem trodimenzionalnog primera, da sa tačnim linijskim traženjem, PRP metod može da divergira. Tako je pretpostavka da veličina koraka teži nuli, neophodna za konvergenciju [73].

Pošto je LS metod identičan PRP metodu za tačno linijsko traženje, iako nije mnogo istraživanja posvećeno ovom metodu, očekuje se da bi tehnike, razvijene za analizu PRP metoda trebalo da se primene i na LS metod [73].

Motivisani smo ovim činjenicama da dokažemo Teoremu 5.3.1 i Teoremu 5.4.1.

5.3 Algoritam LSCD

Algoritam 5.3.1. *Ulazni parametri:* $\epsilon > 0$, x_0, $k := 0$, $0 < \delta \leq \sigma < 1$, $\beta < 1$.

Korak 1. Ako važi $\|g_k\| \leq \epsilon$, *STOP.*

Korak 2. Odrediti najveće j_k, *tako da je za* $t_k = \beta^{j_k}$ *zadovoljeno*

$$f(x_k + t_k d_k) - f(x_k) \leq \delta t_k g_k^T d_k,$$
$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k.$$

Odrediti $x_{k+1} = x_k + t_k d_k$.

Korak 3. Ako je $(g_{k+1}^T g_k)(y_k^T s_k) = 0$, *onda postaviti* $\theta_k = 0$, *inače izračunati* θ_k *kao u (5.2.7).*

Korak 4. Izračunati β_k^{hyb} *prema formuli*

$$\beta_k^{hyb} = (1 - \theta_k) \cdot \frac{g_{k+1}^T y_k}{-g_k^T s_k} + \theta_k \cdot \frac{g_{k+1}^T g_{k+1}}{-g_k^T s_k}. \tag{5.3.1}$$

Korak 5. Ako važi

$$|g_{k+1}^T g_k| > a\|g_{k+1}\|^2, \tag{5.3.2}$$

onda postaviti $d_{k+1} = -g_{k+1}$, *inače* $d_{k+1} = -g_{k+1} + \beta_k^{hyb} \cdot s_k$.

Korak 6. $k := k + 1$, *ići na Korak 1.*

Teorema 5.3.1. *Neka su zadovoljene pretpostavke P_1 i P_2. Neka je konstanta a u Algoritmu 5.3.1 takva da važi*

$$0 < a < \frac{1}{\sigma} - 1. \tag{5.3.3}$$

Onda je Algoritam 5.3.1 dobro definisan i d_k zadovoljava uslov dovoljnog pada za svako k.

Dokaz. Na osnovu Leme 1.3.1 znamo da je Korak 2 Algoritma 5.3.1 dobro definisan ako je d_k opadajući pravac. Pokazaćemo da d_k zadovoljava uslov dovoljnog pada, iz čega će slediti i da je d_k opadajući pravac.

Za $k = 0$, važi $d_0 = -g_0$, te je $g_0^T d_0 = -\|g_0\|^2$, pa uslov dovoljnog pada važi za $k = 0$.

Dalje važi

$$d_{k+1} = -g_{k+1} + \beta_k^{hyb} s_k,$$

to jest,

$$d_{k+1} = -g_{k+1} + ((1-\theta_k)\beta_k^{LS} + \theta_k \beta_k^{CD}) s_k.$$

Možemo pisati

$$d_{k+1} = -(\theta_k g_{k+1} + (1-\theta) g_{k+1}) + ((1-\theta_k)\beta_k^{LS} + \theta_k \beta_k^{CD}) s_k.$$

Sledi

$$d_{k+1} = \theta_k(-g_{k+1} + \beta_k^{CD} s_k) + (1-\theta_k)(-g_{k+1} + \beta_k^{LS} s_k),$$

odakle sledi

$$d_{k+1} = \theta_k d_{k+1}^{CD} + (1-\theta_k) d_{k+1}^{LS}. \tag{5.3.4}$$

Množenjem (5.3.4) pomoću g_{k+1}^T sa leve strane, dobijamo

$$g_{k+1}^T d_{k+1} = \theta_k g_{k+1}^T d_{k+1}^{CD} + (1-\theta_k) g_{k+1}^T d_{k+1}^{LS}. \tag{5.3.5}$$

Neka je, najpre, $\theta_k = 0$. Onda d_{k+1} zadovoljava uslov dovoljnog pada ako i samo ako d_{k+1}^{LS} zadovoljava uslov dovoljnog pada. Podsetimo da važi

$$d_{k+1}^{LS} = -g_{k+1} + \beta_k^{LS} s_k$$

$$\Rightarrow g_{k+1}^T d_{k+1}^{LS} \leq -\|g_{k+1}\|^2 + \frac{(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{-g_k^T s_k}. \tag{5.3.6}$$

Dokazaćemo da važi $\left| \frac{(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{-g_k^T s_k} \right| \leq \mu \|g_{k+1}\|^2$, pri čemu je $0 < \mu < 1$.

Razmatramo apsolutnu vrednost izraza

$$T := \frac{(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{-g_k^T s_k}.$$

Dakle,

$$|T| = \left| \frac{(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{-g_k^T s_k} \right| \le \left| \frac{g_{k+1}^T s_k}{-g_k^T s_k} \right| |g_{k+1}^T y_k|.$$

Na osnovu drugog jakog Volfovog uslova, važi

$$\left| \frac{g_{k+1}^T s_k}{-g_k^T s_k} \right| \le \sigma.$$

Sada važi

$$|T| \le \sigma |g_{k+1}^T y_k|. \tag{5.3.7}$$

Ako je zadovoljena relacija (5.3.2), tada je $d_{k+1} = -g_{k+1}$, te važi $g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2$, čime je dokazano da d_{k+1} zadovoljava uslov dovoljnog pada.

Ukoliko relacija (5.3.2) nije zadovoljena, tada važi

$$|g_{k+1}^T g_k| \le a \|g_{k+1}\|^2. \tag{5.3.8}$$

Pošto je $y_k = g_{k+1} - g_k$, iz relacije (5.3.7) dalje dobijamo

$$\begin{aligned} |T| &\le \sigma |g_{k+1}^T y_k| && (5.3.9)\\ &= \sigma |g_{k+1}^T (g_{k+1} - g_k)| && (5.3.10)\\ &\le \sigma \|g_{k+1}\|^2 + \sigma |g_{k+1}^T g_k|, && (5.3.11) \end{aligned}$$

odakle, primenom relacije (5.3.8), dobijamo da važi

$$|T| \le \sigma \|g_{k+1}\|^2 + \sigma a \|g_{k+1}\|^2,$$

pa, imajući u vidu relaciju (5.3.3), možemo pisati

$$|T| \le \mu \|g_{k+1}\|^2, \text{ pri čemu je } 0 < \mu = \sigma(1+a) < 1.$$

Sada, korišćenjem relacije (5.3.6), dobijamo

$$g_{k+1}^T d_{k+1}^{LS} \le -\|g_{k+1}\|^2 + \mu \|g_{k+1}\|^2,$$

i

$$g_{k+1}^T d_{k+1}^{LS} \le -(1-\mu) \|g_{k+1}\|^2.$$

Označimo $K_1 = (1-\mu)$; tada možemo pisati

$$g_{k+1}^T d_{k+1}^{LS} \le -K_1 \|g_{k+1}\|^2. \tag{5.3.12}$$

Neka je sada $\theta_k = 1$. Onda d_{k+1} zadovoljava uslov dovoljnog pada ako i samo ako d_{k+1}^{CD} zadovoljava uslov dovoljnog pada.

Dalje dokazujemo da za CD metod uslov dovoljnog pada važi u prisustvu jakih Volfovih uslova, kao što je napomenuto u radu [73].

Za $k = 0$ dokaz je trivijalan, imajući u vidu da je $d_0^{CD} = -g_0$ i zbog toga $g_0^T d_0^{CD} = -\|g_0\|^2$.

Imajući u vidu da je

$$d_{k+1}^{CD} = -g_{k+1} + \beta_k^{CD} s_k, \tag{5.3.13}$$

množenjem relacije (5.3.13) izrazom g_{k+1}^T sa leve strane i korišćenjem izraza (CD), dobijamo $g_{k+1}^T d_{k+1}^{CD} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{-g_k^T s_k} \cdot (g_{k+1}^T s_k)$, odakle dobijamo

$$g_{k+1}^T d_{k+1}^{CD} = -\|g_{k+1}\|^2 (1 - \frac{g_{k+1}^T s_k}{-g_k^T s_k}) = -\|g_{k+1}\|^2 \cdot \frac{-g_k^T s_k - g_{k+1}^T s_k}{-g_k^T s_k}.$$

Korišćenjem jakih Volfovih uslova, sada važi

$$\frac{-g_k^T s_k - g_{k+1}^T s_k}{-g_k^T s_k} \geq \frac{(\sigma - 1) g_k^T s_k}{-g_k^T s_k} = 1 - \sigma > 0.$$

Sada važi

$$g_{k+1}^T d_{k+1}^{CD} \leq -(1 - \sigma)\|g_{k+1}\|^2.$$

Uvedimo oznaku $1 - \sigma = K_2 > 0$.

Dakle,

$$g_{k+1}^T d_{k+1}^{CD} \leq -K_2 \|g_{k+1}\|^2. \tag{5.3.14}$$

Pretpostavimo sada da je $0 < \theta_k < 1$, to jest, $0 < a_1 \leq \theta_k \leq a_2 < 1$.

Iz relacije (5.3.5), sada zaključujemo

$$g_{k+1}^T d_{k+1} \leq a_1 g_{k+1}^T d_{k+1}^{CD} + (1 - a_2) g_{k+1}^T d_{k+1}^{LS}. \tag{5.3.15}$$

Označimo $K = a_1 K_1 + (1 - a_2) K_2$, te konačno dobijamo

$$g_{k+1}^T d_{k+1} \leq -K \|g_{k+1}\|^2. \tag{5.3.16}$$

□

Analogno, u radu [52] razmatrana je konveksna kombinacija FR i PRP metoda konjugovanih gradijenata, pri čemu je parametar β_k^{hyb} dat sledećom relacijom

$$\beta_k^{hyb} = (1 - \theta_k) \cdot \beta_k^{PRP} + \theta_k \cdot \beta_k^{FR}.$$

5.4 Analiza konvergencije

Neka važe pretpostavke P_1 i P_2.

Pod ovim pretpostavkama postoji konstanta $\Gamma \geq 0$, takva da važi

$$\|g(x)\| \leq \Gamma \tag{5.4.1}$$

za svako $x \in \mathcal{L}$ [5].

U radu [39] dokazano je da za ma koji metod konjugovanih gradijenata sa jakim Volfovim linijskim traženjem, važi:

Lema 5.4.1. *[39] Neka važe pretpostavke P_1 i P_2.Razmotrimo metod (4.1.5), (4.1.6), gde je d_k opadajući pravac, a t_k dobijeno jakim Volfovim linijskim traženjem. Ako važi*

$$\sum_{k\geq 1} \frac{1}{\|d_k\|^2} = \infty, \tag{5.4.2}$$

tada

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{5.4.3}$$

Teorema 5.4.1. *Razmotrimo iterativni metod, definisan Algoritmom 5.3.1. Neka važe pretpostavke P_1 i P_2. Pretpostavimo da važe jaki Volfovi uslovi (4.1.1)-(4.1.3). Tada je ili $g_k = 0$, za neko k, ili*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{5.4.4}$$

Dokaz. Pretpostavimo da je $g_k \neq 0$, za svako k. Tada treba da dokažemo (5.4.4).

Pretpostavimo suprotno: (5.4.4) ne važi. Onda postoji konstanta $c > 0$, takva da važi

$$\|g_k\| \geq c, \;\text{ za svako }\; k. \tag{5.4.5}$$

Neka D označava dijametar nivo skupa $\mathcal{L}$.

Iz (5.3.1), dobijamo

$$\|\beta_k^{hyb}\| \leq |\beta_k^{LS}| + |\beta_k^{CD}|. \tag{5.4.6}$$

Važi

$$|\beta_k^{LS}| = \left|\frac{g_{k+1}^T y_k}{-g_k^T s_k}\right| \leq \frac{\|g_{k+1}\|\|y_k\|}{|-g_k^T s_k|} \leq \frac{\Gamma\|y_k\|}{|-g_k^T s_k|},$$

pri čemu smo koristili (5.4.1). Primenom pretpostavke P_1, dobijamo

$$|\beta_k^{LS}| \leq \frac{\Gamma L\|s_k\|}{|-g_k^T s_k|}.$$

Pošto je $\|s_k\| \leq D$, sledi

$$|\beta_k^{LS}| \leq \frac{\Gamma L D}{|-g_k^T s_k|}.$$

Na osnovu Teoreme 5.3.1, znamo da je za LS metod zadovoljen uslov dovoljnog pada, pa je moguće zadovoljiti jake Volfove uslove. Neka je $t_k = \beta^{j_k}$ niz dužina koraka tako da je

$$f(x_k + t_k d_k) - f(x_k) \leq \delta t_k g_k^T d_k.$$

Analogno dokazu Teoreme 3.2.1, može se pokazati da je $t_k \geq t* > 0$. To implicira da važi

$$\|s_k\| = \|t_k d_k\| \geq t * \|d_k\|,$$

pa iz uslova dovoljnog pada sledi

$$|\beta_k^{LS}| \leq \frac{\Gamma L D}{K_1 \|g_{k+1}\|^2}, \tag{5.4.7}$$

za neko $K_1 > 0$.

Sada koristimo (5.4.5) i dobijamo

$$|\beta_k^{LS}| \leq \frac{\Gamma L D}{K_1 c^2}. \tag{5.4.8}$$

Pomoću relacije (5.4.8), dobijamo

$$\|d_{k+1}^{LS}\| \leq \|g_{k+1}\| + \frac{\Gamma L D}{K_1 c^2} D \leq \Gamma + \frac{\Gamma L D^2}{K_1 c^2}. \tag{5.4.9}$$

Na drugoj strani,

$$\|d_{k+1}^{CD}\| \leq \|g_{k+1}\| + |\beta_k^{CD}| \|s_k\| \leq \Gamma + |\beta_k^{CD}| D. \tag{5.4.10}$$

Takođe važi

$$|\beta_k^{CD}| = \frac{\|g_{k+1}\|^2}{|-g_k^T s_k|} \leq \frac{\Gamma^2}{|-g_k^T s_k|}. \tag{5.4.11}$$

Uslov dovoljnog pada važi i za CD metod, tako da, analogno, dobijamo

$$|\beta_k^{CD}| \leq \frac{\Gamma^2}{K_2 \|g_k\|^2} \leq \frac{\Gamma^2}{K_2 c^2},$$

pa važi

$$\|d_{k+1}^{CD}\| \leq \Gamma + \frac{\Gamma^2 D}{K_2 c^2}. \tag{5.4.12}$$

Primenom (5.3.4), nalazimo da važi

$$\|d_{k+1}\| \leq \Gamma + \frac{\Gamma L D^2}{K_1 c^2} + \Gamma + \frac{\Gamma^2 D}{K_2 c^2},$$

odakle je

$$\sum_{k \geq 1} \frac{1}{\|d_k\|^2} = \infty, \tag{5.4.13}$$

te primenom Leme 5.4.1, zaključujemo da važi $\liminf_{k \to \infty} \|g_k\| = 0$. Ovo je kontradikcija sa (5.4.5), te smo dokazali (5.4.4). $\square$

Glava 6

Numerički eksperimenti

Implementacija svih razmatranih algoritama rađena je u paketu Mathematica. Kriterijum upoređivanja je vreme rada centralne procesorske jedinice (CPU). Eksperimenti su izvršeni na skupu test funkcija, predstavljenih u [13]. Razmatramo najpre numeričke rezultate, koji se odnose na Poglavlje 3, korišćenjem dimenzija $n = 10$, u Tabeli 1. i $n = 100$ u Tabeli 2. Dalje razmatramo numeričke rezultate, koji se odnose na Poglavlje 5, korišćenjem dimenzija $n = 10$, u Tabeli 3., a $n = 100$ u Tabeli 4. Korišćena je vrednost parametra $\epsilon = 10^{-6}$. U Algoritmu 5.3.1 korišćena je vrednost $a = 0.2$. Početna iteracija označena je sa x_0. Slede odgovarajuće tabele.

Tabela 1. $n = 10$

Funkcija	x_0	GDQN	RGDQN
Extended Penalty	$[1, 2, ..., n]$	0.156	0.046
Perturbed Quadratic	$[0.5, 0.5, ..., 0.5]$	0.141	0.078
Raydan 1	$[1, 1, ..., 1]$	0.172	0.063
Raydan 2	$[1, 1, ..., 1]$	0.063	0.047
Diagonal1	$[\frac{1}{n}, \frac{1}{n}, ... \frac{1}{n}]$	0.124	0.016
Diagonal2	$[1, \frac{1}{2}, ... \frac{1}{n}]$	0.156	0.047
Diagonal3	$[1, 1, ..., 1]$	0.188	0.078

Generalized Tridiagonal-1	$[2, 2, ..., 2]$	0.14	0.047
Hager	$[1, 1, ..., 1]$	0.14	0.062
Extended Tridiagonal-1	$[2, 2, ..., 2]$	0.172	0.046
Extended Three Exponential Terms	$[0.1, 0.1, ..., 0.1]$	0.125	0.063
Diagonal4	$[1, 1, ..., 1]$	0.202	0.047
Diagonal5	$[1.1, 1.1, ..., 1.1]$	0.172	0.063
Extended Himmelblau	$[1, 1, ..., 1]$	0.171	0.047
Extended PSC1	$[3, 0.1, ..., 3, 0.1]$	0.171	0.063
Full Hessian FH2	$[0.01, 0.01, ..., 0.01]$	0.172	0.063
Extended Block Diagonal BD1	$[0.1, 0.1, ..., 0.1]$	0.172	0.062
Quadratic QF1	$[1, 1, ..., 1]$	0.218	0.062
Extended Quadratic Penalty QP1	$[1, 1, ..., 1]$	0.172	0.078
Quadratic QF2	$[0.5, 0.5, ..., 0.5]$	0.187	0.031
Extended EP1	$[1.5, 1.5, ..., 1.5]$	0.14	0.063
Extended Tridiagonal-2	$[1, 1, ..., 1]$	0.202	0.047
Tridia	$[1, 1, ..., 1]$	0.109	0.047
Arwhead	$[1, 1, ..., 1]$	0.203	0.047
Dqdrtic	$[3, 3, ..., 3]$	0.219	0.047
Quartc	$[2, 2, ..., 2]$	0.172	0.063
Dixon3dq	$[-1, -1, ..., -1]$	0.109	0.047
Biggsb1	$[0, 0, ..., 0]$	0.156	0.062
Generalized Quartic	$[1, 1, ..., 1]$	0.218	0.078
Diagonal 7	$[1, 1, ..., 1]$	0.187	0.046
Diagonal 8	$[1, 1, ..., 1]$	0.171	0.062
Full Hessian FH3	$[1, 1, ..., 1]$	0.218	0.078
Himmelbg	$[1.5, 1.5, ..., 1.5]$	0.172	0.047
Extended Freudenstein Roth	$[0.5, -2, ..., 0.5, -2]$	0.156	0.093

Extended Trigonometric	$[0.2, 0.2, ..., 0.2]$	0.187	0.11
Extended Rosenbrock	$[-1.2, 1, ..., -1.2, 1]$	0.172	0.032
Generalized Rosenbrock	$[-1.2, 1, ..., -1.2, 1]$	0.156	0.078
Extended White Holst	$[-1.2, 1, ..., -1.2, 1]$	0.171	0.031
Generalized White Holst	$[-1.2, 1, ..., -1.2, 1]$	0.125	0.031
Extended Beale	$[1, 0.8, ..., 1, 0.8]$	0.078	0.078
Perturbed Quadratic	$[0.5, 0.5, ..., 0.5]$	0.124	0.062
Extended Powell	$[3, -1, 0.1, ..., 3, -1, 0.1]$	0.171	0.063
Extended Maratos	$[1.1, 0.1, ..., 1.1, 0.1]$	0.172	0.078
Extended Cliff	$[0, -1, ..., 0, -1]$	0.141	0.031
Extended Wood	$[-3, -1, ..., -3, -1]$	0.14	0.063
Extended Hiebert	$[0, 0, ..., 0]$	0.124	0.062
Extended Quadratic Penalty QP2	$[1, 1, ..., 1]$	0.171	0.078
Extended Quadratic Exponential EP1	$[1.5, 1.5, ..., 1.5]$	0.156	0.078
FLETCBV3	$[\frac{1}{n+1}, \frac{2}{n+1}, ..., \frac{n}{n+1},]$	0.171	0.078
FLETCHCR	$[0, 0, ..., 0]$	0.125	0.078
BDQRTIC	$[1, 1, ..., 1]$	0.125	0.062
ARGLINB	$[1, 1, ..., 1]$	0.172	0.078
NONDIA	$[-1, -1, ..., -1]$	0.109	0.078
NONDQUAR	$[1, -1, ..., 1, -1]$	0.156	0.078
EG2	$[1, 1, ..., 1]$	0.156	0.062
CURLY20	$\left[\frac{0.001}{n+1}, \frac{0.001}{n+1}, ..., \frac{0.001}{n+1}\right]$	0.156	0.031
DIXMAANA 1	$[2, 2, ..., 2]$	0.156	0.078
DIXMAANA 2	$[2, 2, ..., 2]$	0.141	0.046
DIXMAANA 3	$[2, 2, ..., 2]$	0.156	0.062
DIXMAANA 4	$[2, 2, ..., 2]$	0.156	0.078
DIXMAANA 5	$[2, 2, ..., 2]$	0.172	0.062
DIXMAANA 6	$[2, 2, ..., 2]$	0.094	0.031
DIXMAANA 7	$[2, 2, ..., 2]$	0.171	0.093

DIXMAANA 8	$[2, 2, ..., 2]$	0.109	0.046
DIXMAANA 9	$[2, 2, ..., 2]$	0.094	0.032
DIXMAANA 10	$[2, 2, ..., 2]$	0.187	0.062
DIXMAANA 11	$[2, 2, ..., 2]$	0.172	0.062
DIXMAANA 12	$[2, 2, ..., 2]$	0.156	0.031
Partial Perturbed Quadratic	$[0.5, 0.5, ..., 0.5]$	0.141	0.078
Broyden Tridiagonal	$[-1, -1, ..., -1]$	0.172	0.109
Almost Perturbed Quadratic	$[0.5, 0.5, ..., 0.5]$	0.156	0.062
Perturbed Tridiagonal Quadratic	$[0.5, 0.5, ..., 0.5]$	0.094	0.047
Staircase 1	$[1, 1, ..., 1]$	0.093	0.047
Staircase 2	$[0, 0, ..., 0]$	0.156	0.078
LIARWHD	$[4, 4, ..., 4]$	0.156	0.062
POWER	$[1, 1, ..., 1]$	0.094	0.078
ENGVAL1	$[2, 2, ..., 2]$	0.141	0.047
CRAGGLVY	$[1, 2, ..., 2]$	0.109	0.063
EDENSCH	$[0, 0, ..., 0]$	0.187	0.063
INDEF	$[\frac{1}{n+1}, \frac{1}{n+1}, ..., \frac{1}{n+1}]$	0.156	0.078
CUBE	$[-1.2, 1, ..., -1.2, 1]$	0.141	0.062
EXPLIN1	$[0, 0, ..., 0]$	0.141	0.062
EXPLIN2	$[0, 0, ..., 0]$	0.156	0.047
ARGLINC	$[1, 1, ..., 1]$	0.109	0.078
BDEXP	$[1, 1, ..., 1]$	0.156	0.062
HARKERP2	$[1, 2, ..., n]$	0.156	0.078
GENHUMPS	$[-506, 506.2, ..., 506.2]$	0.093	0.062
MCCORMCK	$[1, 1, ..., 1]$	0.171	0.078
NONSCOMP	$[3, 3, ..., 3]$	0.125	0.062
VARDIM	$[1 - \frac{1}{n}, 1 - \frac{2}{n}, ..., 0]$	0.14	0.031
Diagonal 6	$[1, 1, ..., 1]$	0.093	0.032
SINQUAD	$[0.1, 0.1, ..., 0.1]$	0.172	0.063
Extended DENSCHNB	$[1, 1, ..., 1]$	0.171	0.093
Extended DENSCHNF	$[2, 0, ..., 2, 0]$	0.125	0.078
COSINE	$[1, 1, ..., 1]$	0.171	0.078
SINE	$[1, 1, ..., 1]$	0.156	0.062
SINCOS	$[3, 0.1, ..., 3, 0.1]$	0.14	0.078
Diagonal 9	$[1, 1, ..., 1]$	0.109	0.078
Average		0.1515	0.0615

Tabela 2. $n = 100$

Funkcija	x_0	GDQN	RGDQN
Extended Penalty	$[1, 2, ..., n]$	74.241	50.716
Perturbed Quadratic	$[0.5, 0.5, ..., 0.5]$	80.325	50.529
Raydan 1	$[1, 1, ..., 1]$	65.926	47.362
Raydan 2	$[1, 1, ..., 1]$	11.497	9.049
Diagonal1	$[\frac{1}{n}, \frac{1}{n}, ... \frac{1}{n}]$	12.059	9.313
Diagonal2	$[1, \frac{1}{2}, ... \frac{1}{n}]$	85.489	52.041
Diagonal3	$[1, 1, ..., 1]$	87.111	51.995
Generalized Tridiagonal-1	$[2, 2, ..., 2]$	86.19	23.228
Hager	$[1, 1, ..., 1]$	86.986	23.696
Extended Tridiagonal-1	$[2, 2, ..., 2]$	86.94	24.103
Extended Three Exponential Terms	$[0.1, 0.1, ..., 0.1]$	87.204	23.572
Diagonal4	$[1, 1, ..., 1]$	86.908	24.056
Diagonal5	$[1.1, 1.1, ..., 1.1]$	87.017	24.289
Extended Himmelblau	$[1, 1, ..., 1]$	87.236	23.697
Extended PSC1	$[3, 0.1, ..., 3, 0.1]$	87.532	23.931
Full Hessian FH2	$[0.01, 0.01, ..., 0.01]$	87.33	23.899
Extended Block Diagonal BD1	$[0.1, 0.1, ..., 0.1]$	86.674	24.648
Quadratic QF1	$[1, 1, ..., 1]$	86.986	24.133
Extended Quadratic Penalty QP1	$[1, 1, ..., 1]$	87.329	23.837
Quadratic QF2	$[0.5, 0.5, ..., 0.5]$	87.236	24.056
Extended EP1	$[1.5, 1.5, ..., 1.5]$	87.111	24.554
Extended Tridiagonal-2	$[1, 1, ..., 1]$	86.643	24.398
Tridia	$[1, 1, ..., 1]$	88.063	24.352

Arwhead	$[1, 1, ..., 1]$	86.799	24.398
Dqdrtic	$[3, 3, ..., 3]$	87.361	23.946
Quartc	$[2, 2, ..., 2]$	87.376	24.118
Dixon3dq	$[-1, -1, ..., -1]$	86.783	24.087
Biggsb1	$[0, 0, ..., 0]$	86.924	24.305
Generalized Quartic	$[1, 1, ..., 1]$	87.563	24.055
Diagonal 7	$[1, 1, ..., 1]$	87.22	24.399
Diagonal 8	$[1, 1, ..., 1]$	87.329	24.289
Full Hessian FH3	$[1, 1, ..., 1]$	86.939	24.321
Himmelbg	$[1.5, 1.5, ..., 1.5]$	87.751	24.102
Extended Freudenstein Roth	$[0.5, -2, ..., 0.5, -2]$	87.111	23.962
Extended Trigonometric	$[0.2, 0.2, ..., 0.2]$	87.454	24.445
Extended Rosenbrock	$[-1.2, 1, ..., -1.2, 1]$	87.985	24.18
Generalized Rosenbrock	$[-1.2, 1, ..., -1.2, 1]$	87.594	24.711
Extended White Holst	$[-1.2, 1, ..., -1.2, 1]$	87.173	24.476
Generalized White Holst	$[-1.2, 1, ..., -1.2, 1]$	87.204	23.946
Extended Beale	$[1, 0.8, ..., 1, 0.8]$	87.36	24.352
Perturbed Quadratic	$[0.5, 0.5, ..., 0.5]$	88.686	24.414
Extended Powell	$[3, -1, 0.1, ..., 3, -1, 0.1]$	87.829	24.851
Extended Maratos	$[1.1, 0.1, ..., 1.1, 0.1]$	87.407	23.993
Extended Cliff	$[0, -1, ..., 0, -1]$	87.111	23.805
Extended Wood	$[-3, -1, ..., -3, -1]$	87.11	24.336
Extended Hiebert	$[0, 0, ..., 0]$	87.938	24.29
Extended Quadratic Penalty QP2	$[1, 1, ..., 1]$	88.39	25.226
Extended Quadratic Exponential EP1	$[1.5, 1.5, ..., 1.5]$	87.641	24.789

FLETCBV3	$[\frac{1}{n+1}, \frac{2}{n+1}, ..., \frac{n}{n+1},]$	88.639	23.759
FLETCHCR	$[0, 0, ..., 0]$	87.719	24.336
BDQRTIC	$[1, 1, ..., 1]$	88.218	24.024
ARGLINB	$[1, 1, ..., 1]$	88.063	24.539
NONDIA	$[-1, -1, ..., -1]$	88.718	24.164
NONDQUAR	$[1, -1, ..., 1, -1]$	87.594	24.258
EG2	$[1, 1, ..., 1]$	89.498	24.399
CURLY20	$\left[\frac{0.001}{n+1}, \frac{0.001}{n+1}, ..., \frac{0.001}{n+1}\right]$	88.25	23.868
DIXMAANA 1	$[2, 2, ..., 2]$	89.233	24.056
DIXMAANA 2	$[2, 2, ..., 2]$	88.235	24.024
DIXMAANA 3	$[2, 2, ..., 2]$	88.172	23.65
DIXMAANA 4	$[2, 2, ..., 2]$	90.917	24.866
DIXMAANA 5	$[2, 2, ..., 2]$	90.715	24.071
DIXMAANA 6	$[2, 2, ..., 2]$	88.702	24.414
DIXMAANA 7	$[2, 2, ..., 2]$	89.498	24.305
DIXMAANA 8	$[2, 2, ..., 2]$	88.562	24.289
DIXMAANA 9	$[2, 2, ..., 2]$	88.812	24.57
DIXMAANA 10	$[2, 2, ..., 2]$	88.297	24.43
DIXMAANA 11	$[2, 2, ..., 2]$	87.516	24.507
DIXMAANA 12	$[2, 2, ..., 2]$	88.078	24.337
Partial Perturbed Quadratic	$[0.5, 0.5, ..., 0.5]$	88.562	24.305
Broyden Tridiagonal	$[-1, -1, ..., -1]$	88.	24.071
Almost Perturbed Quadratic	$[0.5, 0.5, ..., 0.5]$	87.532	24.149
Perturbed Tridiagonal Quadratic	$[0.5, 0.5, ..., 0.5]$	89.295	24.508
Staircase 1	$[1, 1, ..., 1]$	89.436	24.57
Staircase 2	$[0, 0, ..., 0]$	88.468	24.664
LIARWHD	$[4, 4, ..., 4]$	88.609	24.336
POWER	$[1, 1, ..., 1]$	91.51	24.071
ENGVAL1	$[2, 2, ..., 2]$	87.626	24.711
CRAGGLVY	$[1, 2, ..., 2]$	88.858	24.18
EDENSCH	$[0, 0, ..., 0]$	87.564	24.414
INDEF	$[\frac{1}{n+1}, \frac{1}{n+1}, ..., \frac{1}{n+1}]$	90.247	23.837
CUBE	$[-1.2, 1, ..., -1.2, 1]$	88.374	24.18
EXPLIN1	$[0, 0, ..., 0]$	88.343	24.399
EXPLIN2	$[0, 0, ..., 0]$	90.169	24.601

ARGLINC	$[1, 1, ..., 1]$	89.575	24.18
BDEXP	$[1, 1, ..., 1]$	88.967	24.102
HARKERP2	$[1, 2, ..., n]$	87.828	24.242
GENHUMPS	$[-506, 506.2, ..., 506.2]$	87.595	24.679
MCCORMCK	$[1, 1, ..., 1]$	88.796	24.43
NONSCOMP	$[3, 3, ..., 3]$	88.157	24.913
VARDIM	$[1 - \frac{1}{n}, 1 - \frac{2}{n}, ..., 0]$	87.423	24.632
Diagonal 6	$[1, 1, ..., 1]$	88.032	24.087
SINQUAD	$[a, ..., a]$ $a = 0.1$	88.032	24.399
Extended DENSCHNB	$[1, 1, ..., 1]$	88.998	24.243
Extended DENSCHNF	$[2, 0, ..., 2, 0]$	89.201	24.398
COSINE	$[1, 1, ..., 1]$	88.749	24.18
SINE	$[1, 1, ..., 1]$	88.468	24.118
SINCOS	$[3, 0.1, ..., 3, 0.1]$	88.453	24.492
Diagonal 9	$[1, 1, ..., 1]$	87.891	24.445
Average		86,0027	25,2924

Tabela 3. $n = 10$

Funkcija	x_0	LSCD	CD	LS	DP	HD
Extended Penalty	$[1, 2, ..., n]$	0.094	0.125	0.125	0.171	0.172
Perturbed Quadratic	$[a, a, ..., a]$ $a = 0.5$	0.093	0.14	0.141	0.14	0.14
Raydan 1	$[1, 1, ..., 1]$	0.14	0.187	0.203	0.187	0.171
Raydan 2	$[1, 1, ..., 1]$	0.063	0.078	0.078	0.093	0.078
Diagonal1	$[\frac{1}{n}, \frac{1}{n}, ... \frac{1}{n}]$	0.093	0.125	0.109	0.14	0.141
Diagonal2	$[1, \frac{1}{2}, ... \frac{1}{n}]$	0.156	0.156	0.172	0.187	0.188
Diagonal3	$[1, 1, ..., 1]$	0.141	0.156	0.218	0.187	0.187
Generalized Tridiagonal-1	$[2, 2, ..., 2]$	0.171	0.203	0.203	0.203	0.188
Hager	$[1, 1, ..., 1]$	0.141	0.203	0.187	0.172	0.171
Extended Tridiagonal-1	$[2, 2, ..., 2]$	0.156	0.187	0.187	0.188	0.156
Extended Three Exponential Terms	$[a, a, ..., a]$ $a = 0.1$	0.156	0.172	0.156	0.171	0.141
Diagonal4	$[1, 1, ..., 1]$	0.14	0.171	0.187	0.187	0.14
Diagonal5	$[a, a, ..., a]$ $a = 1.1$	0.125	0.156	0.156	0.171	0.14
Extended Himmelblau	$[1, 1, ..., 1]$	0.14	0.156	0.188	0.171	0.141
Extended PSC1	$[3, a, ..., 3, a]$ $a = 0.1$	0.125	0.188	0.203	0.172	0.125
Full Hessian FH2	$[a, a, ..., a]$ $a = 0.01$	0.171	0.171	0.187	0.187	0.14
Extended Block Diagonal BD1	$[a, a, ..., a]$ $a = 0.1$	0.141	0.156	0.187	0.171	0.14
Quadratic QF1	$[1, ..., 1]$	0.125	0.188	0.202	0.187	0.172
Extended Quadratic Penalty QP1	$[1, ..., 1]$	0.156	0.203	0.219	0.156	0.125

Quadratic QF2	$[a, ..., a]$ $a = 0.5$	0.156	0.172	0.203	0.203	0.14
Extended EP1	$[a, ..., a]$ $a = 1.5$	0.156	0.202	0.187	0.203	0.109
Extended Tridiagonal-2	$[1, ..., 1]$	0.156	0.188	0.156	0.156	0.141
Tridia	$[1, 1, ..., 1]$	0.14	0.171	0.172	0.172	0.125
Arwhead	$[1, ..., 1]$	0.14	0.187	0.187	0.187	0.125
Dqdrtic	$[3, ..., 3]$	0.141	0.14	0.203	0.172	0.14
Quartc	$[2, ..., 2]$	0.14	0.203	0.14	0.172	0.109
Dixon3dq	$[a, ..., a]$ $a = -1$	0.16	0.2	0.22	0.19	0.12
Biggsb1	$[0, ..., 0]$	0.16	0.17	0.19	0.17	0.16
Generalized Quartic	$[1, ..., 1]$	0.16	0.16	0.22	0.19	0.12
Diagonal 7	$[1, ..., 1]$	0.17	0.2	0.19	0.16	0.12
Diagonal 8	$[1, ..., 1]$	0.16	0.2	0.19	0.17	0.14
Full Hessian FH3	$[1, 1, ..., 1]$	0.16	0.16	0.2	0.2	0.12
Himmelbg	$[a, a, ..., a]$ $a = 1.5$	0.16	0.19	0.17	0.16	0.12
Extended Freudenstein Roth	$[a, -2, ..., a, -2]$ $a = 0.5$	0.16	0.17	0.14	0.17	0.11
Extended Trigonometric	$[a, a, ..., a]$ $a = 0.2$	0.14	0.16	0.16	0.16	0.12
Extended Rosenbrock	$[a, 1, ..., a, 1]$ $a = -1.2$	0.12	0.16	0.09	0.17	0.12
Generalized Rosenbrock	$[a, 1, ..., a, 1]$ $a = -1.2$	0.12	0.14	0.16	0.14	0.14
Extended White Holst	$[a, 1, ..., a, 1]$ $a = -1.2$	0.14	0.17	0.16	0.09	0.12
Generalized White Holst	$[a, 1, ..., a, 1]$ $a = -1.2$	0.14	0.14	0.17	0.14	0.16
Extended Beale	$[1, a, ..., 1, a]$ $a = 0.8$	0.14	0.11	0.14	0.14	0.17
Perturbed Quadratic	$[a, a, ..., a]$ $a = 0.5$	0.12	0.14	0.14	0.11	0.14
Extended Powell	$[3, a, b, ..., 3, a, b]$ $a = -1,\ b = 0.1$	0.14	0.12	0.16	0.16	0.09

Extended Maratos	$[a,b,...,a,b]$ $a=1.1,\ b=0.1$	0.12	0.09	0.16	0.16	0.14
Extended Cliff	$[0,a,...,0,a]$ $a=-1$	0.14	0.12	0.16	0.12	0.12
Extended Wood	$[a,b,...,a,b]$ $a=-3,\ b=-1$	0.16	0.17	0.14	0.16	0.14
Extended Hiebert	$[0,...,0]$	0.14	0.12	0.16	0.16	0.12
Extended Quadratic Penalty QP2	$[1,...,1]$	0.12	0.14	0.12	0.14	0.11
Extended Quadratic Exponential EP1	$[a,...,a]$ $a=1.5$	0.12	0.16	0.12	0.11	0.14
FLETCBV3	$[\frac{1}{n+1},...,\frac{n}{n+1}]$	0.11	0.16	0.141	0.16	0.12
FLETCHCR	$[0,...,0]$	0.16	0.14	0.16	0.11	0.17
BDQRTIC	$[1,...,1]$	0.12	0.17	0.14	0.16	0.14
ARGLINB	$[1,...,1]$	0.12	0.14	0.14	0.11	0.16
NONDIA	$[a,...,a]$ $a=-1$	0.12	0.19	0.14	0.14	0.16
NONDQUAR	$[1,a,...,1,a]$ $a=-1$	0.11	0.14	0.14	0.09	0.14
EG2	$[1,...,1]$	0.11	0.14	0.125	0.17	0.11
CURLY20	$\left[\frac{a}{n+1},...,\frac{a}{n+1}\right]$ $a=0.001$	0.14	0.16	0.16	0.11	0.11
DIXMAANA 1	$[2,...,2]$	0.12	0.14	0.16	0.12	0.14
DIXMAANA 2	$[2,...,2]$	0.12	0.17	0.11	0.17	0.11
DIXMAANA 3	$[2,...,2]$	0.09	0.14	0.12	0.16	0.09
DIXMAANA 4	$[2,...,2]$	0.16	0.17	0.17	0.09	0.16
DIXMAANA 5	$[2,...,2]$	0.14	0.16	0.12	0.12	0.14
DIXMAANA 6	$[2,...,2]$	0.16	0.16	0.14	0.09	0.12
DIXMAANA 7	$[2,...,2]$	0.16	0.17	0.09	0.08	0.16
DIXMAANA 8	$[2,...,2]$	0.12	0.17	0.16	0.09	0.17
DIXMAANA 9	$[2,...,2]$	0.11	0.16	0.17	0.12	0.09
DIXMAANA 10	$[2,...,2]$	0.12	0.17	0.12	0.14	0.16
DIXMAANA 11	$[2,...,2]$	0.14	0.12	0.12	0.16	0.16
DIXMAANA 12	$[2,...,2]$	0.11	0.16	0.16	0.14	0.14
Partial Perturbed Quadratic	$[a,...,a]$ $a=0.5$	0.12	0.14	0.11	0.17	0.09

Broyden Tridiagonal	$[a, ..., a]$ $a = -1$	0.11	0.12	0.14	0.11	0.08
Almost Perturbed Quadratic	$[a, ..., a]$ $a = 0.5$	0.14	0.14	0.16	0.11	0.12
Perturbed Tridiagonal Quadratic	$[a, ..., a]$ $a = 0.5$	0.14	0.16	0.14	0.14	0.11
Staircase 1	$[1, ..., 1]$	0.09	0.14	0.14	0.17	0.09
Staircase 2	$[0, ..., 0]$	0.09	0.09	0.16	0.17	0.09
LIARWHD	$[4, ..., 4]$	0.16	0.14	0.09	0.16	0.14
POWER	$[1, ..., 1]$	0.11	0.11	0.12	0.14	0.12
ENGVAL1	$[2, ..., 2]$	0.14	0.19	0.11	0.17	0.09
CRAGGLVY	$[1, 2, ..., 2]$	0.11	0.12	0.14	0.11	0.14
EDENSCH	$[0, ..., 0]$	0.14	0.14	0.16	0.17	0.19
INDEF	$[\frac{1}{n+1}, ..., \frac{1}{n+1}]$	0.14	0.17	0.17	0.16	0.12
CUBE	$[a, 1, ..., a, 1]$ $a = -1.2$	0.14	0.16	0.16	0.12	0.14
EXPLIN1	$[0, 0, ..., 0]$	0.14	0.16	0.12	0.11	0.14
EXPLIN2	$[0, 0, ..., 0]$	0.11	0.12	0.14	0.16	0.17
ARGLINC	$[1, 1, ..., 1]$	0.16	0.16	0.19	0.11	0.16
BDEXP	$[1, 1, ..., 1]$	0.09	0.16	0.16	0.17	0.14
HARKERP2	$[1, 2, ..., n]$	0.14	0.14	0.09	0.14	0.14
GENHUMPS	$[-506, a, ..., a]$ $a = 506.2$	0.12	0.16	0.18	0.17	0.11
MCCORMCK	$[1, ..., 1]$	0.14	0.16	0.09	0.12	0.17
NONSCOMP	$[3, ..., 3]$	0.12	0.14	0.16	0.09	0.14
VARDIM	$[\frac{a}{n}, \frac{a-1}{n}, ..., 0]$ $a = n - 1$	0.09	0.12	0.11	0.11	0.14
Diagonal6	$[1, ..., 1]$	0.09	0.17	0.11	0.11	0.14
SINQUAD	$[a, ..., a]$ $a = 0.1$	0.14	0.16	0.17	0.16	0.14
Extended DENSCHNB	$[1, ..., 1]$	0.14	0.16	0.16	0.14	0.17
Extended DENSCHNF	$[2, 0, ..., 2, 0]$	0.12	0.16	0.12	0.14	0.16
COSINE	$[1, ..., 1]$	0.14	0.16	0.12	0.11	0.14
SINE	$[1, ..., 1]$	0.12	0.12	0.12	0.14	0.16
SINCOS	$[3, a, ..., 3, a]$ $a = 0.1$	0.12	0.14	0.14	0.19	0.12
Diagonal9	$[1, ..., 1]$	0.14	0.16	0.11	0.12	0.17
Average		0.1332	0.1557	0.153	0,149	0.137

Tabela 4. $n = 100$

Funkcija	x_0	LSCD	CD	LS	DP	HD
Extended Penalty	$[1, ..., n]$	78.37	78.14	80.64	76.16	79.17
Perturbed Quadratic	$[a, ..., a]$ $a = 0.5$	78.36	80.36	81.9	77.64	80.51
Raydan 1	$[1, ..., 1]$	64.21	65.55	67.1	64.04	65.25
Raydan 2	$[1, ..., 1]$	11.25	11.82	11.84	11.5	11.56
Diagonal1	$[\frac{1}{n}, ... \frac{1}{n}]$	11.5	11.68	11.87	11.57	11.68
Diagonal2	$[\frac{1}{1}, ... \frac{1}{n}]$	81.45	83.29	85.94	82.01	83.18
Diagonal3	$[1, ..., 1]$	82.37	84.36	86.44	82.43	84.29
Generalized Tridiagonal-1	$[2, ..., 2]$	82.87	84.75	87.05	82.93	84.54
Hager	$[1, ..., 1]$	83.02	84.85	86.72	83.27	84.91
Extended Tridiagonal-1	$[2, ..., 2]$	81.45	85.07	87.16	83.43	84.65
Extended Three Exponential Terms	$[a, ..., a]$ $a = 0.1$	82.52	84.91	87.27	83.13	85.05
Diagonal4	$[1, ..., 1]$	82.87	85.29	87.03	83.37	84.8
Diagonal5	$[a, ..., a]$ $a = 1.1$	82.852	84.86	87.3	82.62	84.85
Extended Himmelblau	$[1, ..., 1]$	82.95	84.5	86.72	83.51	84.96
Extended PSC1	$[3, a, ..., 3, a]$ $a = 0.1$	82.82	85.24	87.2	83.18	84.97
Full Hessian FH2	$[a, ..., a]$ $a = 0.01$	82.91	85.04	87.28	83.34	85.1
Extended Block Diagonal BD1	$[a, ..., a]$ $a = 0.1$	82.62	85.43	87.05	83.09	84.82
Quadratic QF1	$[1, ..., 1]$	82.87	85.36	87.41	83.15	85
Extended Quadratic Penalty QP1	$[1, ..., 1]$	82.73	85.16	87.09	82.85	85.04

Quadratic QF2	$[a, ..., a]$ $a = 0.5$	83.01	85.25	87.42	83.1	85.13
Extended Quadratic Exponential EP1	$[a, ..., a]$ $a = 1.5$	82.84	84.86	87.3	82.54	84.91
Extended Tridiagonal-2	$[1, ..., 1]$	82.79	85.21	87.36	82.8	84.55
Tridia	$[1, ..., 1]$	82.57	85.22	86.92	83.09	84.52
Arwhead	$[1, ..., 1]$	83.01	85.57	86.64	83.41	84.54
Dqdrtic	$[3, ..., 3]$	83.01	85.05	87.45	83.88	84.82
Quartc	$[2, ..., 2]$	82.99	86.16	87.39	83.6	84.51
Dixon3dq	$[a, ..., a]$ $a = -1$	82.87	85.02	87.16	83.46	84.86
Biggsb1	$[0, ..., 0]$	83.16	84.88	87.14	83.52	84.9
Generalized Quartic	$[1, ..., 1]$	82.71	85.22	87.02	83.43	85.43
Diagonal 7	$[1, ..., 1]$	82.79	84.93	86.97	83.21	84.82
Diagonal 8	$[1, ..., 1]$	82.96	84.97	86.72	83.1	84.83
Full Hessian FH3	$[1, 1, ..., 1]$	82.84	85.29	87.39	82.99	84.35
Himmelbg	$[a, a, ..., a]$ $a = 1.5$	83.2	85.63	87.44	82.82	85.22
Extended Freudenstein Roth	$[a, -2, ..., a, -2]$ $a = 0.5$	81.82	82.1	87.73	82.87	81.68
Extended Trigonometric	$[a, a, ..., a]$ $a = 0.2$	81.87	82.03	88.26	83.15	81.87
Extended Rosenbrock	$[a, 1, ..., a, 1]$ $a = -1.2$	81.26	81.84	87.42	82.59	81.62
Generalized Rosenbrock	$[a, 1, ..., a, 1]$ $a = -1.2$	81.56	81.74	87.52	82.32	81.74
Extended White Holst	$[a, 1, ..., a, 1]$ $a = -1.2$	81.6	83.16	87.3	82.52	81.7
Generalized White Holst	$[a, 1, ..., a, 1]$ $a = -1.2$	82.1	82.24	87.05	82.76	82.02
Extended Beale	$[1, a, ..., 1, a]$ $a = 0.8$	81.98	83.09	87.2	82.26	81.53
Perturbed Quadratic	$[a, a, ..., a]$ $a = 0.5$	82.57	82.82	87.17	82.85	82.01
Extended Powell	$[3, a, b, ..., 3, a, b]$ $a = -1,\ b = 0.1$	81.73	82.07	86.72	81.87	80.76

Extended Maratos	$[a, b, ..., a, b]$ $a = 1.1,\ b = 0.1$	82.23	81.64	87.1	81.73	82.29
Extended Cliff	$[0, a, ..., 0, a]$ $a = -1$	82.76	82.66	87.25	82.37	82.26
Extended Wood	$[a, b, ..., a, b]$ $a = -3,\ b = -1$	81.46	81.57	87.38	82.48	81.96
Extended Hiebert	$[0, ..., 0]$	81.76	81.99	87.38	82.38	81.7
Extended Quadratic Penalty QP2	$[1, ..., 1]$	81.87	81.65	87.58	82.55	81.93
Extended Quadratic Exponential EP1	$[a, ..., a]$ $a = 1.5$	81.43	82.15	87.23	82.98	81.56
FLETCBV3	$[\frac{1}{n+1}, ..., \frac{n}{n+1}]$	81.7	81.74	87.67	82.66	81.64
FLETCHCR	$[0, ..., 0]$	81.62	82.63	87.55	82.88	81.68
BDQRTIC	$[1, ..., 1]$	82.23	82.66	87.28	82.54	81.9
ARGLINB	$[1, ..., 1]$	82.57	81.6	87.66	82.65	82.07
NONDIA	$[a, ..., a]$ $a = -1$	82.16	82.77	87.33	82.48	82.18
NONDQUAR	$[1, a, ..., 1, a]$ $a = -1$	82.23	82.32	87.47	82.45	81.73
EG2	$[1, ..., 1]$	81.68	82.24	86.49	83.19	81.76
CURLY20	$\left[\frac{a}{n+1}, ..., \frac{a}{n+1}\right]$ $a = 0.001$	81.93	82.49	87.59	82.84	81.93
DIXMAANA 1	$[2, ..., 2]$	81.15	82.32	87.92	82.46	81.99
DIXMAANA 2	$[2, ..., 2]$	81.59	82.7	87.83	82.806	81.7
DIXMAANA 3	$[2, ..., 2]$	81.57	82.46	87.48	82.93	81.9
DIXMAANA 4	$[2, ..., 2]$	81.45	82.93	87.13	84.3	82.52
DIXMAANA 5	$[2, ..., 2]$	81.35	83.24	88.53	82.26	81.21
DIXMAANA 6	$[2, ..., 2]$	82.2	83.77	87.55	83.52	82.9
DIXMAANA 7	$[2, ..., 2]$	81.17	83.21	87.61	83.94	81.7
DIXMAANA 8	$[2, ..., 2]$	82.03	83.69	87.66	82.21	82.21
DIXMAANA 9	$[2, ..., 2]$	81.48	83.77	87.98	82.37	82.46
DIXMAANA 10	$[2, ..., 2]$	81.37	83.65	87.24	82.79	81.71
DIXMAANA 11	$[2, ..., 2]$	81.79	83.85	87.67	82.76	82.06
DIXMAANA 12	$[2, ..., 2]$	81.46	83.93	87.97	82.96	82.02

Partial Perturbed Quadratic	$[a, ..., a]$ $a = 0.5$	81.85	83.79	87.88	82.52	81.84
Broyden Tridiagonal	$[a, ..., a]$ $a = -1$	81.74	83.66	87.59	82.85	82.17
Almost Perturbed Quadratic	$[a, ..., a]$ $a = 0.5$	82.26	83.74	87.72	82.88	81.79
Perturbed Tridiagonal Quadratic	$[a, ..., a]$ $a = 0.5$	81.59	83.76	87.173	82.71	81.28
Staircase1	$[1, ..., 1]$	81.87	83.63	88.36	82.88	82.62
Staircase2	$[0, ..., 0]$	82.4	83.85	87.94	83.12	81.63
LIARWHD	$[4, ..., 4]$	81.62	83.6	88.55	83.07	82.23
POWER	$[1, ..., 1]$	82.15	83.59	87.58	82.8	82.31
ENGVAL1	$[2, ..., 2]$	82.34	83.35	87.63	83.33	82.87
CRAGGLVY	$[1, 2, ..., 2]$	81.7	83.46	88.03	83.24	82.16
EDENSCH	$[0, ..., 0]$	81.67	83.32	87.67	83.33	81.62
INDEF	$[\frac{1}{n+1}, ..., \frac{1}{n+1}]$	81.68	83.49	87.45	82.37	82.24
CUBE	$[a, 1, ..., a, 1]$ $a = -1.2$	82.17	83.19	87.58	82.32	81.78
EXPLIN1	$[0, ..., 0]$	81.68	83.49	87.34	82.96	82.63
EXPLIN2	$[0, ..., 0]$	82.29	83.46	87.81	83.27	81.92
ARGLINC	$[1, 1, ..., 1]$	81.95	83.99	87.3	82.1	82.74
BDEXP	$[1, 1, ..., 1]$	82.35	83.66	87.63	82.82	82.37
HARKERP2	$[1, 2, ..., n]$	81.56	83.68	87.45	83.18	82.57
GENHUMPS	$[-506, a, ..., a]$ $a = 506.2$	82.56	83.54	88.11	83.29	82.59
MCCORMCK	$[1, ..., 1]$	82.35	84.01	88	82.77	82.27
NONSCOMP	$[3, ..., 3]$	82.2	83.19	88	82.84	82.63
VARDIM	$[\frac{a}{n}, \frac{a-1}{n}, ..., 0]$ $a = n - 1$	82.46	83.49	87.66	82.93	82.24
Diagonal6	$[1, ..., 1]$	82.04	83.55	87.84	83.62	82.12
SINQUAD	$[a, ..., a]$ $a = 0.1$	82.06	83.6	87.83	83.63	82.27
Extended DENSCHNB	$[1, ..., 1]$	82.1	83.41	88.28	83.24	81.81
Extended DENSCHNF	$[2, 0, ..., 2, 0]$	82.45	83.41	87.66	82.8	82.87
COSINE	$[1, ..., 1]$	82.12	83.27	87.66	82.48	82.21
SINE	$[1, ..., 1]$	81.87	83.83	87.75	82.87	82.31
SINCOS	$[3, a, ..., 3, a]$	82.1	83.26	87.59	82.76	82.2
Diagonal9	$[1, ..., 1]$	82.46	83.24	88.11	83.19	82.21
Average		80,46	81,9	85,576	81,13	81,16

Najpre komentarišemo Tabelu 1.

U Tabeli 1. upoređeni su rezultati dobijeni primenom pomenutih metoda GDQN [113] i RGDQN, pri čemu su prosečni rezultati dati u poslednjoj vrsti Tabele 1. Prosečni rezultati jasno pokazuju da je metod RGDQN mnogo bolji od metoda GDQN. Jedino se u slučaju Extended Beale funkcije RGDQN poklapa sa GDQN.

Sada razmatramo Tabelu 2.

Pregledom rezultata u Tabeli 2., zaključujemo da je metod RGDQN ubedljivo bolji od metoda GDQN. To potvrđuju i prosečni rezultati, dati u poslednjoj vrsti Tabele 2.

Zaključak je da se uvođenjem parametra θ, čije se vrednosti nasumično biraju iz intervala $(0, 1)$, dobijaju mnogo bolji eksperimentalni rezultati u poređenju sa eksperimentalnim rezultatima koji važe za GDQN metod.

Slede Tabela 3. i Tabela 4., u kojima se porede vrednosti metoda LSCD, predstavljenog u Poglavlju 5, sa metodima CD, LS, DYPRP [4], u oznaci DP, HSDY, u oznaci HD iz [12]. Tabela 3. i Tabela 4. sadrže rezultate koji su dobijeni pri korišćenju dimenzija problema $n = 10$, $n = 100$ respektivno.

Sada analiziramo Tabelu 3.

Iz prosečnih rezultata, datih u poslednjoj vrsti Tabele 3., možemo zaključiti da se metod LSCD pokazao boljim od ostalih metoda u tabeli.

Dalje, LSCD metod se pokazao najboljim u slučaju ukupno 23 funkcije, dok kod njih 15 deli najbolji rezultat sa nekim od ostalih metoda iz Tabele 3. U slučaju 14 funkcija se poklapa sa jednim od metoda LS i CD, ali je od onog drugog bolji. Osim u slučaju funkcija LIARWHD i POWER, bolji je od bar jednog od LS i CD metoda. Kod ukupno 65 funkcija, LSCD se pokazao boljim od metoda DYPRP, a kod 14 funkcija je jednak sa ovim metodom. U slučaju 42 funkcije pokazao se boljim od HSDY metoda. Kod 11 funkcija jednak je sa ovim metodom.

Analiziramo sada Tabelu 4., u kojoj su pomenuti metodi LSCD, CD, LS, DYPRP i HSDY upoređivani za $n = 100$. Prosečni rezultati, dati u poslednjoj vrsti Tabele 4., pokazuju da je metod LSCD dao najbolje rezultate među pomenutim metodima.

Dalje, LSCD metod se pokazao najboljim u slučaju ukupno 65 funkcija, dok u slučaju dve funkcije deli najbolji rezultat sa HSDY. U slučaju 91 funkcije bolji je od CD, a kod svih funkcija je bolji od LS. Takođe, u slučaju 91 funkcije bolji je od DYPRP. U slučaju 73 funkcije bolji je od HSDY metoda.

Zaključak je da se formiranjem konveksne kombinacije metoda LS i CD, pri čemu se parametar θ_k bira tako da je zadovoljen uslov konjugacije, svakako poboljšavaju metodi LS i CD, a dobija se metod uporediv sa nekim poznatim hibridnim metodima, i čak bolji od njih.

Literatura

[1] M. Al-Baali, *Descent property and global convergence of the Fletcher-Reeves method with inexact line search*, IMA J. Numer. Anal., 5 (1985), 121-124.

[2] X.-M. An, D.-H. Li, and Y. Xiao, *Sufficient descent directions in unconstrained optimization*, Computational Optimization and Applications, 48, 3 (2011), 515-532.

[3] N. Andrei, *40 Conjugate Gradient Algorithms for unconstrained optimization, A survey on their definition*, ICI Technical Report, 13/08, 2008.

[4] N. Andrei, *New hybrid conjugate gradient algorithms for unconstrained optimization*, Encyclopedia of Optimization, 2560-2571, 2009.

[5] N. Andrei, *A Hybrid Conjugate Gradient Algorithm with Modified Secant Condition for Unconstrained Optimization as a Convex Combination of Hestenes-Stiefel and Dai-Yuan Algorithms*, STUDIES IN INFORMATICS AND CONTROL, 17, 4 (2008), 373-392.

[6] N. Andrei, *Scaled conjugate gradient algorithms for unconstrained optimization*, Computational Optimization and Applications, 38, 3 (2007), 401-416.

[7] N. Andrei, *Scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization*, Optimization Methods and Software, 22, 4 (2007), 561-571.

[8] N. Andrei, *A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization*, Applied Mathematics Letters, 20, 6 (2007), 645650.

[9] N. Andrei, *Accelerated conjugate gradient algorithm with finite difference Hessian/vector product approximation for unconstrained optimization*, Journal of Computational and Applied Mathematics, 230, 2, 15 (2009), 570-582.

[10] N. Andrei, *A New Gradient Descent Method for Unconstrained Optimization*, ICI Technical Report, (2004).

[11] N. Andrei, *An acceleration of gradient descent algorithm with backtracking for unconstrained optimization*, Numer. Algor., 42 (2006), 63-73.

[12] N. Andrei, *A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan*, Studies in Informatics and Control, 17, 1 (2008), 55-70.

[13] N. Andrei, *An unconstrained optimization test functions*, Advanced Modeling and Optimization. An Electronic International Journal, 10 (2008), 147-161.

[14] L. Armijo, *Minimization of functions having Lipschitz continuous partial derivatives*, Pacific J. Mathematics 16 (1966) 1-3.

[15] P. Baptist, J. Stoer, *On the relation between quadratic termination and convergence properties of minimization algorithms, Part II, Applications*, Numer. Math., 28 (1977), 367-392.

[16] J. Barzilai, J. M. Borwein, Two-Point Step Size Gradient Methods, IMA Journal of Numerical Analysis, 8 (1988), 141-148.

[17] E. M. L. Beale, *A Derivative of Conjugate Gradients*, in F.A Lootsma eds., Numerical Methods for Nonlinear Optimization, London, Academic Press (1972), 39-43.

[18] E. Birgin and J.M. Martinez, *A spectral conjugate gradient method for unconstrained optimization*, Applied Math. and Optimization, 43 (2001), 117-128.

[19] E. G. Birgin, N. Krejić, J. M. Martinez, *Globally Convergent Inexact Quasi-Newton Methods for Solving Nonlinear Systems*, Numerical Algorithms, 32 (2003), 249-250.

[20] C. G. Broyden, *The convergence of a class of double-rank minimization algorithms 1, General considerations*, J. Inst. Math. Appl., 6 (1970), 76-90.

[21] S. Buhmiler, N. Krejić, Z. Lužanin, , *Practical Quasi-Newton methods for singular nonlinear systems*, Numerical Algorithms, 55, 4 (2010), 481-502.

[22] W. Burmeister, *Die konvergenzordnung des Fletcher-Powell algorithmus*, Z. Angew. Math. Mech. 53 (1973), 693-699.

[23] A. Cohen, *Rate of convergence of several conjugate gradient algorithms*, SIAM J. Numer. Anal., 9 (1972), 248-259.

[24] H.P. Crowder, P. Wolfe, *Linear convergence of the conjugate gradient method*, IBM J. Res. Develop., 16 (1972), 431-433.

[25] J.W. Daniel, *The conjugate gradient method for linear and nonlinear operator equations*, SIAM J. Numer. Anal., 4 (1967), 10-26.

[26] Y.H. Dai, *Analyses of conjugate gradient methods*, Ph.D. thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997.

[27] Y.H. Dai, *A nonmonotone conjugate gradient algorithm for unconstrained optimization*, J. Syst. Sci. Complex., 15 (2002), 139-145.

[28] Y.H. Dai, *On the nonmonotone line search*, J. Optim. Theory Appl., 112 (2002), 315-330.

[29] Y. H. Dai, R. Fletcher, *Projected Barzilai-Borwein methods for largescale box-constrained quadratic programming*, Numer. Math., 100 (2005), 21-47.

[30] Y.H. Dai, L. Z. Liao, *New conjugacy conditions and related nonlinear conjugate gradient methods*, Appl. Math. Optim.,43 (2001), 87-101.

[31] Y. H. Dai, L. Z. Liao, *R-linear convergence of the Barzilai and Borwein gradient method*, IMA J. Numer. Anal., 22 (2002), 1-10.

[32] Y. H. Dai, L. Z. Liao, D. Li, *An analysis of Barzilai-Borwein gradient method for unsymmetric linear equations*, In K. Teo L. Qi and X. Yang (Eds.), Optimization and Control with Applications, Springer (2005), 183-211.

[33] Y.H. Dai, Y. Yuan, *Further studies on the Polak-Ribiére-Polyak method*, Research report ICM-95-040, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1995.

[34] Y.H. Dai, Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal., 16 (1996), 155-164.

[35] Y.H. Dai, Y. Yuan, *Convergence of the Flecher-Reeves method under a generalized Wolfe search*, J. Comput. Math., 2 (1996), 142-148.

[36] Y.H. Dai, Y. Yuan, *Convergence properties of the conjugate descent method*, Adv. Math. (China), 26 (1996), 552-562.

[37] Y.H. Dai, Y. Yuan, *A class of globally convergent conjugate gradient methods*, Research report ICM-98-030, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998.

[38] Y.H. Dai, Y. Yuan, *Extension of a class of nonlinear conjugate gradient methods*, Research report ICM-98-049, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998.

[39] Y.H. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim., 10 (1999), 177-182.

[40] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin, Y. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, SIAM J. Optim., 10 (1999), 345-358.

[41] Y.H. Dai, Y. Yuan, *Nonlinear conjugate gradient methods*, Shanghai Science and Technology Publisher, Shanghai, 2000.

[42] Y.H. Dai, Y. Yuan, *An efficient hybrid conjugate gradient method for unconstrained optimization*, Annals of Operations Research, 103 (2001), 33-47.

[43] Y.H. Dai, *New properties of a nonlinear conjugate gradient method*, Numer. Math., 89 (2001), 83-98.

[44] Y.H. Dai, Y. Yuan, *A three parameter family of hybrid conjugate gradient method*, Mathematics of Computation, 70 (2001), 1155-1167.

[45] Y.H. Dai, Q. Ni, *Testing different conjugate gradient methods for large-scale unconstrained optimization*, Journal of Computational Mathematics, 21, 3 (2003), 311-320.

[46] Y.H. Dai, Y. Yuan, *A class of globally convergent conjugate gradient methods*, Sci. China Ser. A, 46 (2003), 251-261.

[47] Y. H. Dai, Y. Yuan, *Analysis of monotone gradient methods*, J. Ind. Manag. Optim, 1 (2005), 181-192.

[48] W.C. Davidon, *Variable metric methods for minimization*, Argonne National Labs Report, ANL-5990, 1959.

[49] S. Deng, Z. Wan, X. Chen, *An Improved Spectral Conjugate Gradient Algorithm for Nonconvex Unconstrained Optimization Problems*, Journal of Optimization Theory and Applications, 157, 3 (2013), 820-842.

[50] M.A. Diniz-Ehrhardt, J. M. Martínez, M. Raydan, *A derivative-free nonmonotone line-search technique for unconstrained optimization*, Journal of Computational and Applied Mathematics, 219, 2 (2008), 383-397.

[51] S.S. Đorđević, *Two modifications of the method of the multiplicative parameters in descent gradient methods*, Applied Mathematics and Computation, 218, 17, (2012), 8672-8683.

[52] S.S. Đorđević, *New hybrid conjugate gradient method as a convex combination of FR and PRP methods*, Filomat, u štampi.

[53] E. D. Dolan, J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Programming, 91 (2002), 201-213.

[54] Y. Drori, M. Teboulle, *Performance of first-order methods for smooth convex minimization: a novel approach*, Mathematical Programming Series A (2013), 1-32.

[55] R. Fletcher, M.J.D. Powell, *A rapid convergent descent method for minimization*, Computer Journal, 6 (1963), 163-168.

[56] R. Fletcher, *A new approach to variable metric algorithms*, Computer Journal, 13 (1970), 317-322.

[57] R. Fletcher, *Practical methods of optimization vol. 1: Unconstrained Optimization*, John Wiley and Sons, New York, 1987.

[58] R. Fletcher, C. Reeves, *Function minimization by conjugate gradients*, Computer Journal, 7 (1964), 149-154.

[59] D. H. Li, M. Fukushima, *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*, Opt. Methods Software, 13 (2000), 181-201.

[60] A. Friedlander, J. M. Martinez, B. Molina, *Gradient method with restarts and generalizations*, SIAM J. Numer. Anal., 36 (1999), 275-289.

[61] A. Friedlander, J. M. Martnez, M. Raydan, *A new method for large-scale box constrained convex quadratic minimization problems*, Optimization Methods and Software, 5 (1995), 57-74.

[62] A. Friedlander, J. M. Martnez, S. A. Santos, *On the resolution of large scale linearly constrained convex minimization problems*, SIAM Journal of Optimization, 4 (1994), 331-339.

[63] J.C. Gilbert, J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM Journal of Optimization, 2 (1992), 21-42.

[64] D. Goldfarb, *A family of variable metric methods derived by variation mean*, Mathematics of Computation 23 (1970), 23-26.

[65] A.A. Goldstein, *On steepest descent*, SIAM J. Control 3 (1965), 147-151.

[66] G. H. Golub, D. P. O'Leary, *Some history of the conjugate gradient methods and Lanczos algorithms: 1948 - 1976*, SIAM Rev., 31 (1989), 50-100.

[67] L. Grippo, F. Lampariello, S. Lucidi, *A nononotone line search technique for Newton's method*, SIAM Journal on Numerical Analysis, 23, 4 (1986), 707-716.

[68] L. Grippo, F. Lampariello, S. Lucidi, *A class of nonmonotone stabilization methods in unconstrained optimization*, Numer. Math. 59 (1991), 779-805.

[69] L. Grippo, S. Lucidi, *A globally convergent version of the Polak-Ribiére conjugate gradient method*, Math. Prog., 78 (1997), 375-391.

[70] L. Grippo, S. Lucidi, *Convergence conditions, line search algorithms and trust region implementations for the Polak-Ribiere conjugate gradient method*, Optimization Methods and Software, 20,1 (2005), 7198.

[71] W.W. Hager, H. Zhang, *A new conjugate gradient method with guaranteed descent and an efficient line search*, SIAM J. Optim., 16, 1 (2003), 170-192.

[72] W.W. Hager, H. Zhang, *CG-DESCENT, a conjugate gradient method with guaranteed descent*, ACM Transactions on Mathematical Software, 32, 1 (2006), 113137.

[73] W.W. Hager, H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pacific Journal of Optimization, 2 (2006), 35-58.

[74] H. Zhang, W. W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim. 4 (2004), 1043-1056.

[75] J.Y. Han, G.H. Liu, H.X. Yin, *Convergence properties of conjugate gradient methods with strong Wolfe line search*, Systems Sci. Math. Sci., 11 (1998), 112-116.

[76] L. Han, G. Yu, L. Guan, *Multivariate spectral gradient method for unconstrained optimization*, Appl. Math. Comput., 210 (2008), 621-630.

[77] D. Herceg, N. Krejić, Z. Lužanin, *Quasi-Newton method with correction*, Novi Sad J. Math., 25, 1 (1996), 115-127.

[78] J. B. Hiriart-Urruty, C Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, 1993.

[79] J. B. Hiriart-Urruty, C. Lemaréchal, *Convex Analysis and Minimization Algorithms: Part 1 Fundamentals*, Springer-Verlag, 2011.

[80] N. Krejić, N. Krklec, *Line search methods with variable sample size for unconstrained optimization*, Journal of Computational and Applied Mathematics 245 (2013), 213-231.

[81] N. Krejić, N. Krklec, *Nonmonotone line search methods with variable sample size*, Numerical Algorithms, (2014).

[82] N. Krejić, Z. Lužanin, I. Stojkovska, *Gauss-Newton-based BFGS method with filter for unconstrained optimization*, Applied Mathematics and Computation 211, 1, (2009), 354-362.

[83] D. P. O'Leary, *Conjugate gradients and related KMP algorithms: The beginnings*, In L. Adams and J.L. Nazareth (Eds.) Linear and Nonlinear Conjugate Gradient - Related Methods. SIAM, Philadelphia (1996), 1-8.

[84] M.R. Hestenes, E.L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), 409-436.

[85] Y.F. Hu and C. Storey, *Global convergence result for conjugate gradient methods*, J. Optim. Theory Appl., 71 (1991), 399-405.

[86] C. Lemaréchal, *A view of line searches*, in Optimization and Optimal Control, A. Auslander, W. Oettli, J. Stoer, eds., Lecture Notes in Control and Information Science 30, Springer Verlag, Berlin (1981), 59-78.

[87] G.H. Liu, J.Y. Han, H.X. Yin, *Global convergence of the Fletcher-Reeves algorithm with an inexact line search*, Appl. Math. J. Chinese Univ. Ser. B, 10 (1995), 75-82.

[88] Y. Liu, C. Storey, *Efficient generalized conjugate gradient algorithms, Part 1: Theory*, J. Optim. Theory Appl., 69 (1991), 129-137.

[89] Z. Lužanin, D. Herceg, N. Krejić, *Parameter selection for inexact Newton methods*, Nonlinear Analysis 30 (1997), 17-24.

[90] G. McCormick, K. Ritter, *Alternative proofs of the convergence properties of the conjugate gradient method*, J. Optimization Theory and Applications, 13 (1974) 497-518.

[91] J.J. Moré, D.J. Thuente, *On line search algorithms with guaranteed sufficient decrease*, Mathematics and Computer Science Division Preprint MCS-P153-0590, Argone National Laboratory, Argonne, IL, 1990.

[92] J. L. Nazareth, *A conjugate direction algorithm without line searches*, J. Optim. Theory Appl., 23 (1977), 373-387.

[93] J. L. Nazareth, *Conjugate-gradient methods*, Encyclopedia of Optimization, C. Floudas and P. Pardalos, eds., Kluwer Academic Publishers, Boston, 1999.

[94] Y. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer, Boston, 2004.

[95] J. Nocedal, *Conjugate gradient methods and nonlinear optimization*, In Linear and nonlinear Conjugate Gradient related methods, L. Adams and J.L. Nazareth (eds.), SIAM (1996), 9-23.

[96] J. Nocedal, S. J. Wright, *Numerical Optimization*, Springer, 1999.

[97] J. M. Perry, *A class of conjugate gradient algorithms with a two-step variable-metric memory*, Discussion Paper 269, Center for Mathematical Studies in Economics and Management Sciences, Northwestern University, Evanston, Illinois, 1977.

[98] E. Polak, G. Ribiére, *Note sur la convergence de méthodes de directions conjugués*, Revue Française d'Informatique et de Recherche Opérationnelle, 16 (1969), 35-43.

[99] B.T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comp. Math. Math. Phys., 9 (1969), 94-112.

[100] M.J.D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in D.F. Griffiths, ed., Numerical Analysis Lecture Notes in Mathematics 1066 (Springer-Verlag, Berlin, 1984), 122-141.

[101] M.J.D. Powell, *Restart procedures of the conjugate gradient method*, Math. Prog., 2 (1977), 241-254.

[102] M.J.D. Powell, *Some convergence properties of the conjugate gradient method*, Math. Programming, 11 (1976), 42-49.

[103] M.J.D. Powell, *Convergence properties of algorithms for nonlinear optimization*, SIAM Rev., 28 (1986), 487-500.

[104] M. Raydan, *On the Barzilai and Borwein choice of steplength for the gradient method*, IMA J. Numer. Anal., 13 (1993), 321-325.

[105] M. Raydan, *The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem*, SIAM J. Optim., 7 (1997), 25-33.

[106] M. Raydan, B. F. Svaiter, *Relaxed steepest descent and Cauchy-Barzilai-Borwein method*, Comput. Optim. Appl. 21 (2002), 155-167.

[107] K. Ritter, *On the rate of superlinear convergence of a class of variable metric methods*, Numerische Mathematik, 35 (1980), 293-313.

[108] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1997.

[109] D.F. Shanno, *Conditioning of quasi-Newton methods for function minimization*, Math. Comput., 24 (1970), 647-656.

[110] D.F. Shanno, *On the convergence of a new conjugate gradient algorithm*, SIAM J. Numer. Anal., 15 (1978), 1247-1257.

[111] D. F. Shanno, *Conjugate gradient methods with inexact searches*, Math. Oper. Res., 3 (1978), 244-256.

[112] D. F. Shanno and K. H. Phua, *Algorithm 500, Minimization of unconstrained multivariate functions*, ACM Trans. on Math. Soft., 2 (1976), 87-94.

[113] P. Stanimirović, M. Miladinović, S. Đorđević, *Multiplicative parameters in gradient descent methods*, Filomat, 23, 3 (2009), 23-36.

[114] J. Stoer, *On the relation between quadratic termination and convergence properties of minimization algorithms, Part* I, *Theory*, Numerische Mathematik, 28 (1977), 343-366.

[115] J. Sun, J. Zhang, *Global convergence of conjugate gradient methods without line search*, Ann. Oper. Res., 163 (2001), 161-173.

[116] Wenyu Sun, Ya-Xiang Yuan, *Optimization Theory and Methods: Nonlinear Programming, Springer Optimization and Its Applications, Vol. 1*, 2006.

[117] P. L. Toint, *An assessment of nonmonotone line search techniques for unconstrained optimization*, SIAM J. Scientific Computing, 17, 3 (1996), 725-739.

[118] D. Touati-Ahmed, C. Storey, *Efficient hybrid conjugate gradient techniques*, J. Optim. Theory Appl., 64 (1990), 379-397.

[119] X. Yang, Z. Luo, X. Dai, *A Global Convergence of LS-CD Hybrid Conjugate Gradient Method*, Advances in Numerical Analysis, 2013 (2013), Article ID 517452, 5 pages.

[120] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Review, 11 (1969), 226-235.

[121] P. Wolfe, *Convergence conditions for ascent methods. II: Some corrections*, SIAM Review, 11 (1969), 226-235.

[122] Y. Xiao, Q. Wang and D. Wang, *Notes on the Dai-Yuan-Yuan modified spectral gradient method*, J. Comput. Appl. Math., 234 (2010), 2986-2992.

[123] H Yabe, M. Takano, *Global converence properties of nonlinear conjugate gradient methods with modified secant condition*, Comput. Optim. Appl., 28 (2004), 203-225.

[124] G. Yu, J. Huang, Y. Zhou, *A descent spectral conjugate gradient method for impulse noise removal*, Appl. Math. Lett., 23 (2010), 555-560.

[125] Y. Yuan, *Analysis on the conjugate gradient method*, Optim. Meth. Softw., 2 (1993), 19-29.

[126] J. Z. Zhang, N. Y. Deng, L. H. Chen, *New quasi-Newton equation and related methods for unconstrained optimization*, J. Optim. Theory Appl., 102 (1999), 147-167.

[127] J. Z. Zhang, C. X. Xu, *Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations*, J. Comput. Appl. Math., 137 (2001), 269-278.

[128] G. Zoutendijk, *Nonlinear programming, computational methods*, in Integer and Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam, (1970), 37-86.

Biografija

Rođena sam u Leskovcu, 11.3.1971. godine. Osnovnu školu i Gimnaziju završila sam u Leskovcu, obe sa Vukovom diplomom.
Završila sam u Nišu Filozofski fakultet, grupa Matematika, smer Računarstvo i informatika, sa prosečnom ocenom 7,62 i dobila sam zvanje Diplomirani matematičar za računarstvo i informatiku.
Specijalističke studije završila sam u Nišu, na Prirodno-Matematičkom fakultetu i odbranila specijalistički rad, pod nazivom *Linearno programiranje* juna 2002. godine.
Magistarske studije završila sam u Nišu, na Prirodno-matematičkom fakultetu i odbranila magistarsku tezu pod nazivom Nelinearna optimizacija i pretraživanje po pravcu, januara, 2010. godine.
Trenutno radim na Visokoj strukovnoj školi za tekstil u Leskovcu kao predavač Matematike i Računarske tehnike.

Decembar, 2014. godine Mr Snežana S. Đorđević

UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR
Identifikacioni broj:
IBR
Tip dokumentacije: Monografska dokumentacija
TD
Tip zapisa: Tekstualni štampani materijal
TZ
Vrsta rada: Doktorska disertacija
VR
Autor: Snežana S. Đorđević
AU
Mentor: Prof. dr Nataša Krejić
MN

Naslov rada: Izbor parametara kod gradijentnih metoda za probleme optimizacije bez ograničenja
NR

Jezik publikacije: srpski
JP
Jezik izvoda: srpski/engleski
JI
Zemlja publikovanja: Republika Srbija
ZP
Godina: 2014.
GO

Izdavač: Autorski reprint
IZ
Mesto i adresa: Novi Sad, Prirodno-matematički fakultet, Trg Dositeja Obradovića 4
MA
Fizički opis rada: 6/97/128/4/0/0/0
(broj poglavlja/strana/lit. citata/tabela/slika/grafika/priloga)

FO
Naučna oblast: Matematika
NO
Naučna disciplina: Numerička matematika
ND
Predmetna odrednica/Ključne reči: Nelinearna optimizacija, metodi linijskog pretraživanja, monotono linijsko pretraživanje, aproksimacija matrice hesijana, slučajno izabrani parametar, hibridni metod konjugovanih gradijenata, uslov konjugacije
PO

UDK:
Čuva se: u biblioteci Departmana za matematiku i informatiku, Novi Sad
ČU
Važna napomena:
VN

Izvod:

Posmatra se problem optimizacije bez ograničenja. Za rešavanje problema optimizacije bez ograničenja postoji mnoštvo raznovrsnih metoda. Istraživanje ovde motivisano je potrebom za metodama koje će brzo konvergirati.
Cilj je sistematizacija poznatih rezultata, kao i teorijska i numerička analiza mogućnosti uvođenja parametra u gradijentne metode.
Najpre se razmatra problem minimizacije konveksne funkcije više promenljivih.
Problem minimizacije konveksne funkcije više promenljivih ovde se rešava bez izračunavanja matrice hesijana, što je naročito aktuelno za sisteme velikih dimenzija, kao i za probleme optimizacije kod kojih ne raspolažemo ni tačnom vrednošću funkcije cilja, ni tačnom vrednošću gradijenta. Deo motivacije za istraživanjem ovde leži i u postojanju problema kod kojih je funkcija cilja rezultat simulacija.
Numerički rezultati, predstavljeni u Glavi 6, pokazuju da uvođenje izvesnog parametra može biti korisno, odnosno, dovodi do ubrzanja određenog metoda optimizacije.
Takođe se predstavlja jedan novi hibridni metod konjugovanog gradijenta, kod koga je parametar konjugovanog gradijenta konveksna kombinacija dva poznata parametra konjugovanog gradijenta.

U prvoj glavi opisuje se motivacija kao i osnovni pojmovi potrebni za praćenje preostalih glava.
U drugoj glavi daje se pregled nekih gradijentnih metoda prvog i drugog reda.
Četvrta glava sadrži pregled osnovnih pojmova i nekih rezultata vezanih za metode konjugovanih gradijenata.
Pomenute glave su tu radi pregleda nekih poznatih rezultata, dok se originalni doprinos predstavlja u trećoj, petoj i šestoj glavi.
U trećoj glavi se opisuje izvesna modifikacija određenog metoda u kome se koristi multiplikativni parametar, izabran na slučajan način. Dokazuje se linearna konvergencija tako formiranog novog metoda.
Peta glava sadrži originalne rezultate koji se odnose na metode konjugovanih gradijenata. Naime, u ovoj glavi predstavlja se novi hibridni metod konjugovanih gradijenata, koji je konveksna kombinacija dva poznata metoda konjugovanih gradijenata.
U šestoj glavi se daju rezultati numeričkih eksperimenata, izvršenih na izvesnom skupu test funkcija, koji se odnose na metode iz treće i pete glave. Implementacija svih razmatranih algoritama rađena je u paketu MATHEMATICA. Kriterijum upoređivanja je vreme rada centralne procesorske jedinice.
IZ
Datum prihvatanja teme od strane Nastavnog veća:
25.4.2014. godine
DP

Datum odbrane:
DO
Članovi komisije:

Predsednik: prof. dr Zorana Lužanin, Prirodno-Matematički fakultet u Novom Sadu

Mentor: prof. dr Natasa Krejic, Prirodno-Matematički fakultet u Novom Sadu

Član: prof. dr Zorica Uzelac, Fakultet tehnickih nauka u Novom Sadu
KO

UNIVERSITY OF NOVI SAD
FACULTY OF NATURAL SCIENCES
AND MATHEMATICS
KEY WORD DOCUMENTATION

Accession number:
ANO
Identification number:
INO
Document type: Monograph type
DT
Type of record: Printed text
TR
Contents code: PhD thesis
CC
Author: Snežana S. Đorđević
AU
Mentor: Nataša Krejić, Full Professor
MN
Title: Choice of parameters in gradient methods for the unconstrained optimization problems
TI
Language of text: Serbian
LT
Language of abstract: Serbian/English
LA
Country of publication: Republic of Serbia
CP
Locality of publication:
LP
Publication year: 2014.
PY
Publisher: Author's reprint
PU
Publ. place: Novi Sad, Faculty of Sciences, Trg Dositeja Obradovića 4
PP
Physical description: 6/97/128/4/0/0/0
(chapters/pages/literature/tables/pictures/graphics/appendices)
PD
Scientific field: Mathematics

SF
Scientific discipline: Numerical mathematics
SD
Subject / Key words: Nonlinear optimization, line search methods, monotone line search, approximation of Hessian, randomly chosen parameter, hybrid conjugate gradient method,
conjugation condition
SKW
UC:
Holding data:
HD
Note:
N

Abstract:

The problem under consideration is an unconstrained optimization problem. There are many different methods made in aim to solve the optimization problems. The investigation made here is motivated by the fact that the methods which converge fast are necessary.
The main goal is the systematization of some known results and also theoretical and numerical analysis of the possibilities to introduce some parameters within gradient methods.
Firstly, the minimization problem is considered, where the objective function is a convex, multivariable function. This problem is solved here without the calculation of Hessian, and such solution is very important, for example, when the big dimension systems are solved, and also for solving optimization problems with unknown values of the objective function and its gradient. Partially, this investigation is motivated by the existence of problems where the objective function is the result of simulations.
Numerical results, presented in Chapter 6, show that the introduction of a parameter is useful, i.e., such introduction results by the acceleration of the known optimization method.
Further, one new hybrid conjugate gradient method is presented, in which the conjugate gradient parameter is a convex combination of two known conjugate gradient parameters.
In the first chapter, there is motivation and also the basic concepts which are necessary for the other chapters.

The second chapter contains the survey of some first order and second order gradient methods.
The fourth chapter contains the survey of some basic concepts and results corresponding to conjugate gradient methods.
The first, the second and the fourth chapters are here to help in considering of some known results, and the original results are presented in the chapters 3,5 and 6.
In the third chapter, a modification of one unconstrained optimization method is presented, in which the randomly chosen multiplicative parameter is used. Also, the linear convergence of such modification is proved.
The fifth chapter contains the original results, corresponding to conjugate gradient methods. Namely, one new hybrid conjugate gradient method is presented, and this method is the convex combination of two known conjugate gradient methods.
The sixth chapter consists of the numerical results, performed on a set of test functions, corresponding to methods in the chapters 3 and 5. Implementation of all considered algorithms is made in Mathematica. The comparison criterion is CPU time.
AB

Accepted on Scientific Board on: April 25, 2014
ASB
Defended:
DE

Thesis defend board:

President: Zorana Lužanin, PhD, Full Professor, Faculty of Sciences, University of Novi Sad

Mentor: Nataša Krejić, PhD, Full Professor, Faculty of Sciences,
University of Novi Sad

Member: Zorica Uzelac, PhD, Full Professor, Faculty of Technical Sciences, University of Novi Sad
DB