Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

Dunja Ćosić

Primjena Quasi - Newtonovih metoda na rješavanje sustava nelinearnih jednadžbi

Diplomski rad

Osijek, 2021.

Sveučilište J. J. Strossmayera u Osijeku
Odjel za matematiku

Dunja Ćosić

Primjena Quasi - Newtonovih metoda na rješavanje sustava nelinearnih jednadžbi

Diplomski rad

Mentor: prof. dr. sc. Kristian Sabo

Osijek, 2021.

Sadržaj

1 Uvod

U primjenjenoj matematici često je potrebno izračunati nultočku nelinearne funkcije. Obzirom da nultočku u velikoj većini slučajeva nije moguće egzaktno odrediti moramo posegnuti za numeričkim metodama pomoću kojih aproksimiramo rješenje danog problema. Primjeri takvih metoda su metoda bisekcije, metoda jednostavnih iteracija te Newtonova metoda tangenti. Metode se razlikuju po tome što zahtijevaju različite uvjete na funkciju. Primjerice metoda bisekcije zahtijeva samo neprekidnost funkcije, metoda jednostavnih iteracija zahtijeva da je prva derivacija funkcije neprekidna, dok Newtonova metoda zahtijeva neprekidnost druge derivacije funkcije. S druge strane metode se razlikuju po brzini konvergencije. Obično veći zahtjev na funkciju za posljedicu ima veću brzinu konvergencije metode, ali se često javlja problem sa složenim računanjem derivacije. Primjerice Newtonova metoda od navedenih metoda ima najveću brzinu konvergencije, ali je u njoj osim računanja vrijednosti funcije potrebno računati i derivaciju funkcije. Iz tog razloga razvijene su modifikacije Newtonove metode kod kojih se umjesto derivacije, koriste njezine aproksimacije, kod kojih se računa samo vrijednosti funkcije, a istovremeno nastoje se sačuvati dobra svojstva Newtonove metode. Primjer takve modifikacije je Newtonova metoda sekanti. Sve ove metode mogu se prirodno generalizirati na rješavanje sustava nelinearnih jednadžbi. Posebni naglasak u ovom radu stavljen je na modifikacije Newtonove metode, koje se u literaturi nazivaju Quasi-Newtonove metode.

Osim Uvoda, rad se sastoji od tri poglavlja. U drugom poglavlju upoznat ćemo se s Newtonovom metodom tangenti te s metodom sekanti za traženje nultočki nelinearnih jednadžbi i riješiti pripadne primjere. Treće poglavlje će nas uvesti u sustave nelinearnih jednadžbi. Prisjetit ćemo se pojmova poput Jacobijeve matrice, regularne matrice te Lispchitz-neprekidne funkcije. Izvest ćemo Newtonovu metodu za sustave nelinearnih jednadžbi te riješiti primjer. Najvažnije, a ujedno i zadnje poglavlje, bavi se Quasi-Newtonovim metodama. Ovo poglavlje upoznat će nas s Broydenovom, Davidon-Fletcher-Powellovom i Broyden-Fletcher-Goldfarb-Schannovom metodom. Za svaku od navedenih metoda navest ćemo formule i osnovna svojstva te riješiti po jedan primjer.

2 Rješavanje nelinearnih jednadžbi

Neka je $f : [a, b] \to \mathbb{R}$ neprekidna funkcija. Tražimo $\xi \in [a, b]$ tako da je

$$f(\xi) = 0. \tag{2.1}$$

Ako je funkcija f neprekidna na intervalu $[a, b]$ te ako je $f(a) \cdot f(b) < 0$ tada postoji barem jedna točka $\xi \in [a, b]$ takva da vrijedi $f(\xi) = 0$. Ukoliko je f derivabilna i prva derivacija f' stalnog predznaka na intervalu $I = [a, b]$, onda je to ujedno i jedina nultočka funkcije f na tom intervalu. Traženje realnog rješenja jednadžbe (2.1) sada se provodi kroz sljedeća dva koraka:

1. Separiranje intervala I u kojemu funkcija f ima nultočku ξ.

2. Određivanje aproksimacije nultočke ξ s unaprijed zadanom točnošću putem neke iterativne metode.

Kao što smo već ranije naveli postoje brojne metode za rješavanje jednadžbe (2.1), a mi ćemo u ovome radu opisati dvije koje će nam biti podloga za Quasi - Newtonove metode kojima ćemo se kasnije baviti.

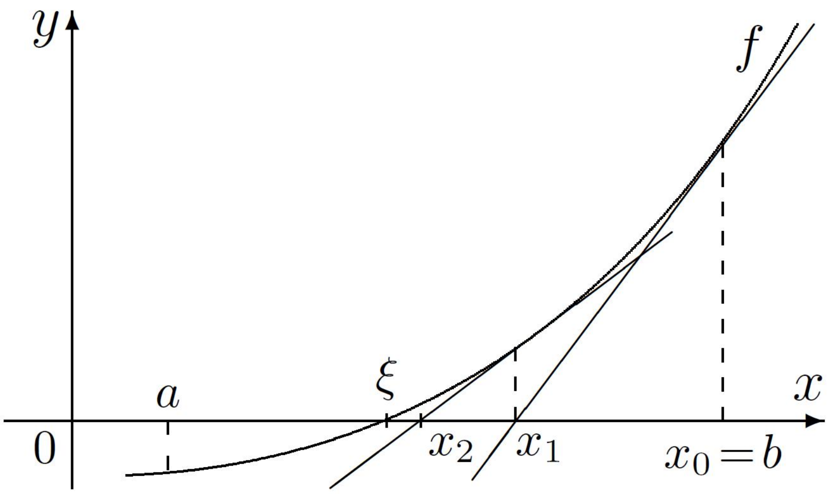
2.1 Newtonova metoda tangenti

Kod Newtonove metode funkciju f aproksimiramo njezinom tangentom u nekoj odabranoj točki $(x_0, f(x_0))$, a nultočku funkcije aproksimiramo nultočkom tangente. Geometrijski gledano, umjesto da tražimo sjecište grafa funkcije f i osi x tražit ćemo sjecište tangente i osi x.

Dakle, pretpostavimo da smo odredili interval $I = [a, b]$ u kojemu se nalazi jedinstvena nultočka neprekidne funkcije f za koju vrijedi $f'(x) \neq 0$, za svaki $x \in I$. Uzmimo $x_0 \in I$ kao početnu aproksimaciju. Razvijmo sada funkciju f u Taylorov red u okolini točke x_0 i promatrajmo samo linearan član. Razvojem funkcije f u Taylorov red dobili smo

$$f_1(x) = f(x_0) + (x - x_0) f'(x_0). \tag{2.2}$$

Na ovaj način smo funkciju f u okolini točke x_0 aproksimirali linearnom funkcijom čiji graf je tangenta na graf funkcije f u točki $(x_0, f(x_0))$ (Slika 2.1).



Slika 2.1 Newtonova metoda tangenti

Iz tog razloga ćemo rješavati jednadžbu $f_1(x) = 0$ umjesto početne jednadžbe (2.1). Označimo s x_1 rješenje jednadžbe $f_1(x) = 0$. Tada je

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

Povucimo sada tangentu na graf funkcije f kroz točku $(x_1, f(x_1))$. Dobivenu točku označimo s x_2 i dobivamo

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}.$$

Ponavljajući ovaj postupak povlačenja tangenti redom kroz točke $(x_0, f(x_0))$, $(x_1, f(x_1))$, ..., dobivamo niz x_0, x_1, x_2, ..., zadan rekurzivnom formulom

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},\ n = 0, 1, 2, \dots \tag{2.3}$$

Sada ćemo kroz sljedeći teorem, pokazati da uz početnu aproksimaciju x_0 te uz određene uvjete na funkciju f niz definiran s (2.3) konvergira prema jedinstvenom rješenju jednadžbe $f(x) = 0$.

Teorem 2.1. ***(vidi[4])*** *Neka funkcija $f : I \to \mathbb{R}$ ima neprekidnu drugu derivaciju na intervalu $I = [a, b]$. Neka je nadalje, $f(a) \cdot f(b) < 0$, a prva i druga derivacija funkcije f na intervalu I imaju stalan predznak.*

Tada, ako je $x_0 \in I$ izabran tako da je

$$f(x_0) \cdot f''(x_0) > 0, \tag{2.4}$$

onda niz definiran s (2.3) konvergira prema jedinstvenom rješenju ξ jednadžbe $f(x) = 0$.

Pri tome vrijedi ocjena pogreške aproksimacije

$$|\xi - x_n| \le \frac{M_2}{2m_1}(x_n - x_{n-1})^2 \tag{2.5}$$

gdje su

$$m_1 = \min_{x \in I} |f'(x)|,$$
$$M_2 = \max_{x \in I} |f''(x)|$$

te vrijedi

$$|\xi - x_{n+1}| \leq \frac{M_2}{2m_1}(\xi - x_n)^2. \tag{2.6}$$

Prije dokaza Teorema 2.1 uvest ćemo definiciju brzine konkvergencije.

Definicija 2.1. *Neka niz (x_n), dobiven nekom iterativnom metodom, konvergira prema $\xi \in \mathbb{R}$ i neka je $\Delta x_n = |\xi - x_n|$ apsolutna pogreška n-te aproksimacije. Tada, ako postoje dvije pozitivne konstante A, $r \in \mathbb{R}_+$, takve da vrijedi*

$$\lim_{n \to \infty} \frac{\Delta x_{n+1}}{(\Delta x_n)^r} = A,$$

kažemo da metoda ima brzinu konvergencije r.

Specijalno, za $r = 1$ kažemo da metoda ima linearnu brzinu konvergencije, a za $r = 2$ da ima kvadratnu brzinu konvergencije. Ako je $r \in \langle 1, 2 \rangle$, onda kažemo da metoda ima superlinearnu brzinu konvergencije. Primijetimo da Definicija 2.1. kaže kako metoda ima brzinu konvergencije r ako postoji $A \in \mathbb{R}_+$ i $n_0 \in \mathbb{N}$, tako da je

$$\Delta x_{n+1} \approx A(\Delta x_n)^r, \quad n \geq n_0.$$

Sada, kada smo definirali brzinu konvergencije, iz (2.6) u iskazu prethodnog teorema vidimo da Newtonova metoda ima kvadratnu brzinu konvergencije.

U nastavku ćemo dokazati Teorem 2.1. U tu svrhu koristi se dokaz odgovarajućeg teorema iz [4], uz određene modifikacije.

Dokaz. Teorema 2.1. Bez smanjenja općenitosti možemo pretpostaviti da je $f(a) > 0$, $f(b) < 0$, $f'(x) < 0$, $f''(x) < 0$, za svaki $x \in I$. Primijetimo da onda postoji jedinstveni $\xi \in I$ takav da je $f(\xi) = 0$ i izaberimo $x_0 \in I$ tako da je zadovoljeno (2.4) (npr. možemo uzeti $x_0 = b$). Obzirom da je prema pretpostavci $f''(x) < 0$, onda je i $f(x_0) < 0$ zbog $f(x_0) \cdot f''(x_0) > 0$ pa je $x_0 > \xi$, jer je f padajuća funkcija ($f(x_0) < f(\xi)$).

Ideja dokaza je najprije pokazati da je niz (x_n), odnosno niz definiran rekurzivno s (2.3), konvergentan.

Pokažimo prvo da je niz ograničen odozdo s ξ. Dokaz provodimo metodom matematičke indukcije. U tu svrhu pretpostavimo da je $x_n > \xi$ te dokažimo da je $x_{n+1} > \xi$. Taylorov razvoj funkcije f u okolini točke x_n glasi:

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{f''(c_1)}{2}(x - x_n)^2,$$

pri čemu je c_1 neki broj između x i x_n. Specijalno, za $x = \xi$ dobivamo

$$f(\xi) = f(x_n) + f'(x_n)(\xi - x_n) + \frac{f''(c_2)}{2}(\xi - x_n)^2,$$

odnosno

$$0 = f(x_n) + f'(x_n)(\xi - x_n) + \frac{f''(c_2)}{2}(\xi - x_n)^2,$$

pri čemu je c_2 neki broj između ξ i x_n. Obzirom da je $f''(c_2) < 0$ i da zbog pretpostavke $x_n > \xi$ vrijedi $\xi - x_n < 0$, slijedi da je $\frac{f''(c_2)}{2}(\xi - x_n)^2 < 0$ pa mora biti

$$f(x_n) + f'(x_n)(\xi - x_n) > 0.$$

Iz te nejednakosti sada dobivamo

$$f'(x_n)\xi > f'(x_n)x_n - f(x_n),$$

odnosno (zbog $f'(x_n) < 0$)

$$\xi < x_n - \frac{f(x_n)}{f'(x_n)} = x_{n+1}$$

pa je onda i $x_{n+1} > \xi$. Dakle, niz (x_n) je ograničen odozdo s ξ.

Sada, kada smo pokazali da je niz ograničen, trebamo pokazati i da je monoton. Znamo da je $x_n > \xi$ za sve $n \in \mathbb{N}$ pa onda zbog strogog monotonog pada funkcije f mora vrijediti $f(x_n) < f(\xi)$. Obzirom da je $f(\xi) = 0$, slijedi $f(x_n) < 0$, za svaki $n \in \mathbb{N}$. Zato iz (2.3) imamo

$$x_n - x_{n+1} = x_n - \left(x_n - \frac{f(x_n)}{f'(x_n)} \right) = \frac{f(x_n)}{f'(x_n)} > 0,$$

a to znači da je niz (x_n) monotono padajući.

Dakle, pokazali smo da je niz (x_n) konvergentan. Kako se cijeli niz (x_n) nalazi u I zbog toga mora postojati realan broj $\epsilon \in I$ takav da je $\epsilon = \lim_{n\to\infty} x_n$. Kada u (2.3) pustimo $n \to \infty$ dobivamo

$$\epsilon = \epsilon - \frac{f(\epsilon)}{f'(\epsilon)},$$

odakle slijedi $f(\epsilon) = 0$ (zbog $f'(\epsilon) < 0$). Kako je ϵ jedinstvena nultočka funkcije f mora biti $\epsilon = \xi$.

Preostalo je dokazati ocjenu pogreške (2.5). Razvojem funkcije f u Taylorov red u okolini točke x_{n-1} dobivamo

$$f(x_n) = f(x_{n-1}) + f'(x_{n-1})(x_n - x_{n-1}) + \frac{f''(c)}{2}(x_n - x_{n-1})^2.$$

Zbog (2.3) imamo

$$f(x_n) = f(x_{n-1}) + f'(x_{n-1}) \left(x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} - x_{n-1} \right) + \frac{f''(c)}{2}(x_n - x_{n-1})^2,$$

odnosno

$$f(x_n) = \frac{f''(c)}{2}(x_n - x_{n-1})^2.$$

Odavde je

$$|f(x_n)| = \frac{1}{2}|f''(c)|(x_n - x_{n-1})^2 \le \frac{M_2}{2}(x_n - x_{n-1})^2. \tag{2.7}$$

Sada iz

$$|\xi - x_n| \le \frac{|f(x_n)|}{m_1}$$

i izraza (2.7) dobivamo ocjenu (2.5).

Dokažimo još da metoda ima kvadratnu brzinu konvergencije. Razvojem funkcije f u Taylorov red u okolini tčke x_n imamo

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{f''(c)}{2}(x - x_n)^2.$$

Stavimo $x = \xi$ i dobivamo

$$0 = f(\xi) = f(x_n) + f'(x_n)(\xi - x_n) + \frac{f''(c)}{2}(\xi - x_n)^2,$$

gdje je c realan broj između ξ i x_n. Dijeljenjem s $f'(x_n)$ dobivamo

$$0 = \frac{f(x_n)}{f'(x_n)} + (\xi - x_n) + \frac{f''(c)}{2f'(x_n)}(\xi - x_n)^2.$$

Iskoristimo sada (2.3) i dobivamo

$$x_{n+1} - \xi = \frac{f''(c)}{2f'(x_n)}(\xi - x_n)^2,$$

odnosno imamo

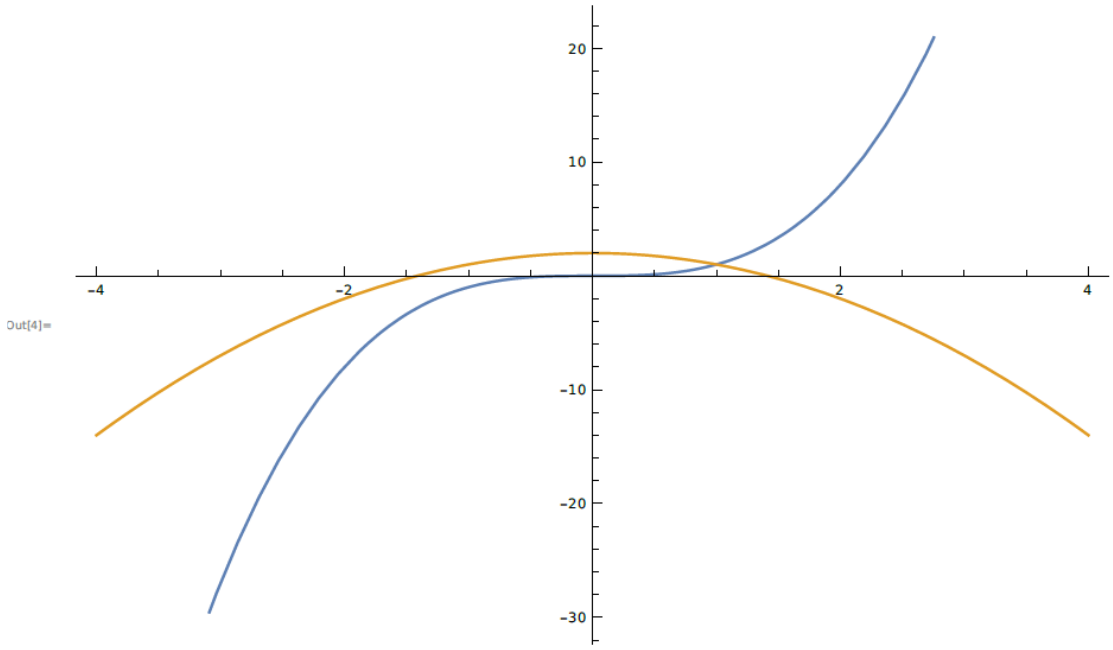
$$|x_{n+1} - \xi| = \frac{1}{2}\frac{|f''(c)|}{|f'(x_n)|}(x_n - \xi)^2 \le \frac{M_2}{2m_1}(x_n - \xi)^2$$

što je tvrdnja koju smo trebali dokazati. □

U sljedećem primjeru ilustriramo primjenu Newtonove metode.

Primjer 2.1. *Odredite nultočku funkcije $f(x) = x^3 + x^2 - 2$ koristeći Newtonovu metodu tangenti te postupak prekinite u trenutku $|x_n - \xi| < \epsilon = 0.0005$ gdje je ξ prava nultočka funkcije f.*

Rješenje: Prvo ćemo separirati rješenje jednadžbe kako bi mogli odrediti interval I u kojemu se ono nalazi. Promotrimo Sliku 2.2 koja predstavlja separaciju rješenja.



Slika 2.2 Separacija rješenja funkcije $f(x) = x^3 + x^2 - 2$

Sa slike je vidljivo kako za naš interval I možemo uzeti $I = [0.5, 3]$. Provjerimo zadovoljava li ovaj interval uvjete Teorema 2.1. Prvo što mora biti zadovoljeno je $f(a) \cdot f(b) < 0$. U našem slučaju imamo

$$f(0.5) \cdot f(3) = (-1.625) \cdot 34 < 0$$

pa je prvi uvjet zadovoljen. Provjerimo još imaju li f' i f'' na našem intervalu I stalan predznak. Prva derivacija funkcije f je

$$f'(x) = 3x^2 + 2x > 0, \quad \forall x \in I,$$

a druga derivacija

$$f''(x) = 6x + 2 > 0, \quad \forall x \in I.$$

Dakle, pokazali smo da su f' i f'' stalnog predznaka na $I = [0.5, 3]$.
Odredimo sada m_1 i M_2:

$$m_1 = \min_{x \in I} |f'(x)| = 1.75,$$
$$M_2 = \max_{x \in I} |f''(x)| = 20.$$

Znamo da ova funkcija ima jedinstvenu nultočku i to je $\xi = 1$. Za x_0 ćemo uzeti $x_0 = 3$ jer mora vrijediti $f(x_0) \cdot f''(x_0) > 0$.
Za $n = 0$ računamo x_1 po formuli

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

i dobivamo

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 3 - \frac{34}{33} = 1.9697.$$

Potrebno je izračunati i $|x_n - \xi|$. Uvrštavanjem dobivamo

$$|x_1 - 1| = 0.9697$$

što je veće od $\epsilon = 0.0005$ pa postupak ponavljamo za $n = 1$.
U slučaju kada je $n = 1$ imamo

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 1.3585$$

i

$$|x_2 - 1| = 0.3585.$$

Vidimo da je dobivena vrijednost opet veća od 0.0005 te nastavljamo postupak.
Za x_3 dobivamo

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 1.07345.$$

i

$$|x_3 - 1| = 0.07345,$$

a to je opet veće od 0.0005 pa nastavljamo postupak.
Sljedeća iteracija je

$$x_4 = x_3 - \frac{f(x_3)}{f'(x_3)} = 1.00399$$

i vrijedi

$$|x_4 - 1| = 0.00399 > \epsilon.$$

Za x_5 dobivamo

$$x_5 = x_4 - \frac{f(x_4)}{f'(x_4)} = 1.00001$$

i

$$|x_5 - 1| = 0.00001,$$

a to je manje od 0.0005 pa je naša aproksimacija $x_5 = 1.00001$.

2.2 Metoda sekanti

Jedna modifikacija Newtonove metode je metoda sekanti. Metodom sekanti, za razliku od Newtonove metode, izbjegavamo izračunavanje derivacije funkcije u svakoj pojedinoj iteraciji što uvelike olakšava postupak. Na intervalu $I = [a, b]$ izabiremo dvije početne aproksimacije x_0 i x_1 te kroz točke $(x_0, f(x_0))$ i $(x_1, f(x_1))$ povlačimo sekantu na graf funkcije f. Jednadžba te sekante je

$$y = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0).$$

Aproksimaciju x_2 dobit ćemo kao sjecište sekante s osi x, odnosno izjednačavajući prethodnu jednadžbu s nulom (Slika 2.3):

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)},$$

ako je $f(x_1) \neq f(x_0)$. Iterativnim ponavljanjem ovog postupka dobivamo niz definiran formulom

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

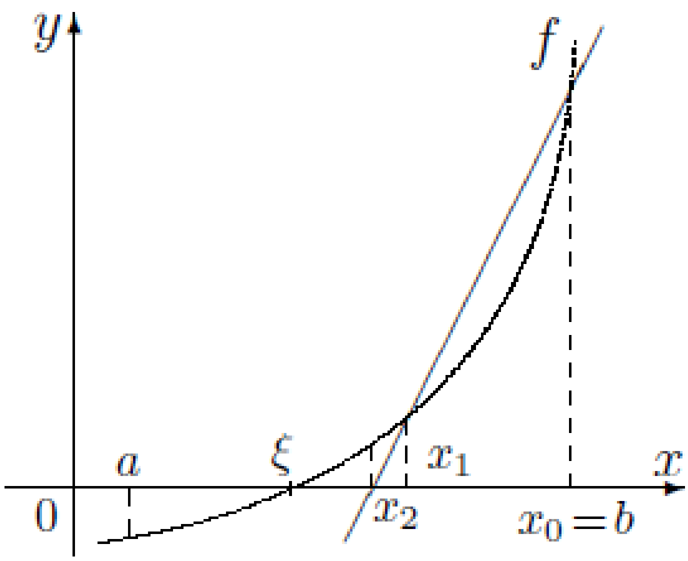
za $f(x_n) \neq f(x_{n-1})$, $n = 1, 2, \ldots$

Ocjena apsolutne pogreške ove metode je

$$|\xi - x_n| \leq \frac{|f(x_n)|}{m_1},$$

pri čemu je

$$m_1 = \min_{x \in I} |f'(x)|.$$



Slika 2.3 Metoda sekanti

U nastavku ćemo pokazati kako Newtonova metoda sekanti ima superlinarnu brzunu konvergencije. Odnosno ako su početne aproksimacije x_0 i x_1 izabrane dovoljno blizu nultočke $x^\star$, onda niz iteracija (x_n) dobiven metodom sekante konvergira prema $x^\star$ s brzinom konvergencije p, gdje je

$$p = \frac{1 + \sqrt{5}}{2} \approx 1.618.$$

Neka je x_n aproksimacija dobivena metodom sekante. Označimo li s $e_n = x_n - x^\star$, razvojem funkcije f u Taylorov red u okolini točke $x^\star$ te uvrštavanjem $x = x_{n-1}$, odnosno $x = x_n$ dobivamo:

$$f(x_{n-1}) = f(e_{n-1} + x^\star) = f(x^\star) + e_{n-1} f'(x^\star) + \frac{e_{n-1}^2}{2} f''(x^\star) + O(e_{n-1}^3)$$

te

$$f(x_n) = f(e_n + x^\star) = f(x^\star) + e_n f'(x^\star) + \frac{e_n^2}{2} f''(x^\star) + O(e_n^3).$$

Uvrštavanjem $e_n = x_n - x^\star$ u metodu sekante

$$x_{n+1} = x_n - f(x_n) \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

dobivamo

$$e_{n+1} + x^\star = e_n + x^\star - f(e_n + x^\star) \frac{(e_n + x^\star) - (e_{n-1} + x^\star)}{f(e_n + x^\star) - f(e_{n-1} + x^\star)},$$

odnosno

$$e_{n+1} = e_n - \frac{f(e_n + x^\star)(e_n - e_{n-1})}{f(e_n + x^\star) - f(e_{n-1} + x^\star)}. \tag{2.9}$$

Uvedemo li oznaku $M := \frac{f''(x^\star)}{2f'(x^\star)}$, vrijedi $\frac{f''(x^\star)}{2} = M f'(x^\star)$.

Raspisivanjem brojnika u izrazu (2.9), dobivamo

$$f(e_n + x^\star)(e_n - e_{n-1}) = \left(f(x^\star) + e_n f'(x^\star) + \frac{e_n^2}{2} f''(x^\star) \right)(e_n - e_{n-1}),$$

tj.

$$\begin{aligned} f(e_n + x^\star)(e_n - e_{n-1}) &= \left(e_n f(x^\star) + M e_n^2 f'(x^\star) \right)(e_n - e_{n-1}) \\ &= e_n f'(x^\star)(1 + M e_n)(e_n - e_{n-1}). \end{aligned} \tag{2.10}$$

Raspišimo sada nazivnik izraza (2.9). Imamo

$$\begin{aligned} f(e_n + x^\star) - f(e_{n-1} + x^\star) &= f(x^\star) + e_n f'(x^\star) + \frac{e_n^2}{2} f''(x^\star) - f(x^\star) - e_{n-1} f'(x^\star) \\ &\quad + \frac{e_{n-1}^2}{2} f''(x^\star) \\ &= f'(x^\star)(e_n - e_{n-1}) + M f'(x^\star)(e_n^2 - e_{n-1}^2) \\ &= f'(x^\star)(e_n - e_{n-1})\left(1 + M(e_n + e_{n-1})\right). \end{aligned} \tag{2.11}$$

Sada iz (2.10) i (2.11) izraz (2.9) možemo zapisati kao

$$e_{n+1} = e_n - \frac{e_n f'(x^\star)(1 + M e_n)(e_n - e_{n-1})}{f'(x^\star)(e_n - e_{n-1})\left(1 + M(e_n + e_{n-1})\right)},$$

tj.

$$e_{n+1} = e_n - \frac{e_n(1 + M e_n)}{1 + M(e_n + e_{n-1})} = \frac{M e_n e_{n-1}}{1 + M(e_n + e_{n-1})}.$$

Dakle, gornji izraz implicira[1]

$$e_{n+1} \approx M e_n e_{n-1} = \frac{f''(x^\star)}{2f'(x^\star)} e_n e_{n-1}.$$

Preostalo nam je izračunati eksponent p. Iz definicije brzine konvergencije je $|e_{n+1}| \approx C|e_n|^p$. U našem slučaju je

$$|e_{n+1}| \approx |M||e_n||e_{n-1}|,$$

odnosno

$$\begin{aligned} C|e_n|^p &\approx |M||e_n||e_{n-1}| \\ |e_n|^{p-1} &\approx \frac{|M|}{C}|e_{n-1}| \\ |e_n| &\approx \left(\frac{|M|}{C}\right)^{\frac{1}{p-1}} |e_{n-1}|^{\frac{1}{p-1}}. \end{aligned}$$

Dakle, mora biti $p = \frac{1}{p-1}$. Obzirom da je $p > 0$ jedino rješenje ove jednadžbe je

$$p = \frac{1+\sqrt{5}}{2} \approx 1.618.$$

Riješimo sada isti primjer koji smo riješili pomoću Newtonove metode tangenti kako bi se uvjerili da je metoda sekanti uistinu sporija.

Primjer 2.2. *Odredite nultočku funkcije $f(x) = x^3 + x^2 - 2$ koristeći metodu sekanti te postupak prekinite u trenutku kada je $|x_n - \xi| < \epsilon$.*

Rješenje: Obzirom da je riječ o istoj funkciji kao i u Primjeru 2.1. ne moramo ponovno separirati rješenje već uzimamo da nam je interval I isti kao u prethodnom primjeru, tj. $I = [0.5, 3]$. Također, ranije smo već pokazali kako je ispunjen uvjet $f(a) \cdot f(b) < 0$, tj.

$$f(0.5) \cdot f(3) = (-1.625) \cdot 34 < 0.$$

Obzirom da se i m_1 računa na isti način kao i kod Newtonove metode tangenti imamo

$$m_1 = \min_{x \in I} |f'(x)| = 1.75.$$

---

[1] Ovo je posljedica činjenice da se funkcija $\psi(x,y) = \frac{Mxy}{1+M(x+y)}$ u okolini točke $(0,0)$ može aproksimirati s funkcijom $\phi(x,y) = Mxy$.

Razvojem funkcije ψ u Taylorov red u okolini točke $(0,0)$ do kvadratnog člana dobiva se

$$\begin{aligned} \frac{Mxy}{1+M(x+y)} &\approx \psi(0,0) + \psi'_x(0,0)(x-0) + \psi'_y(0,0)(y-0) \\ &+ \frac{1}{2}\left[\psi''_{xx}(0,0)(x-0)^2 + 2\psi''_{xy}(0,0)(x-0)(y-0) + \psi''_{yy}(0,0)(y-0)^2\right] \\ &= \frac{1}{2} \cdot 2Mxy = Mxy. \end{aligned} \tag{2.12}$$

Derivacije $f'_x(0,0)$, $f'_y(0,0)$,$f''_{xx}(0,0)$, $f''_{yy}(0,0)$ su jednake nuli.

Za razliku od metode tangenti, ovdje su nam potrebne dvije početne aproksimacije x_0 i x_1. Za x_0 uzimamo vrijednost b iz intervala $I = [a, b] = [0.5, 3]$, tj. $x_0 = 3$, a za x_1 uzmimo $x_1 = 2.5$. Računamo x_2 po formuli

$$x_{n+1} = \frac{x_{n-1} f(x_n) - x_n f(x_{n-1})}{f(x_n) - f(x_{n-1})}$$

pa dobivamo

$$x_2 = \frac{x_0 f(x_1) - x_1 f(x_0)}{f(x_1) - f(x_0)} = \frac{3 \cdot 19.875 - 2.5 \cdot 34}{19.875 - 34} = 1.79646.$$

Izračunajmo sada $|x_n - \xi|$ za $x_n = x_2$. Uvrštavanjem dobivamo

$$|x_2 - 1| = 0.79646.$$

Obzirom da je dobivena vrijednost 0.79646 veća od $\epsilon = 0.0005$ postupak ponavljamo za x_3. Računanjem dobivamo

$$x_3 = \frac{x_1 f(x_2) - x_2 f(x_1)}{f(x_2) - f(x_1)} = 1.41185$$

i

$$|x_3 - 1| = 0.41185.$$

Dobivena vrijednost opet je veća od 0.0005 te nastavljamo postupak.
Za x_4 dobivamo

$$x_4 = \frac{x_2 f(x_3) - x_3 f(x_2)}{f(x_3) - f(x_2)} = 1.15580$$

pa slijedi

$$|x_4 - 1| = 0.15580 > 0.0005.$$

U sljedećoj iteraciji dobivamo

$$x_5 = \frac{x_3 f(x_4) - x_4 f(x_3)}{f(x_4) - f(x_3)} = 1.03893$$

i

$$|x_5 - 1| = 0.03893 > 0.0005.$$

Za x_6 ćemo dobiti $x_6 = 1.00438$ te će nam pripadno odstupanje $|x_6 - 1| = 0.00438$ opet biti veće od ϵ. Aproksimacija x_7 jednaka je

$$x_7 = 1.00013$$

pa dobivamo

$$|x_7 - 1| = 0.00013,$$

a ta je vrijednost manja od $\epsilon = 0.0005$ pa ovdje završavamo postupak. Dobivena iteracija je $x_7 = 1.00013$.

Pogledajmo sada tablicu aproksimacija nultočki funkcije $f(x) = x^3 + x^2 - 2$ dobivenih pomoću Newtonove metode tangenti te metode sekanti za $\epsilon = 0.0005$.

n	x_n	$\|x_n - \xi\|$	x_n	$\|x_n - \xi\|$
0	3	2	3	2
1	1.9697	0.9697	2.5	1.5
2	1.3585	0.3585	1.79646	0.79646
3	1.07345	0.07345	1.41185	0.41185
4	1.00399	0.00399	1.15580	0.15580
5	1.00001	0.00001	1.03893	0.03893
6	-	-	1.00438	0.00438
7	-	-	1.00013	0.00013

Tablica 2.1: Newtonova metoda tangenti i metoda sekanti za $f(x) = x^3 + x^2 - 2$

Iz tablice je vidljivo kako je Newtonova metoda tangenti do rješenja došla u 5 koraka, a metoda sekanti u 7 iteracija.

3 Sustavi nelinearnih jednadžbi

Pretpostavimo da rješavamo sustav nelinearnih jednadžbi

$$f_i(x_1, x_2, ..., x_n) = 0, \tag{3.1}$$

pri čemu je $i \in \{1, ..., n\}$ i $f_i : \mathbb{R}^n \to \mathbb{R}$. Ukoliko uvedemo vektorske oznake onda sustav (3.1) možemo zapisati kao $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{0}$, pri čemu je $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ te

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ f_2(\boldsymbol{x}) \\ \vdots \\ f_n(\boldsymbol{x}) \end{bmatrix}, \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Ovakve sustave rješavamo iterativnim metodama. Prije nego krenemo opisivati pojedinu metodu za rješavanje sustava nelinearnih jednadžbi, definirat ćemo pojmove koji će nam biti potrebni.

Definicija 3.1. *Neka je* $\boldsymbol{F} = (f_1, ..., f_n) : \mathbb{R}^n \to \mathbb{R}^n$ *neprekidna diferencijabilna funkcija i* $\boldsymbol{x} = (x_1, ..., x_n)$. *Matricu*

$$\mathbb{J} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \cdots & \vdots & \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

nazivamo Jacobijeva matrica.

Definicija 3.2. ***(vidi [5])*** *Za matricu* $\boldsymbol{A} \in M_n$ *kažemo da je regularna (nesingularna) ako postoji matrica* $\boldsymbol{B} \in M_n$ *takva da je*

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}.$$

Kažemo da je matrica $\boldsymbol{C} \in M_n$ *singularna ako nije regularna.*

Definicija 3.3. *Kažemo da je funkcija* $f : \mathscr{D} \to \mathbb{R}$, $\mathscr{D} = [a, b]$ *Lipsichtz-neprekidna s konstantom* $L > 0$ *na* $\mathscr{D}$ *i pišemo* $f \in Lip_L\mathscr{D}$ *ako za sve* $x, y \in \mathscr{D}$ *vrijedi*

$$|f(x) - f(y)| \leq L|x - y|.$$

Konstantu $L > 0$ *nazivamo Lipschitzova konstanta.*

3.1 Newtonova metoda

Newtonova metoda je jedna od najpoznatijih metoda za rješavanje sustava nelinearnih jednadžbi. Uočimo kako je ova metoda generalizacija Newtonove metode za rješavanje nelinearnih jednadžbi. U nastavku ćemo opisati Newtonovu metodu.

Izaberimo početnu aproksimaciju $\boldsymbol{x}^{(0)}=[x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}]^T \in \mathbb{R}^n$. Svaku od funkcija f_1, $f_2, \ldots, f_n$ razvijmo u Taylorov red u okolini $\boldsymbol{x}^{(0)}$. Dobivenu odgovarajuću linearnu aproksimaciju označimo s $\tilde{f}_i$:

$$\tilde{f}_i(\boldsymbol{x}) = f_i(\boldsymbol{x}^{(0)}) + \sum_{j=1}^{n} \frac{\partial f_i(\boldsymbol{x}^{(0)})}{\partial x_j}(x_j - x_j^{(0)}),$$

za sve $i = 1, \ldots, n$. Sada više ne rješavamo sustav (3.1) već umjesto njega rješavamo sustav $\tilde{f}_i(\boldsymbol{x}) = 0$, $i = 1, \ldots, n$. Sustav $\tilde{f}_i(\boldsymbol{x}) = 0$ možemo zapisati i u matričnom obliku te tada imamo

$$\mathbb{J}^{(0)}\boldsymbol{s} = -\boldsymbol{f}(\boldsymbol{x}^{(0)}), \tag{3.2}$$

pri čemu je $\mathbb{J}^{(0)}=\begin{bmatrix} \frac{\partial f_1(\boldsymbol{x}^{(0)})}{\partial x_1} & \ldots & \frac{\partial f_1(\boldsymbol{x}^{(0)})}{\partial x_n} \\ \vdots & \ldots & \vdots \\ \frac{\partial f_n(\boldsymbol{x}^{(0)})}{\partial x_1} & \ldots & \frac{\partial f_n(\boldsymbol{x}^{(0)})}{\partial x_n} \end{bmatrix}$ Jacobijeva matrica u točki $\boldsymbol{x}^{(0)}$, $\boldsymbol{s}=\begin{bmatrix} x_1 - x_1^{(0)} \\ \vdots \\ x_n - x_n^{(0)} \end{bmatrix}$, $\boldsymbol{f}(\boldsymbol{x}^{(0)})=\begin{bmatrix} f_1(\boldsymbol{x}^{(0)}) \\ \vdots \\ f_n(\boldsymbol{x}^{(0)}) \end{bmatrix}$. Na ovaj način smo došli do nove aproksimacije rješenja koja je oblika

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} + \boldsymbol{s}^{(0)},$$

pri čemu je $\boldsymbol{s}^{(0)}$ rješenje sustava (3.2). Ponavljajući ovaj postupak za $\boldsymbol{x}^{(1)}$, dobivamo aproksimaciju $\boldsymbol{x}^{(2)}$

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} + \boldsymbol{s}^{(1)},$$

gdje je $\boldsymbol{s}^{(1)}$ rješenje sustava

$$\mathbb{J}^{(1)}\boldsymbol{s} = -\boldsymbol{f}(\boldsymbol{x}^{(1)}).$$

Općenito, dolazimo do iterativnog postupka

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{s}^{(k)}, \tag{3.3}$$

$k = 0, 1, \ldots$, gdje je $\boldsymbol{s}^{(k)}$ rješenje sustava

$$\mathbb{J}^{(k)}\boldsymbol{s} = -\boldsymbol{f}(\boldsymbol{x}^{(k)}), \tag{3.4}$$

a $\mathbb{J}^{(k)}$ Jacobijeva matrica u točki $\boldsymbol{x}^{(k)} \in \mathbb{R}^n$.

Ukoliko pretpostavimo da je Jacobijeva matrica $\mathbb{J}^{(k)}$ regularna onda umjesto (3.3) i (3.4) možemo pisati

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - (\mathbb{J}^{(k)})^{-1}\boldsymbol{f}(\boldsymbol{x}^{(k)}), \quad k = 0, 1, \ldots \tag{3.5}$$

Primjetimo da izraz (3.5) podsjeća na Newtonovu metodu tangente (vidi [3]).

U slučaju da je Jacobijan nesingularan u svakoj iteraciji te ako početnu aproksimaciju $\boldsymbol{x}^{(0)}$ izaberemo dovoljno blizu rješenja $\boldsymbol{x}^*$ onda Newtonova metoda kvadratnom brzinom konvergira prema rješenju sustava (3.1). To bi značilo da Newtonova metoda ne mora uvijek konvergirati, ali ukoliko dobro procijenimo početnu aproksimaciju, Newtonova metoda će brzo doći do pravog rješenja.

Teorem 3.1. ***(vidi[6])*** *Neka je $\boldsymbol{F}:\mathbb{R}^n \to \mathbb{R}^n$ neprekidno diferencijabilna na otvorenom konveksnom skupu $\mathscr{D}$, tj $\boldsymbol{F} \in C^1(\mathscr{D})$ i neka postoji $\boldsymbol{x}^\star \in \mathscr{D}$ tako da je $\boldsymbol{F}(\boldsymbol{x}^\star) = 0$ i $\boldsymbol{F}'(\boldsymbol{x}^\star)$ nesingularna matrica. Nadalje, neka je $\boldsymbol{F}'$ Lipschitz neprekidna pri čemu je $\mathscr{D}$ kugla oko $\boldsymbol{x}^\star$ polumjera r. Također, pretpostavimo da je $||\boldsymbol{F}'(\boldsymbol{x}^\star)^{-1}|| \leq \beta$. Ako početnu iteraciju $\boldsymbol{x}^{(0)}$ izaberemo tako da $||\boldsymbol{x}^{(0)} - \boldsymbol{x}^\star|| \leq \min\{r, 1/(2\beta L)\}$, tada je niz $(\boldsymbol{x}_n)$ definiran Newtonovim iterativnim postupkom dobro definiran i konvergira prema $\boldsymbol{x}^\star$ kvadratnom brzinom, tj.*

$$||\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^\star|| \leq \beta L ||\boldsymbol{x}^{(k)} - \boldsymbol{x}^\star||^2, \qquad k = 0, 1, \ldots$$

Primjer 3.1. *Pomoću Newtonove metode riješite sustav nelinearnih jednadžbi*

$$x_1^2 + x_2^2 - 4 = 0$$

$$x_1^3 + x_2 = 0$$

uz početnu aproksimaciju $\boldsymbol{x}^{(0)} = [1, -1]^T$. Postupak prekinite nakon treće iteracije.

Rješenje: Stavimo

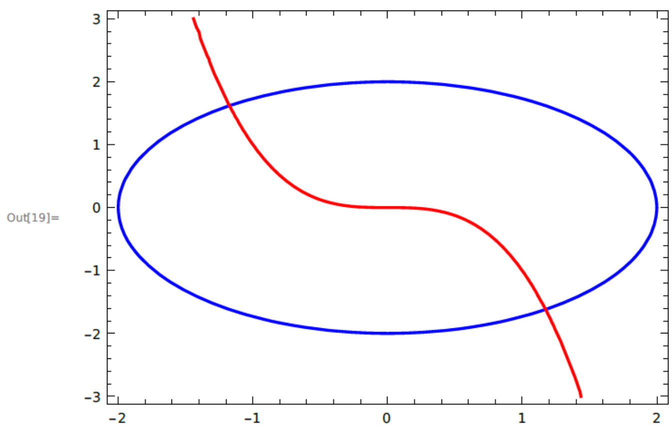
$$f_1(x_1, x_2) = x_1^2 + x_2^2 - 4$$

$$f_2(x_1, x_2) = x_1^3 + x_2.$$

Tada je

$\boldsymbol{f}(\boldsymbol{x}^{(k)}) = \begin{bmatrix} f_1(x_1^{(k)}, x_2^{(k)}) \\ f_2(x_1^{(k)}, x_2^{(k)}) \end{bmatrix}$, za $k = 0, 1, 2, \ldots$

Pogledajmo grafički prikaz ovih funkcija na Slici 3.1.



Slika 3.1 Contour Plot funkcija f_1 i f_2

Potrebno je izračunati i Jacobijan $\mathbb{J}^{(k)}$ pa u tu svrhu računamo

$$\frac{\partial f_1(x_1,x_2)}{\partial x_1} = 2x_1, \qquad \frac{\partial f_1(x_1,x_2)}{\partial x_2} = 2x_2,$$

$$\frac{\partial f_2(x_1,x_2)}{\partial x_1} = 3x_1^2, \qquad \frac{\partial f_2(x_1,x_2)}{\partial x_2} = 1.$$

Sada, za $k = 0,1,2,...$, imamo $\mathbb{J}^{(k)} = \begin{bmatrix} 2x_1^{(k)} & 2x_2^{(k)} \\ 3(x_1^{(k)})^2 & 1 \end{bmatrix}$.

Konkretno, za $k = 0$, odnosno za početnu aproksimaciju $\boldsymbol{x}^{(0)} = [1,-1]^T$ imamo

$\mathbb{J}^{(0)} = \begin{bmatrix} 2x_1^{(0)} & 2x_2^{(0)} \\ 3(x_1^{(0)})^2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -2 \\ 3 & 1 \end{bmatrix}$.

Svaku sljedeću aproksimaciju računamo po formuli (3.5). Za $k = 0$ dobivamo aproksimaciju $\boldsymbol{x}^{(1)}$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - (\mathbb{J}^{(0)})^{-1}\boldsymbol{f}(\boldsymbol{x}^{(0)})$$

$$\boldsymbol{x}^{(1)} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \begin{bmatrix} 2 & -2 \\ 3 & 1 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - \frac{1}{8}\begin{bmatrix} 1 & 2 \\ -3 & 2 \end{bmatrix}\begin{bmatrix} -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.25 \\ -1.75 \end{bmatrix}.$$

Sada, za $\boldsymbol{x}^{(1)} = \begin{bmatrix} 1.25 \\ -1.75 \end{bmatrix}$, računamo $\mathbb{J}^{(1)}$ i dobivamo

$\mathbb{J}^{(1)} = \begin{bmatrix} 2.5 & -3.5 \\ 4.6875 & 1 \end{bmatrix}$.

Aproksimacija $\boldsymbol{x}^{(2)}$ tada iznosi

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} - (\mathbb{J}^{(1)})^{-1}\boldsymbol{f}(\boldsymbol{x}^{(1)}) = \begin{bmatrix} 1.25 \\ -1.75 \end{bmatrix} - \begin{bmatrix} 2.5 & -3.5 \\ 4.6875 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.625 \\ 0.2031 \end{bmatrix} = \begin{bmatrix} 1.1793 \\ -1.6219 \end{bmatrix}.$$

U zadatku se traže prve tri iteracije pa nam preostaje izračunati aproksimaciju $\boldsymbol{x}^{(3)}$. U tu svrhu računamo pripadajući Jacobijan $\mathbb{J}^{(2)}$ te dobivamo

$\mathbb{J}^{(2)} = \begin{bmatrix} 2.3587 & -3.2438 \\ 4.1725 & 1 \end{bmatrix}$.

Na isti način na koji smo dobili aproksimacije $\boldsymbol{x}^{(1)}$ i $\boldsymbol{x}^{(2)}$ dobivamo i aproksimaciju $\boldsymbol{x}^{(3)}$ pa imamo

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} - (\mathbb{J}^{(2)})^{-1}\boldsymbol{f}(\boldsymbol{x}^{(2)}) = \begin{bmatrix} 1.1793 \\ -1.6219 \end{bmatrix} - \begin{bmatrix} 2.3587 & -3.2438 \\ 4.1725 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.0214 \\ 0.0184 \end{bmatrix} = \begin{bmatrix} 1.1742 \\ -1.619 \end{bmatrix}.$$

Dakle, treća iteracija danog sustava iznosi $\boldsymbol{x}^{(3)} = [1.1742,-1.619]^T$.

4 Quasi-Newtonove metode

Glavna tema ovoga rada su upravo Quasi-Newtonove metode koje zauzimaju vrlo važno mjesto među metodama za rješavanje sustava nelinearnih jednadžbi. Najveći nedostatak Newtonove metode je u tome što za svaku iteraciju treba računati Jacobijan te inverznu Jacobijevu matricu što znatno usporava postupak. Također, nedostatak ove metode je i taj što je vrlo važno da se izabere dobra početna aproksimacija $\boldsymbol{x}^{(0)}$. Taj problem rješavaju tzv. Quasi-Newtonove metode, koje je uveo C. G. Broyden 1965. godine. Ranije, 1959. godine, Davidon je uveo klasu Quasi-Newtonovih metoda za rješavanje problema minimizacije funkcije više varijabli bez ograničenja. Quasi-Newtonove metode nastale su kao generalizacija metode sekanti.

4.1 Broydenova metoda

Broydenova metoda je najjednostavnija i najpoznatija Quasi-Newtonova metoda. Pokazali smo kako aproksimaciju funkcije $\boldsymbol{f}$ u okolini točke $\boldsymbol{x}^{(0)}$ možemo zapisati kao

$$\tilde{\boldsymbol{f}}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}^{(0)}) + \mathbb{J}(\boldsymbol{x}^{(0)})(\boldsymbol{x} - \boldsymbol{x}^{(0)}).$$

Oznake koje koristimo bit će jednake oznakama u Newtonovoj metodi. Kako bi u svakoj iteraciji izbjegli računanje Jacobijana Broydenova metoda traži korištenje aproksimacije Jacobijana, što znači da će iterativni prostupak umjesto

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - (\mathbb{J}^{(k)})^{-1} \boldsymbol{f}(\boldsymbol{x}^{(k)}), \quad k = 0, 1, \ldots,$$

glasiti

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - (\boldsymbol{B}_k)^{-1} \boldsymbol{f}(\boldsymbol{x}^{(k)}), \quad k = 0, 1, \ldots,$$

pri čemu je $\boldsymbol{B}_k$ aproksimacija Jacobijana.

Podsjetimo se kako smo u slučaju funkcije jedne varijable kod metode sekanti primijenili Newtonovu metodu samo što smo umjesto derivacije $f'(x_k)$, koristili aproskimaciju derivacije u točki x_k:

$$f'(x_k) \approx \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} =: b_k,$$

odnosno

$$b_k(x_k - x_{k-1}) = f(x_k) - f(x_{k-1}).$$

Slično želimo i ovdje, odnosno umjesto Jacobijana funkcije $\mathbb{J}(\boldsymbol{x}^{(k)})$ želimo koristiti njegovu aproksimaciju koja zadovoljava uvjet

$$\mathbb{B}_k(\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)}) = \boldsymbol{f}(\boldsymbol{x}^{(k)}) - \boldsymbol{f}(\boldsymbol{x}^{(k-1)}),$$

odnosno uz oznake

$$\boldsymbol{s}_{k-1} = \boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)}, \quad \boldsymbol{y}_{k-1} = \boldsymbol{f}(\boldsymbol{x}^{(k)}) - \boldsymbol{f}(\boldsymbol{x}^{(k-1)}),$$

$$\boldsymbol{B}_k \boldsymbol{s}_{k-1} = \boldsymbol{y}_{k-1}. \tag{4.1}$$

Temeljna ideja Broydenove metode je iz poznate matrice $\boldsymbol{B}_{k-1}$, odrediti matricu $\boldsymbol{B}_k$, tako da ona bude najbolja moguća aproksimacija matrice $\boldsymbol{B}_{k-1}$ u smislu Froobeniusove norme te da zadovoljava uvjet (4.1), odnosno da je rješenje sljedećeg problema uvjetne optimizacije:

$$\begin{aligned} \|\boldsymbol{B}-\boldsymbol{B}_{k-1}\|_F^2 &\to \min_{\boldsymbol{B}} \\ \boldsymbol{B}\boldsymbol{s}_{k-1} &= \boldsymbol{y}_{k-1}. \end{aligned}$$

U svrhu rješavanja ovog problema definiramo Lagrangeovu funkciju

$$\mathscr{L}(\boldsymbol{B},\lambda) = \|\boldsymbol{B}-\boldsymbol{B}_{k-1}\|_F^2 + \lambda^T(\boldsymbol{B}\boldsymbol{s}_{k-1}-\boldsymbol{y}_{k-1}).$$

Deriviranjem Lagrangeove funkcije po $\boldsymbol{B}$ dobivamo

$$\frac{\partial \mathscr{L}}{\partial \boldsymbol{B}} = \boldsymbol{B}-\boldsymbol{B}_{k-1} + \frac{\partial \lambda^T \boldsymbol{B}\boldsymbol{s}_{k-1}}{\partial \boldsymbol{B}} = \boldsymbol{B}-\boldsymbol{B}_{k-1} + \lambda(\boldsymbol{s}_{k-1})^T,$$

a deriviranjem po λ

$$\frac{\partial \mathscr{L}}{\partial \lambda} = \boldsymbol{B}\boldsymbol{s}_{k-1}-\boldsymbol{y}_{k-1}.$$

Ukoliko $\frac{\partial \mathscr{L}}{\partial \boldsymbol{B}}$ izjednačimo s nulom, tj.

$$\frac{\partial \mathscr{L}}{\partial \boldsymbol{B}} = \boldsymbol{B}-\boldsymbol{B}_{k-1} + \lambda(\boldsymbol{s}_{k-1})^T = 0$$

dobivamo

$$\boldsymbol{B} = \boldsymbol{B}_{k-1} - \lambda(\boldsymbol{s}_{k-1})^T.$$

Izjednačavanjem $\frac{\partial \mathscr{L}}{\partial \lambda}$ s nulom dobivamo

$$\boldsymbol{y}_{k-1} = \boldsymbol{B}\boldsymbol{s}_{k-1}.$$

Ukoliko u prethodnom izrazu $\boldsymbol{B}$ zamijenimo s $\boldsymbol{B} = \boldsymbol{B}_{k-1} - \lambda(\boldsymbol{s}_{k-1})^T$ slijedi

$$\boldsymbol{y}_{k-1} = \left(\boldsymbol{B}_{k-1} - \lambda(\boldsymbol{s}_{k-1})^T\right)\boldsymbol{s}_{k-1},$$

odnosno

$$\boldsymbol{y}_{k-1} = \boldsymbol{B}_{k-1}\boldsymbol{s}_{k-1} - \lambda(\boldsymbol{s}_{k-1})^T\boldsymbol{s}_{k-1}.$$

Sređivanjem gornjeg izraza imamo

$$\lambda(\boldsymbol{s}_{k-1})^T\boldsymbol{s}_{k-1} = \boldsymbol{B}_{k-1}\boldsymbol{s}_{k-1} - \boldsymbol{y}_{k-1}$$

odakle slijedi

$$\lambda = (\boldsymbol{B}_{k-1}\boldsymbol{s}_{k-1} - \boldsymbol{y}_{k-1})\left((\boldsymbol{s}_{k-1})^T\boldsymbol{s}_{k-1}\right)^{-1}.$$

Uvrštavanjem λ u izraz $\boldsymbol{B} = \boldsymbol{B}_{k-1} - \lambda(\boldsymbol{s}_{k-1})^T$ dobivamo sljedeće

$$\boldsymbol{B} = \boldsymbol{B}_{k-1} - (\boldsymbol{B}_{k-1}\boldsymbol{s}_{k-1} - \boldsymbol{y}_{k-1})\left((\boldsymbol{s}_{k-1})^T \boldsymbol{s}_{k-1}\right)^{-1}(\boldsymbol{s}_{k-1})^T.$$

Sređivanjem gornjeg izraza slijedi

$$\boldsymbol{B} = \boldsymbol{B}_{k-1} - \frac{(\boldsymbol{B}_{k-1}\boldsymbol{s}_{k-1} - \boldsymbol{y}_{k-1})(\boldsymbol{s}_{k-1})^T}{(\boldsymbol{s}_{k-1})^T \boldsymbol{s}_{k-1}}$$

odnosno

$$\boldsymbol{B} = \boldsymbol{B}_{k-1} + \frac{(\boldsymbol{y}_{k-1} - \boldsymbol{B}_{k-1}\boldsymbol{s}_{k-1})(\boldsymbol{s}_{k-1})^T}{(\boldsymbol{s}_{k-1})^T \boldsymbol{s}_{k-1}}.$$

Na taj smo način dobili direktnu Broydenovu metodu:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \boldsymbol{B}_k^{-1}\boldsymbol{f}(\boldsymbol{x}^{(k)}), \quad k = 0, 1, \ldots, \tag{4.2}$$

gdje je

$$\boldsymbol{y}_k = \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \boldsymbol{f}(\boldsymbol{x}^{(k)}), \quad \boldsymbol{s}_k = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)},$$

a matrice $\boldsymbol{B}_k$ se izračunavaju iz rekurzivne formule

$$\boldsymbol{B}_{k+1} = \boldsymbol{B}_k + \frac{(\boldsymbol{y}_k - \boldsymbol{B}_k\boldsymbol{s}_k)\boldsymbol{s}_k^T}{\langle \boldsymbol{s}_k, \boldsymbol{s}_k \rangle}, \quad k = 0, 1, \ldots \tag{4.3}$$

Napomena: Uobičajeno je za aproksimaciju Jacobijana u točki $\boldsymbol{x}^{(0)}$ uzeti jediničnu matricu reda n. U jednodimenzionalnom slučaju se Broydenova metoda svodi na metodu sekante.

Kako bi se izbjeglo računanje inverzne matrice $\boldsymbol{B}_k^{-1}$, umjesto (4.2) možemo rješavati

$$\boldsymbol{B}_k\boldsymbol{s}_k = -\boldsymbol{f}(\boldsymbol{x}^{(k)}). \tag{4.4}$$

Drugi način da se izbjegne računanje $\boldsymbol{B}_k^{-1}$ je primjenom sljedećeg teorema.

Teorem 4.1. ***(Sherman-Morrison-Woodbury)*** *Neka su $\boldsymbol{u}$, $\boldsymbol{v} \in \mathbb{R}^n$, a $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ regularna matrica. Tada je $\tilde{\boldsymbol{B}} = \boldsymbol{B} + \boldsymbol{u}\boldsymbol{v}^T$ regularna onda i samo onda ako je $\sigma = 1 + \boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{u} \neq 0$. Ako je $\sigma \neq 0$, vrijedi*

$$\tilde{\boldsymbol{B}}^{-1} = (\boldsymbol{B} + \boldsymbol{u}\boldsymbol{v}^T)^{-1} = \boldsymbol{B}^{-1} - \frac{1}{\sigma}\boldsymbol{B}^{-1}\boldsymbol{u}\boldsymbol{v}^T\boldsymbol{B}^{-1}.$$

Dokaz. Neka je za bilo koji vektor $\boldsymbol{y} \in \mathbb{R}^n$ dan $\boldsymbol{x} = \tilde{\boldsymbol{B}}^{-1}\boldsymbol{y}$, odnosno $\tilde{\boldsymbol{B}}\boldsymbol{x} = \boldsymbol{y}$. Uz tako definirane vektore imamo

$$\boldsymbol{B}\boldsymbol{x} + (\boldsymbol{v}^T\boldsymbol{x})\boldsymbol{u} = \boldsymbol{y}. \tag{4.5}$$

Ukoliko gornju jednakost pomnožimo s lijeva s $\boldsymbol{v}^T\boldsymbol{B}^{-1}$ dobivamo

$$\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{x} + \boldsymbol{v}^T\boldsymbol{B}^{-1}(\boldsymbol{v}^T\boldsymbol{x})\boldsymbol{u} = \boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{y},$$

odnosno

$$\boldsymbol{v}^T\boldsymbol{x}(1 + \boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{u}) = \boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{y}.$$

Kako je prema pretpostavci $1+\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{u}\neq 0$ imamo

$$\boldsymbol{v}^T\boldsymbol{x}=\frac{\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{y}}{1+\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{u}}.$$

Sada, iz prethodnog izraza i iz izraza (4.5) dobivamo

$$\boldsymbol{B}\boldsymbol{x}=\boldsymbol{y}-(\boldsymbol{v}^T\boldsymbol{x})\boldsymbol{u}=\boldsymbol{y}-\frac{\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{y}}{1+\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{u}}\boldsymbol{u}.$$

Množenjem gornje jednakosti s lijeva s $\boldsymbol{B}^{-1}$ imamo

$$\boldsymbol{B}^{-1}\boldsymbol{B}\boldsymbol{x}=\boldsymbol{B}^{-1}\boldsymbol{y}-\frac{\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{y}}{1+\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{u}}\boldsymbol{B}^{-1}\boldsymbol{u}.$$

Korištenjem $\boldsymbol{x}=\tilde{\boldsymbol{B}}^{-1}\boldsymbol{y}$ slijedi

$$\boldsymbol{x}=\tilde{\boldsymbol{B}}^{-1}\boldsymbol{y}=\left(\boldsymbol{B}^{-1}-\frac{\boldsymbol{B}^{-1}\boldsymbol{u}\boldsymbol{v}^T\boldsymbol{B}^{-1}}{1+\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{u}}\right)\boldsymbol{y}.$$

Kako je $\sigma=1+\boldsymbol{v}^T\boldsymbol{B}^{-1}\boldsymbol{u}$ slijedi

$$\tilde{\boldsymbol{B}}^{-1}=(\boldsymbol{B}+\boldsymbol{u}\boldsymbol{v}^T)^{-1}=\boldsymbol{B}^{-1}-\frac{1}{\sigma}\boldsymbol{B}^{-1}\boldsymbol{u}\boldsymbol{v}^T\boldsymbol{B}^{-1}.$$

□

Korištenjem prethodnog teorema izračunat ćemo $\boldsymbol{B}_{k+1}^{-1}$ iz izraza (4.3). Ukoliko stavimo oznake

$$\boldsymbol{u}=\frac{\boldsymbol{y}_k-\boldsymbol{B}_k\boldsymbol{s}_k}{\langle\boldsymbol{s}_k,\boldsymbol{s}_k\rangle},\quad \boldsymbol{v}=\boldsymbol{s}_k$$

imamo

$$\sigma=1+\left\langle\boldsymbol{s}_k,\boldsymbol{B}_k^{-1}\frac{\boldsymbol{y}_k-\boldsymbol{B}_k\boldsymbol{s}_k}{\langle\boldsymbol{s}_k,\boldsymbol{s}_k\rangle}\right\rangle=\frac{\langle\boldsymbol{s}_k,\boldsymbol{B}_k^{-1}\boldsymbol{y}_k\rangle}{\langle\boldsymbol{s}_k,\boldsymbol{s}_k\rangle}.$$

Sada, korištenjem Teorema 4.2 imamo

$$\boldsymbol{B}_{k+1}^{-1}=\boldsymbol{B}_k^{-1}-\frac{\langle\boldsymbol{s}_k,\boldsymbol{s}_k\rangle}{\langle\boldsymbol{s}_k,\boldsymbol{B}_k^{-1}\boldsymbol{y}_k\rangle}\boldsymbol{B}_k^{-1}\frac{\boldsymbol{y}_k-\boldsymbol{B}_k\boldsymbol{s}_k}{\langle\boldsymbol{s}_k,\boldsymbol{s}_k\rangle}\boldsymbol{s}_k^T\boldsymbol{B}_k^{-1}=\boldsymbol{B}_k^{-1}-\frac{\boldsymbol{B}_k^{-1}(\boldsymbol{y}_k-\boldsymbol{B}_k\boldsymbol{s}_k)\boldsymbol{s}_k^T\boldsymbol{B}_k^{-1}}{\langle\boldsymbol{s}_k,\boldsymbol{B}_k^{-1}\boldsymbol{y}_k\rangle},$$

odnosno

$$\boldsymbol{B}_{k+1}^{-1}=\boldsymbol{B}_k^{-1}-\frac{(\boldsymbol{B}_k^{-1}\boldsymbol{y}_k-\boldsymbol{s}_k)\boldsymbol{s}_k^T\boldsymbol{B}_k^{-1}}{\langle\boldsymbol{s}_k,\boldsymbol{B}_k^{-1}\boldsymbol{y}_k\rangle}=\boldsymbol{B}_k^{-1}+\frac{(\boldsymbol{s}_k-\boldsymbol{B}_k^{-1}\boldsymbol{y}_k)\boldsymbol{s}_k^T\boldsymbol{B}_k^{-1}}{\langle\boldsymbol{s}_k,\boldsymbol{B}_k^{-1}\boldsymbol{y}_k\rangle}.$$

Ako označimo $\boldsymbol{H}_k:=\boldsymbol{B}_k^{-1}$, onda je $\boldsymbol{H}_{k+1}:=\boldsymbol{B}_{k+1}^{-1}$ definiran s

$$\boldsymbol{H}_{k+1}=\boldsymbol{H}_k+\frac{(\boldsymbol{s}_k-\boldsymbol{H}_k\boldsymbol{y}_k)\boldsymbol{s}_k^T\boldsymbol{H}_k}{\langle\boldsymbol{s}_k,\boldsymbol{H}_k\boldsymbol{y}_k\rangle},$$

uz uvjet $\langle\boldsymbol{s}_k,\boldsymbol{H}_k\boldsymbol{y}_k\rangle\neq 0$.

Na taj smo način definirali inverznu Broydenovu metodu:

$$\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}-\boldsymbol{H}_k\boldsymbol{f}(\boldsymbol{x}^{(k)}),\quad k=0,1,\ldots$$

Broydenova metoda sporija je od Newtonove metode.

Teorem 4.2. *(vidi[6]) Neka je $\boldsymbol{F}:\mathbb{R}^n \to \mathbb{R}^n$ neprekidno diferencijabilna na otvorenom konveksnom skupu $\mathcal{D}$, tj $\boldsymbol{F}\in C^1(\mathcal{D})$ i neka postoji $\boldsymbol{x}^\star\in\mathcal{D}$ tako da je $\boldsymbol{F}(\boldsymbol{x}^\star)=0$ i $\boldsymbol{F}'(\boldsymbol{x}^\star)$ nesingularna matrica. Nadalje, neka je $\boldsymbol{F}'$ Lipschitz-neprekidna funkcija.*

1. *Tada postoji $\delta>0$ takav da ako je $||\boldsymbol{B}_k-\boldsymbol{F}'(\boldsymbol{x}^{(k)})||\le\delta$ i $||\boldsymbol{x}^{(0)}-\boldsymbol{x}^\star||\le\delta$, niz definiran s $\boldsymbol{x}^{(k+1)}=\boldsymbol{x}^{(k)}-\boldsymbol{B}_k^{-1}\boldsymbol{F}(\boldsymbol{x}^{(k)})$ je dobro definiran i konvergira ka $\boldsymbol{x}^\star$.*

2. *Ako pri tome vrijedi da $||\boldsymbol{B}_k-\boldsymbol{F}'(\boldsymbol{x}^{(k)})||\to 0$, tada je konvergencija superlinearna.*

3. *Ako postoji konstanta C takva da $||\boldsymbol{B}_k-\boldsymbol{F}'(\boldsymbol{x}^{(k)})||\le C||\boldsymbol{F}(\boldsymbol{x}^{(k)})||$, tada je konvergencija kvadratična.*

Ukoliko je $\boldsymbol{x}^{(0)}$ dovoljno blizu nultočki $\boldsymbol{x}^\star$, gdje je $\mathbb{J}(\boldsymbol{x}^\star)$ nesingularna i gdje je $\boldsymbol{B}_0$ dovoljno blizu $\mathbb{J}(\boldsymbol{x}^{(0)})$, tada niz $\{\boldsymbol{x}^{(k)}\}$ dobiven Broydenovom metodom konvergira superlinearno prema $\boldsymbol{x}^\star$.

Primjer 4.1. *Pomoću Broydenove metode riješite sustav nelinearnih jednadžbi*

$$(x_1-1)^2+(x_2-1)^2+x_3^2=1$$

$$(x_1-1)^2+x_2^2+(x_3-1)^2=1$$

$$x_1^2+(x_2-1)^2+(x_3-1)^2=1$$

uz početnu aproksimaciju $\boldsymbol{x}^{(0)}=[0,0,0]^T$. Postupak prekinite nakon treće iteracije.

Rješenje: Stavimo

$$f_1(x_1,x_2,x_3)=(x_1-1)^2+(x_2-1)^2+x_3^2-1$$

$$f_2(x_1,x_2,x_3)=(x_1-1)^2+x_2^2+(x_3-1)^2-1$$

$$f_3(x_1,x_2,x_3)=x_1^2+(x_2-1)^2+(x_3-1)^2-1$$

Tada je

$$\boldsymbol{f}(\boldsymbol{x}^{(k)})=\begin{bmatrix} f_1(x_1^{(k)},x_2^{(k)},x_3^{(k)})\\ f_2(x_1^{(k)},x_2^{(k)},x_3^{(k)})\\ f_3(x_1^{(k)},x_2^{(k)},x_3^{(k)})\end{bmatrix},\ \text{za } k=0,1,2,...$$

Za razliku od Newtonove metode sada nećemo morati računati Jacobijan funkcije u svakoj iteraciji već ćemo ga aproksimirati matricama $\boldsymbol{B}_k$, $k=0,1,2,...$

Za $\boldsymbol{B}_0$ uzimamo jediničnu matricu reda n, odnosno reda 3. Dakle, $\boldsymbol{B}_0=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}$.

Prvo rješavamo sustav $\boldsymbol{B}_0\boldsymbol{s}_0 = -\boldsymbol{f}(\boldsymbol{x}^{(0)})$. Uvrštavanjem $\boldsymbol{B}_0$ i $\boldsymbol{f}(\boldsymbol{x}^{(0)})$ dobivamo

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1^{(0)} \\ s_2^{(0)} \\ s_3^{(0)} \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix},$$

odakle slijedi $s_1^{(0)} = -1, s_2^{(0)} = -1, s_3^{(0)} = -1$. Dakle, $\boldsymbol{s}_0 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$.

Sada računamo $\boldsymbol{x}^{(1)}$. Imamo

$$\boldsymbol{x}^{(1)} = \boldsymbol{s}_0 + \boldsymbol{x}^{(0)} = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}.$$

Da bismo mogli izračunati aproksimaciju $\boldsymbol{B}_1$ prvo moramo izračunati $\boldsymbol{y}_0$:

$$\boldsymbol{y}_0 = \boldsymbol{f}(\boldsymbol{x}^{(1)}) - \boldsymbol{f}(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 8 \\ 8 \\ 8 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix}.$$

Sada je

$$\boldsymbol{B}_1 = \boldsymbol{B}_0 + \frac{(\boldsymbol{y}_0 - \boldsymbol{B}_0\boldsymbol{s}_0)\boldsymbol{s}_0^T}{\langle \boldsymbol{s}_0, \boldsymbol{s}_0 \rangle} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{\left(\begin{bmatrix} 7 \\ 7 \\ 7 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} \right) \begin{bmatrix} -1 & -1 & -1 \end{bmatrix}}{3},$$

odnosno

$$\boldsymbol{B}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{\begin{bmatrix} -8 & -8 & -8 \\ -8 & -8 & -8 \\ -8 & -8 & -8 \end{bmatrix}}{3} = \begin{bmatrix} -1.66667 & -2.66667 & -2.66667 \\ -2.66667 & -1.66667 & -2.66667 \\ -2.66667 & -2.66667 & -1.66667 \end{bmatrix}.$$

Kako bismo mogli izračunati $\boldsymbol{x}^{(2)}$ ponovno rješavamo sustav $\boldsymbol{B}_k\boldsymbol{s}_k = -\boldsymbol{f}(\boldsymbol{x}^{(k)})$, ali ovaj puta za $k = 1$. Imamo

$$\boldsymbol{B}_1\boldsymbol{s}_1 = -\boldsymbol{f}(\boldsymbol{x}^{(1)}),$$

odnosno

$$\begin{bmatrix} -1.66667 & -2.66667 & -2.66667 \\ -2.66667 & -1.66667 & -2.66667 \\ -2.66667 & -2.66667 & -1.66667 \end{bmatrix} \begin{bmatrix} s_1^{(1)} \\ s_2^{(1)} \\ s_3^{(1)} \end{bmatrix} = \begin{bmatrix} -8 \\ -8 \\ -8 \end{bmatrix}.$$

Dakle, dobili smo sustav jednadžbi

$$-1.66667 s_1^{(1)} - 2.66667 s_2^{(1)} - 2.66667 s_3^{(1)} = -8$$

$$-2.66667s_1^{(1)} - 1.66667s_2^{(1)} - 2.66667s_3^{(1)} = -8$$

$$-2.66667s_1^{(1)} - 2.66667s_2^{(1)} - 1.66667s_3^{(1)} = -8.$$

Rješavanjem sustava dobivamo da je $s_1^{(1)} = 1.14286$, $s_2^{(1)} = 1.14286$, $s_3^{(1)} = 1.14286$, odnosno $\boldsymbol{s}_1 = \begin{bmatrix} 1.14286 \\ 1.14286 \\ 1.14286 \end{bmatrix}$.

Aproksimacija $\boldsymbol{x}^{(2)}$ sada iznosi

$$\boldsymbol{x}^{(2)} = \boldsymbol{s}_1 + \boldsymbol{x}^{(1)} = \begin{bmatrix} 1.14286 \\ 1.14286 \\ 1.14286 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0.14286 \\ 0.14286 \\ 0.14286 \end{bmatrix}.$$

Računamo

$$\boldsymbol{y}_1 = \boldsymbol{f}(\boldsymbol{x}^{(2)}) - \boldsymbol{f}(\boldsymbol{x}^{(1)}) = \begin{bmatrix} -7.51021 \\ -7.51021 \\ -7.51021 \end{bmatrix}.$$

Sada imamo sve što nam je potrebno kako bi mogli izračunati $\boldsymbol{B}_2$:

$$\boldsymbol{B}_2 = \boldsymbol{B}_1 + \frac{(\boldsymbol{y}_1 - \boldsymbol{B}_1\boldsymbol{s}_1)\boldsymbol{s}_1^T}{\langle \boldsymbol{s}_1, \boldsymbol{s}_1 \rangle} = \begin{bmatrix} -1.66667 & -2.66667 & -2.66667 \\ -2.66667 & -1.66667 & -2.66667 \\ -2.66667 & -2.66667 & -1.66667 \end{bmatrix} + \frac{\begin{bmatrix} 0.560 & 0.560 & 0.560 \\ 0.560 & 0.560 & 0.560 \\ 0.560 & 0.560 & 0.560 \end{bmatrix}}{3.91839},$$

odakle slijedi

$$\boldsymbol{B}_2 = \begin{bmatrix} -1.52375 & -2.52375 & -2.52375 \\ -2.52375 & -1.52375 & -2.52375 \\ -2.52375 & -2.52375 & -1.52375 \end{bmatrix}.$$

Preostalo nam je još izračunati $\boldsymbol{s}_2$ kako bismo mogli dobiti $\boldsymbol{x}^{(3)}$. Rješavamo

$$\boldsymbol{B}_2\boldsymbol{s}_2 = -\boldsymbol{f}(\boldsymbol{x}^{(2)}),$$

odnosno

$$\begin{bmatrix} -1.52375 & -2.52375 & -2.52375 \\ -2.52375 & -1.52375 & -2.52375 \\ -2.52375 & -2.52375 & -1.52375 \end{bmatrix} \begin{bmatrix} s_1^{(2)} \\ s_2^{(2)} \\ s_3^{(2)} \end{bmatrix} = \begin{bmatrix} -0.48979 \\ -0.48979 \\ -0.48979 \end{bmatrix}.$$

Rješenje gornjeg sustava je $\boldsymbol{s}_2 = \begin{bmatrix} 0.07454 \\ 0.07454 \\ 0.07454 \end{bmatrix}$ pa je traženi $\boldsymbol{x}^{(3)} = \boldsymbol{s}_2 + \boldsymbol{x}^{(2)} = \begin{bmatrix} 0.21739 \\ 0.21739 \\ 0.21739 \end{bmatrix}$.

4.2 Davidon-Fletcher-Powellova metoda (DFP metoda)

Prije nego krenemo analizirati Davidon-Fletcher-Powellovu (DFP) metodu te Broyden-Fletcher-Goldfarb-Schannovu (BFGS) metodu prisjetimo se pojmova i svojstava koji će nam biti potrebni u nastavku.

Definicija 4.1. *Neka je* $\boldsymbol{f} = [f_1, f_2, ..., f_n] : \mathbb{R}^n \to \mathbb{R}^n$ *funkcija koja ima neprekidnu drugu derivaciju na* $\mathbb{R}^n$ *i neka je* $\boldsymbol{x} = [x_1, x_2, ..., x_n] \in \mathbb{R}^n$. *Tada matricu*

$$\boldsymbol{H} = \begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1^2} & \frac{\partial^2 f_1}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f_1}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f_2}{\partial x_2 \partial x_1} & \frac{\partial^2 f_2}{\partial x_2^2} & \cdots & \frac{\partial^2 f_2}{\partial x_2 \partial x_n} \\ \vdots & \cdots & \vdots & \\ \frac{\partial^2 f_n}{\partial x_n \partial x_1} & \frac{\partial^2 f_n}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f_n}{\partial x_n^2} \end{bmatrix}$$

nazivamo Hessijan ili Hesseova matrica.

Definicija 4.2. *Kažemo da funkcija* $\boldsymbol{f} : \mathscr{D} \to \mathbb{R}, \mathscr{D} \subseteq \mathbb{R}^n$ *u točki* $\boldsymbol{x}^\star \in \mathscr{D}$ *postiže lokalni minimum ako postoji okolina* $\mathscr{O}(\boldsymbol{x}^\star)$ *takva da je* $\boldsymbol{f}(\boldsymbol{x}) \geqslant \boldsymbol{f}(\boldsymbol{x}^\star)$ *za sve* $\boldsymbol{x} \in \mathscr{O}(\boldsymbol{x}^\star) \cap \mathscr{D}$. *Točku* $\boldsymbol{x}^\star$ *zovemo točka lokalnog minimuma ili lokalni minimizator funkcije* $\boldsymbol{f}$. *Kažemo da je* $\boldsymbol{x}^\star$ *točka strogog lokalnog minimuma ili strogi lokalni minimizator funkcije* $\boldsymbol{f}$ *ako postoji okolina* $\mathscr{O}(\boldsymbol{x}^\star)$ *takva da je* $\boldsymbol{f}(\boldsymbol{x}) > \boldsymbol{f}(\boldsymbol{x}^\star)$ *za sve* $\boldsymbol{x} \in \mathscr{O}(\boldsymbol{x}^\star) \cap \mathscr{D} \setminus \{\boldsymbol{x}^\star\}$.

Teorem 4.3. *Neka je* $\boldsymbol{f} \in C^1(\mathscr{D})$ *neprekidno diferencijabilna funkcija definirana na otvorenom skupu* $\mathscr{D} \subset \mathbb{R}^n$ *i neka je* $\boldsymbol{x}^\star$ *točka lokalnog minimuma od* $\boldsymbol{f}$. *Tada vrijedi*
(i) $\boldsymbol{x}^\star$ *je stacionarna točka funkcije* $\boldsymbol{f}$, *tj.* $\nabla \boldsymbol{f}(\boldsymbol{x}^\star) = 0$,
(ii) ako je $\boldsymbol{f} \in C^2(\mathscr{D})$, *njezin Hessijan* $\nabla^2 \boldsymbol{f}(\boldsymbol{x}^\star)$ *u točki* $\boldsymbol{x}^\star$ *je pozitivno semidefinitna matrica.*

Za razliku od Broydenove metode koja služi za rješavanje sustava nelinearnih jednadžbi, DFP i BFGS se koriste za optimizaciju funkcije. Quasi-Newtonove optimizacijske metode kompliciranije su od onih za rješavanje sustava nelinearnih jednadžbi jer metode za rješavanje sustava nelinearnih jednadžbi ne čuvaju simetričnost i pozitivnu definitnost koji su potrebni za metode linijskim pretraživanjem u optimizaciji što bi moglo potaknuti konvergenciju prema lokalnom maksimumu. Kod metode linijskim pretraživanjem za danu aproksimaciju $\boldsymbol{x}^{(k)} \in \mathbb{R}^n$, aproksimaciju $\boldsymbol{x}^{(k+1)}$ tražimo u obliku $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{p}_k$, $k = 0, 1, 2, ...$, gdje je $\boldsymbol{x}^{(0)}$ početna aproksimacija, $\boldsymbol{p}_k \in \mathbb{R}^n$ vektor smjera, a $\lambda_k > 0$ duljina koraka.
Za diferencijabilnu funkciju $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}$ problem optimizacije možemo promatrati kao problem rješavanja jednadžbe

$$\boldsymbol{F}(\boldsymbol{x}) = 0$$

pri čemu je $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ takva da je $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{f}'(\boldsymbol{x})$. Dakle, problem možemo riješiti nekom od metoda za rješavanje sustava nelinearnih jednadžbi, ali i pomoću iterativnih metoda za rješavanje optimizacijskog problema. Jedna takva metoda je i Davidon-Fletcher-Powellova metoda. DFP metodom dobivamo aproksimaciju Hessijana iterativnim postupkom. Promatramo metode za koje je smjer silaska definiran s $\boldsymbol{q}_k = -\boldsymbol{H}_k \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})$, gdje je $\boldsymbol{H}_k$ pozitivno definitna matrica. Također, interesantne su nam one metode koje matricu $\boldsymbol{H}_{k+1}$ relativno jednostavno računaju iz matrice $\boldsymbol{H}_k$. U tom slučaju ne moramo u svakoj iteraciji rješavati sustav $\nabla^2 \boldsymbol{f}(\boldsymbol{x}^{(k)})\boldsymbol{q}_k = -\nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})$ već u svakoj iteraciji množimo matricu $\boldsymbol{H}_k$ s vektorom $\nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})$.

Matricu $\boldsymbol{H}_{k+1}$ ćemo u svakom koraku generirati tako da vrijedi

$$\boldsymbol{H}_{k+1}\Big(\nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})\Big) = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}.$$

Označimo

$$\boldsymbol{y}_k = \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \qquad \boldsymbol{s}_k = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}.$$

Želimo pronaći pozitivno definitnu matricu $\boldsymbol{H}_{k+1}$ koja zadovoljava

$$\boldsymbol{H}_{k+1}\boldsymbol{y}_k = \boldsymbol{s}_k. \tag{4.6}$$

Kao što smo već ranije rekli, želimo da se matrica $\boldsymbol{H}_{k+1}$ lako računa iz matrice $\boldsymbol{H}_k$. To bi značilo da $\boldsymbol{H}_{k+1}$ dobijemo tako da matrici $\boldsymbol{H}_k$ pribrojimo neku jednostavnu matricu $\boldsymbol{E}$. Tada imamo

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \boldsymbol{E}.$$

Ako matricu $\boldsymbol{H}_{k+1}$ definiranu kao u prethodnom izrazu uvrstimo u (4.6) dobivamo

$$(\boldsymbol{H}_k + \boldsymbol{E})\boldsymbol{y}_k = \boldsymbol{s}_k.$$

Na ovaj način smo problem sveli na određivanje matrice $\boldsymbol{E}$ koja zadovoljava

$$\boldsymbol{E}\boldsymbol{y}_k = \boldsymbol{s}_k - \boldsymbol{H}_k\boldsymbol{y}_k.$$

Matrica $\boldsymbol{E}$ je simetrična jer su simetrične i matrice $\boldsymbol{H}_k$ i $\boldsymbol{H}_{k+1}$. Najjednostavnija simetrična matrica $\boldsymbol{M}$ koja zadovoljava

$$\boldsymbol{M}\boldsymbol{x} = \boldsymbol{y}$$

ima oblik

$$\boldsymbol{M} = \frac{\boldsymbol{y}\boldsymbol{y}^T}{\boldsymbol{y}^T\boldsymbol{x}} = \frac{\boldsymbol{y}\boldsymbol{y}^T}{\langle \boldsymbol{y}, \boldsymbol{x}\rangle},$$

gdje su $\boldsymbol{x}$ i $\boldsymbol{y}$ zadani vektori. Pokažimo da ovako definirana matrica $\boldsymbol{M}$ zadovoljava $\boldsymbol{M}\boldsymbol{x} = \boldsymbol{y}$.

U $\boldsymbol{M}\boldsymbol{x} = \boldsymbol{y}$ uvrstimo $\boldsymbol{M} = \frac{\boldsymbol{y}\boldsymbol{y}^T}{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}$ pa imamo

$$\boldsymbol{M}\boldsymbol{x} = \frac{\boldsymbol{y}\boldsymbol{y}^T}{\langle \boldsymbol{y}, \boldsymbol{x}\rangle}\boldsymbol{x} = \frac{\boldsymbol{y}(\boldsymbol{y}^T\boldsymbol{x})}{\langle \boldsymbol{y}, \boldsymbol{x}\rangle} = \frac{\boldsymbol{y}\langle \boldsymbol{y}, \boldsymbol{x}\rangle}{\langle \boldsymbol{y}, \boldsymbol{x}\rangle} = \boldsymbol{y}.$$

Tada za matricu $\boldsymbol{F}$ definiranu kao

$$\boldsymbol{F} = \frac{\boldsymbol{s}_k\boldsymbol{s}_k^T}{\boldsymbol{s}_k^T\boldsymbol{y}_k} = \frac{\boldsymbol{s}_k\boldsymbol{s}_k^T}{\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle}$$

vrijedi $\boldsymbol{F}\boldsymbol{y}_k = \boldsymbol{s}_k$, a za matricu

$$\boldsymbol{G} = \frac{\boldsymbol{H}_k\boldsymbol{y}_k(\boldsymbol{H}_k\boldsymbol{y}_k)^T}{(\boldsymbol{H}_k\boldsymbol{y}_k)^T\boldsymbol{y}_k} = \frac{\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^T\boldsymbol{H}_k}{\boldsymbol{y}_k^T\boldsymbol{H}_k\boldsymbol{y}_k} = \frac{\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^T\boldsymbol{H}_k}{\langle \boldsymbol{y}_k, \boldsymbol{H}_k\boldsymbol{y}_k\rangle}$$

vrijedi

$$\boldsymbol{G}\boldsymbol{y}_k = \frac{\boldsymbol{H}_k\boldsymbol{y}_k(\boldsymbol{H}_k\boldsymbol{y}_k)^T}{(\boldsymbol{H}_k\boldsymbol{y}_k)^T\boldsymbol{y}_k}\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{y}_k.$$

Kako je $\boldsymbol{F}\boldsymbol{y}_k = \boldsymbol{s}_k$ i $\boldsymbol{G}\boldsymbol{y}_k = \boldsymbol{H}_k\boldsymbol{y}_k$ onda umjesto

$$\boldsymbol{E}\boldsymbol{y}_k = \boldsymbol{s}_k - \boldsymbol{H}_k\boldsymbol{y}_k$$

imamo

$$\boldsymbol{E}\boldsymbol{y}_k = \boldsymbol{F}\boldsymbol{y}_k - \boldsymbol{G}\boldsymbol{y}_k,$$

odnosno

$$\boldsymbol{E} = \boldsymbol{F} - \boldsymbol{G}.$$

Dakle, matrica $\boldsymbol{E}$ je definirana pomoću $\boldsymbol{F}$ i $\boldsymbol{G}$ pa imamo

$$\boldsymbol{E} = \boldsymbol{F} - \boldsymbol{G} = \frac{\boldsymbol{s}_k\boldsymbol{s}_k^T}{\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle} - \frac{\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^T\boldsymbol{H}_k}{\langle \boldsymbol{y}_k, \boldsymbol{H}_k\boldsymbol{y}_k\rangle}.$$

Konačno, matrica $\boldsymbol{H}_{k+1}$ definirana je s

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{s}_k\boldsymbol{s}_k^T}{\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle} - \frac{\boldsymbol{H}_k\boldsymbol{y}_k\boldsymbol{y}_k^T\boldsymbol{H}_k}{\langle \boldsymbol{y}_k, \boldsymbol{H}_k\boldsymbol{y}_k\rangle}, \tag{4.7}$$

gdje je

$$\boldsymbol{y}_k = \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \qquad \boldsymbol{s}_k = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}.$$

Ovako definiran $\boldsymbol{H}_{k+1}$ zadovoljava

$$\boldsymbol{H}_{k+1}\Big(\nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})\Big) = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}.$$

Ova metoda se naziva Davidon-Fletcher-Powellova metoda. (vidi [7])

Pokažimo da je $\boldsymbol{H}_{k+1}$ dobro definiran, odnosno da su nazivnici $\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle$ i $\langle \boldsymbol{y}_k, \boldsymbol{H}_k\boldsymbol{y}_k\rangle$ u izrazu (4.7) različiti od nule. Ukoliko je $\nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}) \neq 0$ nazivnici u izrazu (4.7) biti će strogo veći od 0.

Ako imamo egzaktan izbor koraka $\boldsymbol{f}(\boldsymbol{x}^{(k)} + \lambda_k\boldsymbol{p}_k)) = \min_{\lambda \geqslant 0} \boldsymbol{f}(\boldsymbol{x}^{(k)} + \lambda\boldsymbol{p}_k))$ vektori $\boldsymbol{p}_k$ i $\nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)})$ biti će okomiti. Pokažimo to.

Obzirom da je λ_k točka minimuma, imamo

$$0 = \frac{d}{d\lambda}\boldsymbol{f}(\boldsymbol{x}^{(k)} + \lambda\boldsymbol{p}_k)\Big|_{\lambda=\lambda_k} = \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)} + \lambda_k\boldsymbol{p}_k), \boldsymbol{p}_k\rangle = \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}), \boldsymbol{p}_k\rangle.$$

Gornja jednakost neće vrijediti u slučaju da korak biramo nekom numeričkom metodom, a ne egzaktno. Tada je jednoznačno određen $\mu_k \neq 0$ i vrijedi

$$\langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}), \boldsymbol{p}_k \rangle = \mu_k \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \boldsymbol{p}_k \rangle.$$

Ukoliko duljinu koraka ne biramo egzaktno vrijedit će $\mu_k \approx 0$, a ukoliko je izbor koraka egzaktan onda je $\mu_k = 0$ te dobivamo tvrdnju koju smo već dokazali.

Sada, za $\boldsymbol{y}_k = \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})$, $\boldsymbol{s}_k = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$ i $\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} = \lambda_k \boldsymbol{p}_k$ imamo

$$\boldsymbol{s}_k^T \boldsymbol{y}_k = \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)} \rangle = \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \lambda_k \boldsymbol{p}_k \rangle.$$

Sređivanjem gornjeg izraza i korištenjem dokazanih tvrdnji dobivamo

$$\boldsymbol{s}_k^T \boldsymbol{y}_k = \lambda_k [\langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}), \boldsymbol{p}_k \rangle - \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \boldsymbol{p}_k \rangle] = \lambda_k [\mu_k \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \boldsymbol{p}_k \rangle - \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \boldsymbol{p}_k \rangle],$$

odnosno

$$\boldsymbol{s}_k^T \boldsymbol{y}_k = \lambda_k (\mu_k - 1) \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \boldsymbol{p}_k \rangle = -\lambda_k (\mu_k - 1) \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \boldsymbol{H}_k \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}) \rangle.$$

Obzirom da je $\nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}) \neq 0$ te da je $\boldsymbol{H}_k$ pozitivno definitna matrica onda je gornji skalarni produkt strogo pozitivan. Ukoliko je pri tome i $\mu_k < 1$ dobivamo tvrdnju koju smo željeli pokazati, tj. $\boldsymbol{s}_k^T \boldsymbol{y}_k > 0$.

Preostaje nam još pokazati da je $\boldsymbol{y}_k^T \boldsymbol{H}_k \boldsymbol{y}_k > 0$. Obzirom da je matrica $\boldsymbol{H}_k$ pozitivno definitna, vrijedit će

$$\boldsymbol{y}_k^T \boldsymbol{H}_k \boldsymbol{y}_k \geq 0.$$

Izraz $\boldsymbol{y}_k^T \boldsymbol{H}_k \boldsymbol{y}_k$ će biti jednak nuli samo ako je $\nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) = \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})$, odnosno $\boldsymbol{y}_k = 0$. U tom slučaju je

$$\langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}), \boldsymbol{p}_k \rangle = \langle \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}), \boldsymbol{p}_k \rangle,$$

odnosno $\mu_k = 1$.

Kao i u prethodnom slučaju, ako je $\mu_k < 1$ vrijedit će

$$\boldsymbol{y}_k^T \boldsymbol{H}_k \boldsymbol{y}_k > 0.$$

Dokazivanjem ove dvije tvrdnje pokazali smo da je matrica $\boldsymbol{H}_{k+1}$ dobro definirana. Znamo da je matrica $\boldsymbol{H}_k$ pozitivno definitna pa je onda i simetrična. Ono što znamo iz definicije matrice $\boldsymbol{H}_{k+1}$ je da je ta matrica simetrična, ali ne možemo ništa reći o tome je li pozitivno definitna ili ne. Neka je $\boldsymbol{q} \in \mathbb{R}^n$ proizvoljan, takav da je $\boldsymbol{q} \neq \boldsymbol{0}$. Obzirom da je $\boldsymbol{H}_k$ pozitivno definitna matrica, možemo je zapisati kao

$$\boldsymbol{H}_k = \boldsymbol{L}\boldsymbol{L}^T,$$

pri čemu je $\boldsymbol{L}$ regularna donjetrokutasta matrica. Uvođenjem oznaka $\boldsymbol{u} = \boldsymbol{L}^T\boldsymbol{q}$ i $\boldsymbol{v} = \boldsymbol{L}^T\boldsymbol{y}_k$ dobivamo

$$\langle \boldsymbol{H}_{k+1}\boldsymbol{q}, \boldsymbol{q}\rangle = \langle \boldsymbol{u}, \boldsymbol{u}\rangle + \frac{\langle \boldsymbol{s}_k, \boldsymbol{q}\rangle^2}{\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle} - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle^2}{\langle \boldsymbol{v}, \boldsymbol{v}\rangle}.$$

Znamo da je $\frac{\langle \boldsymbol{s}_k, \boldsymbol{q}\rangle^2}{\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle} \geq 0$ jer je $\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle > 0$. Također, kao posljedica Schwartzove nejednakosti vrijedi i

$$\langle \boldsymbol{u}, \boldsymbol{u}\rangle - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle^2}{\langle \boldsymbol{v}, \boldsymbol{v}\rangle} = ||\boldsymbol{u}||^2 - \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle^2}{||\boldsymbol{v}||^2} = \frac{||u||^2||\boldsymbol{v}||^2 - \langle \boldsymbol{u}, \boldsymbol{v}\rangle^2}{||\boldsymbol{v}||^2} \geq 0.$$

Iz zadnja dva izraza slijedi da je $\langle \boldsymbol{H}_{k+1}\boldsymbol{q}, \boldsymbol{q}\rangle \geq 0$. Pokažimo da je $\langle \boldsymbol{H}_{k+1}\boldsymbol{q}, \boldsymbol{q}\rangle \neq 0$. Ukoliko je $\langle \boldsymbol{H}_{k+1}\boldsymbol{q}, \boldsymbol{q}\rangle = 0$, onda zbog ranije dokazanih tvrdnji mora vrijediti

$$||u||^2||\boldsymbol{v}||^2 - \langle \boldsymbol{u}, \boldsymbol{v}\rangle^2 = 0.$$

Gornji izraz će vrijediti samo ako su $\boldsymbol{u}$ i $\boldsymbol{v}$ kolinearni, tj. $\boldsymbol{u} = \alpha\boldsymbol{v}$. U tom slučaju je i $\boldsymbol{q} = \alpha\boldsymbol{y}_k$ pri čemu je $\alpha \neq 0$ jer je $\boldsymbol{q} \neq \boldsymbol{0}$. Također, ako je $\langle \boldsymbol{H}_{k+1}\boldsymbol{q}, \boldsymbol{q}\rangle = 0$ onda mora vrijediti i sljedeći izraz

$$0 = \frac{(\boldsymbol{s}_k^T\boldsymbol{q})^2}{\boldsymbol{s}_k^T\boldsymbol{y}_k} = \frac{\langle \boldsymbol{s}_k, \alpha\boldsymbol{y}_k\rangle^2}{\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle} = \alpha^2\langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle.$$

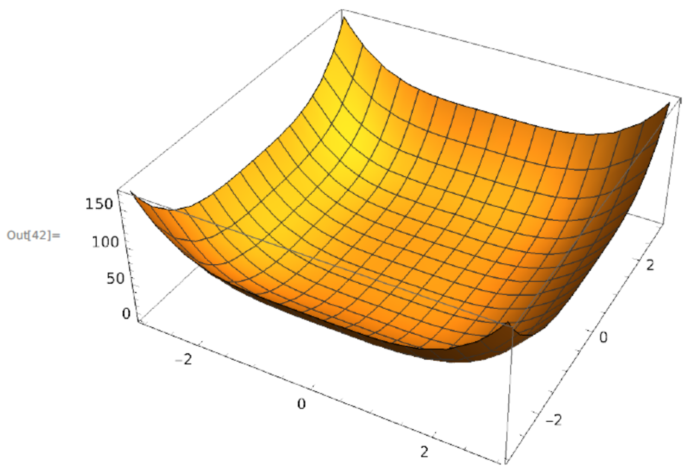
Došli smo do kontradikcije s $0 < \boldsymbol{s}_k^T\boldsymbol{y}_k = \langle \boldsymbol{s}_k, \boldsymbol{y}_k\rangle$ čime smo dokazali da je $\langle \boldsymbol{H}_{k+1}\boldsymbol{q}, \boldsymbol{q}\rangle > 0$.

Primjer 4.2. *Pomoću Davidon-Fletcher-Powellove metode odredite prve četiri aproksimacije minimuma funkcije*

$$\boldsymbol{f}(x_1, x_2) = x_1^4 + x_2^4 + x_2^2$$

uz početnu aproksimaciju $\boldsymbol{x}^{(0)} = [0.5, -0.5]^T$.

Rješenje:



Slika 4.1 Grafički prikaz funkcije $\boldsymbol{f}(x_1, x_2) = x_1^4 + x_2^4 + x_2^2$

Kao i u Broydenovoj metodi, za $\boldsymbol{H}_0$ uzimamo jediničnu matricu čije dimenzije odgovaraju broju varijabli u funkciji $\boldsymbol{f}$, tj. jediničnu matricu reda 2. Dakle,

$$\boldsymbol{H}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Odredimo prvo gradijent funkcije $\boldsymbol{f}$ u nekoj točki $\boldsymbol{x}^{(k)}$:

$$\nabla \boldsymbol{f}(x_1^{(k)}, x_2^{(k)}) = \begin{bmatrix} \frac{\partial \boldsymbol{f}(x_1^{(k)}, x_2^{(k)})}{\partial x_1^{(k)}} \\ \frac{\partial \boldsymbol{f}(x_1^{(k)}, x_2^{(k)})}{\partial x_2^{(k)}} \end{bmatrix} = \begin{bmatrix} 4x_1^3 \\ 4x_2^3 + 2x_2 \end{bmatrix}.$$

Gradijent u točki $\boldsymbol{x}^{(0)}$ iznosi

$$\nabla \boldsymbol{f}(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 0.5 \\ -1.5 \end{bmatrix}.$$

Aproksimaciju minimuma u svakoj sljedećoj iteraciji računamo kao

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \boldsymbol{H}_k \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})$$

pa je onda, za $k = 0$

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \boldsymbol{H}_0 \nabla \boldsymbol{f}(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5 \\ -1.5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Sada možemo izračunati $\boldsymbol{s}_0$. Općenito,

$$\boldsymbol{s}_k = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}.$$

Dakle, $\boldsymbol{s}_0 = \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(0)}$. Uvrštavanjem dobivamo $\boldsymbol{s}_0 = \begin{bmatrix} -0.5 \\ 1.5 \end{bmatrix}$. Kako bismo mogli izračunati aproksimaciju Hessijana u točki $\boldsymbol{x}^{(1)}$ potrebno je još odrediti $\boldsymbol{y}_0$. Općenito je

$$\boldsymbol{y}_k = \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}).$$

Za $k = 0$ uvrštavanjem dobivamo

$$\boldsymbol{y}_0 = \nabla \boldsymbol{f}(\boldsymbol{x}^{(1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 0 \\ 6 \end{bmatrix} - \begin{bmatrix} 0.5 \\ -1.5 \end{bmatrix} = \begin{bmatrix} -0.5 \\ 7.5 \end{bmatrix}.$$

Sada imamo izračunato sve što nam je potrebno da bismo mogli odrediti $\boldsymbol{H}_1$. Aproksimaciju Hessijana računamo po formuli

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^T}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle} - \frac{\boldsymbol{H}_k \boldsymbol{y}_k \boldsymbol{y}_k^T \boldsymbol{H}_k}{\langle \boldsymbol{y}_k, \boldsymbol{H}_k \boldsymbol{y}_k \rangle}.$$

Dakle, za $k = 0$ imamo

$$\boldsymbol{H}_1 = \boldsymbol{H}_0 + \frac{\boldsymbol{s}_0 \boldsymbol{s}_0^T}{\langle \boldsymbol{y}_0, \boldsymbol{s}_0 \rangle} - \frac{\boldsymbol{H}_0 \boldsymbol{y}_0 \boldsymbol{y}_0^T \boldsymbol{H}_0}{\langle \boldsymbol{y}_0, \boldsymbol{H}_0 \boldsymbol{y}_0 \rangle} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{\begin{bmatrix} 0.25 & -0.75 \\ -0.75 & 0.25 \end{bmatrix}}{11.5} - \frac{\begin{bmatrix} 0.25 & -3.75 \\ -3.75 & 56.25 \end{bmatrix}}{56.5},$$

odnosno

$$\boldsymbol{H}_1 = \begin{bmatrix} 1.01732 & 0.00115 \\ 0.00115 & 0.02174 \end{bmatrix}.$$

Sada možemo izračunati aproksimaciju minimuma $\boldsymbol{x}^{(2)}$. Slijedi,

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} - \boldsymbol{H}_1 \nabla \boldsymbol{f}(\boldsymbol{x}^{(1)}) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 1.01732 & 0.00115 \\ 0.00115 & 0.02174 \end{bmatrix} \begin{bmatrix} 0 \\ 6 \end{bmatrix} = \begin{bmatrix} -0.0069 \\ 0.86956 \end{bmatrix}.$$

Uvrštavanjem $\boldsymbol{x}^{(1)}$ i $\boldsymbol{x}^{(2)}$ u $\boldsymbol{s}_1 = \boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)}$ dobivamo $\boldsymbol{s}_1 = \begin{bmatrix} -0.0069 \\ -0.13044 \end{bmatrix}$, a uvrštavanjem $\nabla \boldsymbol{f}(\boldsymbol{x}^{(1)})$ i $\nabla \boldsymbol{f}(\boldsymbol{x}^{(2)})$ u $\boldsymbol{y}_1 = \nabla \boldsymbol{f}(\boldsymbol{x}^{(2)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(1)})$ dobivamo $\boldsymbol{y}_1 = \begin{bmatrix} -0.000001314 \\ -1.63086 \end{bmatrix}$. Aproksimacija $\boldsymbol{H}_2$ računa se po formuli

$$\boldsymbol{H}_2 = \boldsymbol{H}_1 + \frac{\boldsymbol{s}_1 \boldsymbol{s}_1^T}{\langle \boldsymbol{y}_1, \boldsymbol{s}_1 \rangle} - \frac{\boldsymbol{H}_1 \boldsymbol{y}_1 \boldsymbol{y}_1^T \boldsymbol{H}_1}{\langle \boldsymbol{y}_1, \boldsymbol{H}_1 \boldsymbol{y}_1 \rangle}$$

pa uvrštavanjem imamo

$$\boldsymbol{H}_2 = \begin{bmatrix} 1.01732 & 0.00115 \\ 0.00115 & 0.02174 \end{bmatrix} + \frac{\begin{bmatrix} 0.00004761 & 0.00090004 \\ 0.00090004 & 0.017015 \end{bmatrix}}{0.21273} - \frac{\begin{bmatrix} 0.000003522 & 0.000066542 \\ 0.000066543 & 0.0012571 \end{bmatrix}}{0.05782},$$

odnosno

$$\boldsymbol{H}_2 = \begin{bmatrix} 1.01704 & -0.00423 \\ -0.00423 & -0.07998 \end{bmatrix}.$$

Ponavljamo postupak pa računamo $\boldsymbol{x}^{(3)}$. Aproksimacija $\boldsymbol{x}^{(3)}$ iznosi

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} - \boldsymbol{H}_2 \nabla \boldsymbol{f}(\boldsymbol{x}^{(2)}) = \begin{bmatrix} 0.011583 \\ 1.219 \end{bmatrix}.$$

Sada računamo $\boldsymbol{s}_2$ i $\boldsymbol{y}_2$ pa imamo

$$\boldsymbol{s}_1 = \boldsymbol{x}^{(3)} - \boldsymbol{x}^{(2)} = \begin{bmatrix} 0.011583 \\ 1.219 \end{bmatrix} - \begin{bmatrix} -0.0069 \\ 0.86956 \end{bmatrix} = \begin{bmatrix} 0.018483 \\ 0.34944 \end{bmatrix},$$

$$\boldsymbol{y}_2 = \nabla \boldsymbol{f}(\boldsymbol{x}^{(3)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(2)}) = \begin{bmatrix} 0.000006216 \\ 9.68355 \end{bmatrix} - \begin{bmatrix} -0.000001314 \\ 4.36914 \end{bmatrix} = \begin{bmatrix} 0.00000753 \\ 5.31441 \end{bmatrix}.$$

Računamo još $\boldsymbol{H}_3$. Slijedi,

$$\boldsymbol{H}_3 = \boldsymbol{H}_2 + \frac{\boldsymbol{s}_2 \boldsymbol{s}_2^T}{\langle \boldsymbol{y}_2, \boldsymbol{s}_2 \rangle} - \frac{\boldsymbol{H}_2 \boldsymbol{y}_2 \boldsymbol{y}_2^T \boldsymbol{H}_2}{\langle \boldsymbol{y}_2, \boldsymbol{H}_2 \boldsymbol{y}_2 \rangle} = \begin{bmatrix} 1.01708 & -0.00348 \\ -0.00348 & -0.06575 \end{bmatrix}.$$

Konačno, tražena aproksimacija minimuma za $k = 3$ iznosi

$$\boldsymbol{x}^{(4)} = \boldsymbol{x}^{(3)} - \boldsymbol{H}_3 \nabla \boldsymbol{f}(\boldsymbol{x}^{(3)}) = \begin{bmatrix} 0.04528 \\ 1.85569 \end{bmatrix}.$$

4.3 Broyden-Fletcher-Goldfarb-Schannova metoda (BFGS metoda)

Za razliku od prethodne metode u kojoj smo dobili rekurzivnu formulu za matricu koja sve više nalikuje Hessijanu minimizirajuće funkcije sada ćemo tražiti rekurzivnu formulu za aproksimaciju matrice koja će nalikovati na inverzni Hessijan minimizirajuće funkcije.
Pretpostavimo da imamo $\boldsymbol{H}_k^{-1} = [\nabla^2 \boldsymbol{f}(\boldsymbol{x}^{(k)})]^{-1}$ što je aproksimacija inverznog Hessijana. Želimo pronaći aproksimaciju inverznog Hessijana $\boldsymbol{H}_{k+1}^{-1} = [\nabla^2 \boldsymbol{f}(\boldsymbol{x}^{(k+1)})]^{-1}$, gdje je $\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{s}_k$. Quasi-Newtonov uvjet bio je oblika $\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k$ pa je inverzni Quasi-Newtonov uvjet

$$\boldsymbol{H}_{k+1}^{-1}\boldsymbol{y}_k = \boldsymbol{s}_k.$$

BFGS metoda dobiva se iz DFP metode tako da se $\boldsymbol{s}_k$ i $\boldsymbol{y}_k$ zamjene. Aproksimacija Hessijana u točki $\boldsymbol{x}^{(k+1)}$ računa se na način da se aproksimaciji Hessijana $\boldsymbol{H}_k$ dodaju dvije simetrične matrice pri čemu je svaka od njih matrica ranga 1:

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \boldsymbol{U}_k + \boldsymbol{V}_k.$$

Dakle, vrijedi

$$\operatorname{rank}(\boldsymbol{H}_{k+1} - \boldsymbol{H}_k) \leq 2$$

pa zato ova metoda pripada u metode reda 2.

Aproksimacija Hessijana se računa po formuli

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{y}_k\boldsymbol{y}_k^T}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k\rangle} - \frac{\boldsymbol{H}_k\boldsymbol{s}_k\boldsymbol{s}_k^T\boldsymbol{H}_k}{\langle \boldsymbol{s}_k, \boldsymbol{H}_k\boldsymbol{s}_k\rangle},$$

gdje je $\boldsymbol{y}_k = \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)})$ i $\boldsymbol{s}_k = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}$. Iz gornjeg izraza, primjenom Sherman-Morrisonove formule, dobivamo inverzni Hessijan

$$\boldsymbol{H}_{k+1}^{-1} = \left(I - \frac{\boldsymbol{s}_k\boldsymbol{y}_k^T}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k\rangle}\right)\boldsymbol{H}_k^{-1}\left(I - \frac{\boldsymbol{y}_k\boldsymbol{s}_k^T}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k\rangle}\right) + \frac{\boldsymbol{s}_k\boldsymbol{s}_k^T}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k\rangle}, \qquad k = 0, 1, 2, \ldots \tag{4.8}$$

Ova metoda se naziva Broyden-Fletcher-Goldfarb-Schannova metoda.

Primjer 4.3. *Pomoću Broyden-Fletcher-Goldfarb-Schannove metode odredite prve tri aproksimacije minimuma funkcije*

$$\boldsymbol{f}(x_1, x_2) = x_1 - x_2 + 2x_1^2 + 2x_1x_2 + x_2^2$$

uz početnu aproksimaciju $\boldsymbol{x}^{(0)} = [0, 0]^T$.

Rješenje: Kao i u prethodnim metodama, za $\boldsymbol{H}_0$ uzet ćemo jediničnu matricu. Dakle,

$$\boldsymbol{H}_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Gradijent funkcije $\boldsymbol{f}$ u točki $\boldsymbol{x}^{(k)}$ je

$$\nabla \boldsymbol{f}(x_1^{(k)}, x_2^{(k)}) = \begin{bmatrix} \frac{\partial \boldsymbol{f}(x_1^{(k)}, x_2^{(k)})}{\partial x_1^{(k)}} \\ \frac{\partial \boldsymbol{f}(x_1^{(k)}, x_2^{(k)})}{\partial x_2^{(k)}} \end{bmatrix} = \begin{bmatrix} 1+4x_1+2x_2 \\ -1+2x_1+2x_2 \end{bmatrix}.$$

Gradijent u točki $\boldsymbol{x}^{(0)}$ iznosi

$$\nabla \boldsymbol{f}(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Aproksimaciju minimuma u svakoj sljedećoj iteraciji računamo po formuli

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} - \boldsymbol{H}_k^{-1} \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}).$$

Izračunajmo sada aproksimaciju $\boldsymbol{x}^{(1)}$.

$$\boldsymbol{x}^{(1)} = \boldsymbol{x}^{(0)} - \boldsymbol{H}_0^{-1} \nabla \boldsymbol{f}(\boldsymbol{x}^{(0)}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Na isti način kao i u DFP metodi računamo $\boldsymbol{s}_k$. Vrijedi,

$$\boldsymbol{s}_k = \boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}.$$

Dakle, za $k = 0$ slijedi

$$\boldsymbol{s}_0 = \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(0)} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

Izračunajmo sada $\boldsymbol{y}_0$. Općenito vrijedi

$$\boldsymbol{y}_k = \nabla \boldsymbol{f}(\boldsymbol{x}^{(k+1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(k)}).$$

Dakle,

$$\boldsymbol{y}_0 = \nabla \boldsymbol{f}(\boldsymbol{x}^{(1)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(0)}) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}.$$

Za razliku od DFP metode ovdje ćemo računati aproksimaciju inverznog Hessijana i to po formuli

$$\boldsymbol{H}_{k+1}^{-1} = \left(I - \frac{\boldsymbol{s}_k \boldsymbol{y}_k^T}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle} \right) \boldsymbol{H}_k^{-1} \left(I - \frac{\boldsymbol{y}_k \boldsymbol{s}_k^T}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle} \right) + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^T}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle}$$

pri čemu je $k = 0, 1, 2, \ldots$

Izračunajmo sada $\boldsymbol{H}_1^{-1}$ iz

$$\boldsymbol{H}_1^{-1} = \left(I - \frac{\boldsymbol{s}_0 \boldsymbol{y}_0^T}{\langle \boldsymbol{y}_0, \boldsymbol{s}_0 \rangle} \right) \boldsymbol{H}_0^{-1} \left(I - \frac{\boldsymbol{y}_0 \boldsymbol{s}_0^T}{\langle \boldsymbol{y}_0, \boldsymbol{s}_0 \rangle} \right) + \frac{\boldsymbol{s}_0 \boldsymbol{s}_0^T}{\langle \boldsymbol{y}_0, \boldsymbol{s}_0 \rangle}$$

Slijedi,

$$\boldsymbol{H}_1^{-1} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\begin{bmatrix} 2 & 0 \\ -2 & 0 \end{bmatrix}}{2} \right) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\begin{bmatrix} 2 & -2 \\ 0 & 0 \end{bmatrix}}{2} \right) + \frac{\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}}{2},$$

odnosno

$$\boldsymbol{H}_1^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 2.5 \end{bmatrix}.$$

Izračunajmo sada aproksimaciju minimuma $\boldsymbol{x}^{(2)}$. Imamo,

$$\boldsymbol{x}^{(2)} = \boldsymbol{x}^{(1)} - \boldsymbol{H}_1^{-1} \nabla \boldsymbol{f}(\boldsymbol{x}^{(1)}) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 2.5 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix}.$$

Sada, preko formula za $\boldsymbol{s}_k$ i $\boldsymbol{y}_k$ računamo

$$\boldsymbol{s}_1 = \boldsymbol{x}^{(2)} - \boldsymbol{x}^{(1)} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} - \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

te

$$\boldsymbol{y}_1 = \nabla \boldsymbol{f}(\boldsymbol{x}^{(2)}) - \nabla \boldsymbol{f}(\boldsymbol{x}^{(1)}) = \begin{bmatrix} 3 \\ 3 \end{bmatrix} - \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}.$$

Izračunali smo sve što nam je potrebno kako bismo odredili aproksimaciju $\boldsymbol{H}_2^{-1}$ pa imamo

$$\boldsymbol{H}_2^{-1} = \left(I - \frac{\boldsymbol{s}_1 \boldsymbol{y}_1^T}{\langle \boldsymbol{y}_1, \boldsymbol{s}_1 \rangle} \right) \boldsymbol{H}_1^{-1} \left(I - \frac{\boldsymbol{y}_1 \boldsymbol{s}_1^T}{\langle \boldsymbol{y}_1, \boldsymbol{s}_1 \rangle} \right) + \frac{\boldsymbol{s}_1 \boldsymbol{s}_1^T}{\langle \boldsymbol{y}_1, \boldsymbol{s}_1 \rangle},$$

odakle uvrštavanjem dobivamo

$$\boldsymbol{H}_2^{-1} = \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 & 0 \\ 8 & 8 \end{bmatrix}}{8} \right) \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 2.5 \end{bmatrix} \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 & 8 \\ 0 & 8 \end{bmatrix}}{8} \right) + \frac{\begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}}{8},$$

odnosno

$$\boldsymbol{H}_2^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 1 \end{bmatrix}.$$

Sada možemo izračunati traženu aproksimaciju minimuma za $k = 2$, tj.

$$\boldsymbol{x}^{(3)} = \boldsymbol{x}^{(2)} - \boldsymbol{H}_2^{-1} \nabla \boldsymbol{f}(\boldsymbol{x}^{(2)}) = \begin{bmatrix} -1 \\ 3 \end{bmatrix} - \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ 1.5 \end{bmatrix}.$$

Literatura

[1] B. P. Demidovič i suradnici, Zadaci i riješeni primjeri iz više matematike s primjenom na tehničke nauke, Tehnička knjiga, Zagreb, 1978.

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

[3] M. Milinčević, Quasi-Newtonove metode, Završni rad, Odjel za matematiku, 2016.

[4] R. Scitovski, Numerička matematika, Izmijenjeno i dopunjeno izdanje, Odjel za matematiku, Sveučilište Josipa Jurja Strossmayera u Osijeku, Osijek, 2015.

[5] R. Scitovski, D. Marković, D. Brajković, Linearna algebra I, nastavni materijali, Odjel za matematiku, Sveučilište u Osijeku, 2020.

[6] R. Scitovski, N. Truhar, Z. Tomljanović, Metode optimizacije, Odjel za matematiku, Sveučilište Josipa Jurja Strossmayera, Osijek, 2012.

[7] Z. Drmač, M. Marušić, S. Singer, V. Hari, M. Rogina, S. Singer, Numerička analiza, Predavanja i vježbe, PMF-Matematički odjel, Sveučilište u Zagrebu, Zagreb, 2003.

Sažetak

U ovome radu ukratko smo se upoznali s Newtonovom metodom tangente i metodom sekante za rješavanje nelinearnih jednadžbi. Za obje metode izveli smo pripadne formule te riješili primjere. Drugi dio rada bavi se rješavanjem sustava nelinearnih jednadžbi. U ovome dijelu prisjetili smo se nekih važnijih pojmova poput Jacobijeve matrice i regularne matrice te smo izveli Newtonovu metodu. Naveli smo osnovna svojstva ove metode te riješili pripadni primjer. U glavnom dijelu rada upoznali smo se s Quasi-Newtonovim metodama, naveli razlog njihova korištenja te izveli Broydenovu, Davidon-Fletcher-Powellovu i Broyden-Fletcher-Goldfarb-Schannovu metodu. Za svaku od ove tri metode naveli smo neka osnovna svojstva te riješili po jedan primjer.

Ključne riječi: nelinearne jednadžbe, Newtonova metoda, metoda sekante, sustav nelinearnih jednadžbi, Jacobijeva matrica, Broydenova metoda, Davidon-Fletcher-Powellova metoda, Broyden-Fletcher-Goldfarb-Schannova metoda

Summary

In this paper we introduced to the Newton's tangent method and secant method for solving nonlinear equations. For both methods we derived formulas and solved examples. Second part of paper refers to solving systems of nonlinear equations. In this part we remembered some important terms such as Jacobian matrix and regular matrix and we derived formula for Newton's method. We gave a main properties of Newton's method and solved one example. In the main part of paper we introduced to the Quasi-Newton's methods, we gave the reason why we use them and derived formulas for Broyden, Davidon-Fletcher-Powell and Broyden-Fletcher-Goldfarb-Schanno method. For each of these three methods we gave some important properties and we solved examples.

Key words: nonlinear equations, Newton's method, secant method, system of nonlinear equations, Jacobian matrix, Broyden method, Davidon-Fletcher-Powell method, Broyden-Fletcher-Goldfarb-Schanno method

Životopis

Rođena sam 4.10.1995. godine u Osijeku. Pohađala sam Osnovnu školu Fran Krsto Frankopan nakon koje upisujem III. gimnaziju u Osijeku. Tijekom srednje škole sudjelovala sam u natjecanju mladih Hrvatskog Crvenog križa te sam, zajedno s ekipom, osvojila četvrto mjesto na državnom natjecanju. Više od 10 godina volontiram u Crvenom križu te sam sudjelovala u brojnim aktivnostima kao što su radionice pružanja prve pomoći u srednjim školama, dežuranje na Svjetskom gimnastičkom kupu (World Cup Osijek), dežuranje na ispraćaju maturanata te prikupljanje i sortiranje pomoći za stradale u potresu. Godine 2014. upisujem preddiplomski studij matematike na Odjelu za matematiku u Osijeku. Preddiplomski studij završila sam 2018. godine uz završni rad na temu Runge Kutta metode za numeričko rješavanje diferencijalnih jedandžbi. Iste godine upisujem diplomski studij Financijske matematike i statistike. Stručnu praksu odradila sam u Zagrebačkoj banci. Krajem 2020. radila sam na zamjeni iz matematike u Osnovnoj školi Antunovac. Godine 2021. završila sam Pedagoško psihološko didaktičko metodičku izobrazbu na Filozofskom fakultetu u Osijeku.