Nelinearne jednadžbe:

- Uvod
- Jednodimenzionalna minimizacija
- Metoda zlatnog reza
- Višedimenzionalna minimizacija
- Gradijentna metoda
- Modificirana Newtonova metoda
- Kvazi-Newtonove metode
- Konvergencija
- Primjer

Optimizacija
Traženje minimuma funkcije

Zadana funkcija

$$f : \mathbb{R}^n \to \mathbb{R}$$

Cilj: Naći x^*, točku minimuma funkcije f:

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

- Problem je jednostavno opisati

- Rješavanje je jedno od najtežih područja numeričke analize.

- Cilj optimizacije je definirati prikladne, tj. što brže postupke, za rješavanje ovog tipa problema.

Nužan uvjet minimuma:

$$\nabla f(\alpha) = 0.$$

Umjesto minimizacije možemo tražiti nultočku gradijenta?

Nemamo garanciju konvergencije prema minimumu. (maksimum, infleksija)

Nekada je bolje umjesto

$$f(x) = 0$$

rješavati ekvivalentni problem

$$f^2(x) \to \min \quad \text{ili} \quad \sum_{i=1}^{n} f_i^2(x) \to \min$$

- Neke poteškoće:

  - nepoznata analitička svojstva funkcije f

  - globalni minimum

Traženje minimuma

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

je potuno analogan problem nalaženju maksimuma funkcije f.

Traženje maksimuma od f ekvivalentno je traženju minimuma funkcije $-f$:

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

$\rightarrow$ Termin optimizacija ravnopravno se koristi s terminom **minimizacija funkcija**.

Često se minimum funkcije f ne traži na cijelom području $\mathbb{R}^n$, već na nekom podskupu $D \subset \mathbf{R^n}$:

$$\min_{x \in D} f(x).$$

$\rightarrow$ Problem **minimizacije s ograničenjima**.

- Problem izgleda jednostavniji (zbog manjeg područja minimizacije),ali metode za njegovo rješavanje su složenije nego za polazni problem.

-Uz funkciju f moramo voditi računa i o rubu područja D (često definiranog pomoću skupa funkcija).

Primjer: Funkcija $f(x) = x^2$ ima globalni minimum u 0.

Ako tražimo minimimum na intervalu $[a, b]$:

minimum funkcije f jednak 0 ako je $0 \in [a, b]$ ili

$\min\{f(a), f(b)\}$ ako $0 \notin [a, b]$.

- u razmatranje smo trebali uzeti i rubove intervala.

U praksi se često javlja problem optimizacije s posebnim oblikom funkcije f i skupa D. Npr.,

- D presjek poluravnina u $\mathbb{R}^n$
- f je linearna funkcija

$\rightarrow$ **problem linearnog programiranja**.

- f kvadratična funkcija (polinom)

$\rightarrow$ **problem kvadratičnog programiranja**.

Jednodimenzionalna minimizacija. Metoda zlatnog reza

$f : \mathbb{R} \to \mathbb{R}$

Zadane su tri točke $a < b < c$

Poznate su vrijednosti funkcije f u njima: $f(a)$, $f(b)$ i $f(c)$

Vrijedi: $f(b) \leq f(a)$ i $f(b) \leq f(c)$

- vrijednost funkcije f je najmanja u srednjoj točki b.

- točka lokalnog minimuma nalazi u intervalu $[a, c]$

-Izaberimo novu točku x iz intervala $[b, c]$.

- Pretpostavit ćemo da je $x \in (b, c)$.

- Dvije mogućnosti: $f(x) \geq f(b)$ ili $f(x) < f(b)$.

- Ako je $f(x) \geq f(b)$
  - minimum nalazi u $[a, x]$ (jer je $f(b) \leq f(a)$ i $f(b) \leq f(x)$)
  - novu točku biramo iz trojke (a, b, x).

- Ako je $f(x) < f(b)$
  - minimum se nalazi u intervalu $[b, c]$
  - novu točku biramo iz trojke (b, x, c).

Opisani postupak je osnova metode zlatnog reza.

Preostalo je još za razmotriti način izbora nove točke x.

Neka je w omjer u kojem b dijeli interval $[a, c]$:

$$w = \frac{b-a}{c-a} \quad \text{i} \quad \frac{c-b}{c-a} = 1 - w.$$

Izaberemo točku $x \in (b, c)$. Označimo

$$z = \frac{x-b}{c-a}.$$

- Za $f(x) \geq f(b)$ minimum je lociran na intervalu širine $x - a$
- Za $f(x) < f(b)$ minimum je lociran na intervalu širine $c - b$.

1. zahtjev. I u jednom i u drugom slučaju širina intervala je jednaka:

$$x - a = c - b,$$

$$\Rightarrow \quad z + w = \frac{x-b+b-a}{c-a} = \frac{x-a}{c-a} = \frac{c-b}{c-a} = 1 - w.$$

$$\Rightarrow \quad z = 1 - 2w$$

- točke x i b smještene su simetrično s obzirom na središte intervala $[a, c]$.

Povoljnim izborom početne točke b možemo postići da vrijedi

2. zahtjev. x dijeli interval $[b, c]$ u istom omjeru kao što b dijeli interval $[a, c]$.

$$\Rightarrow \quad \frac{c-b}{c-a} = \frac{c-x}{c-b},$$

$$\Rightarrow \quad 1 - w = 1 - \frac{z}{1-w}.$$

$$z = 1 - 2w \quad \Rightarrow \quad w^2 - 3w + 1 = 0.$$

- dva rješenja$\rightarrow$ zanima nas ono koje je manje od 1:

$$w = \frac{3-\sqrt{5}}{2} \approx 0.38197.$$

Gornji je broj poznat kao zlatni broj pa se metoda s ovakvim izborom nove točke x naziva metodom zlatnog reza.

Uočimo da vrijedi

$$|c^{(i+1)} - a^{(i+1)}| = (1 - w)|c^{(i)} - a^{(i)}| \approx 0.61803|c^{(i)} - a^{(i)}|$$

- metoda konvergira linearno prema točki minimuma x^*.

Za inicijalno određivanje trojke $(a^{(0)}, b^{(0)}, c^{(0)})$ krenemo od bilo kojeg para točaka a i b. Bez smanjenja općenitosti pretpostavimo da je $f(a) \geq f(b)$. Točku c odredimo tako da b dijeli interval $[a, c]$ u zlatnom omjeru. Ukoliko je $f(b) \leq f(c)$ našli smo traženu trojku. U suprotnom slučaju ponovimo postupak s točkama a i c (ili b i c).

Oprez. Općenito nema garancije da ćemo naći početnu trojku.

Brentova metoda

- Metoda inverzne parabole.
- U blizini minimuma očekujemo da parabola dobro aproksimira funkcijau.
- ista ideja kao kod traženja nultočke.
- U n-toj iteraciji tri točke: a_n, b_n, c_n.
- U srednjoj točki je vrijednost funkcije najmanja.
- Nova točka je tjeme interpoliraju 'ce parabole.
- Zadržimo tri točke koje garantiraju postojanje minimuma.
- Konvergira brže nego metoda zlatnog reza.
- Može se kombinirati ove dvije metode: prvo metoda zlatnog reza a zatim Brentova metoda.

Višedimenzionalna minimizacija

Zadana je funkcija $f : \mathbb{R}^n \to \mathbb{R}$
Cilj je odrediti

$$\min_{x \in \mathbb{R}^n} f(x).$$

Ideja:

Za neku zadanu točku

$$x_0 \in \mathbb{R}^n$$

odredi točku

$$x_1 \in \mathbb{R}^n$$

takva da je

$$f(x_1) < f(x_0).$$

- postupak ponovimo.

Za točku $x_i \in \mathbb{R}^n$, tražimo točku $x_{i+1} \in \mathbb{R}^n$ za koju je

$$f(x_{i+1}) < f(x_i).$$

Generiramo niz točaka $x_0, x_1, x_2, \ldots$ za koje je

$$f(x_0) > f(x_1) > f(x_2) > \cdots$$

Ako je niz $(f(x_i))_i$ odozdo ograničen, tada postoji $\lim\limits_i f(x_i)$.

- $\lim\limits_i x_i$ ne treba nužno postojati.

Npr., za

$$f(x) = e^{-x}, \qquad x_i = i$$

je $\lim\limits_i f(x_i) = 0$, no $\lim\limits_i x_i$ ne postoji.

Za egzistenciju $\lim_i x_i$ nužno je da postoji $x_0 \in \mathbb{R}^n$ takav da je skup

$$\{x \mid f(x) \leq f(x_0)\}$$

kompaktan.

Osnovno je pitanje kako izabrati x_{i+1}.

s - smjer u kojem funkcija f lokalno pada u okolini točke x: .

$$f(x + \lambda s) < f(x), \quad \text{za mali } \lambda, \quad \lambda \in \mathbb{R}, \quad \lambda > 0.$$

Promatrajmo funkciju $g : \mathbb{R} \to \mathbb{R}$ definiranu s

$$g(\lambda) = f(x + \lambda s).$$

Vrijedi: $g(0) = f(x)$.

f glatka funkcija ($f \in C^1$) $\quad \Rightarrow \quad g \in C^1$.

f pada u smjeru s oko točke x $\quad \Leftrightarrow \quad$ g pada u 0, tj. $g'(0) < 0$.

Jer

$$g'(\lambda) = \langle \nabla f(x + \lambda s), s \rangle, \quad \Rightarrow \quad g'(0) = \langle \nabla f(x), s \rangle < 0.$$

Definiramo skup

$$D(x) = \{s \mid \langle \nabla f(x), s \rangle < 0\}.$$

Ovaj skup nazivamo **skup smjerova silaska** a njegove elemente **smjerovi silaska**.

Osnova niza minimizacijskih metoda.

Za zadani $x_i \in \mathbb{R}^n$ definiramo

$$x_{i+1} = x_i + \lambda_i s_i$$

gdje je $s_i \in D(x_i)$.

Konstantu λ_i nazivamo **korakom** minimizacije.

Odabiremo je tako da bude zadovoljeno $f(x_{i+1}) < f(x_i)$.

To je moguće jer je s_i smjer silaska.

Izbor smjera silaska ovisi o izboru metode minimizacije.

Različite metode na različite načine određuju izbor smjera silaska.

Veličina koraka minimizacije je bitna za konvergenciju metode.

Nije dovoljno odrediti točku x_{i+1} koja zadovoljava $f(x_{i+1}) < f(x_i)$.

Primjer. Minimizacija funkcije $f(x) = x^2$. Izborom

$$x_0 = 2, \quad x_1 = 1.5, \quad \ldots, \quad x_i = 1 + \frac{1}{i+1},$$

- niz točaka $(x_i)_i$ konvergira

- niz padajućih vrijednosti $(f(x_i))_i$ također konvergira.

- ovi nizovi ne konvergiraju prema točki minimuma, odnosno minimumu funkcije f.

Jedan od načina izbora koraka λ_i je **maksimalno spuštanje** u smjeru s_i. Tada je

$$\lambda_i = \arg\min_{\lambda>0} f(x_i + \lambda s_i).$$

- Ovakav pristup zahtijeva primjenu analitičkih metoda.

- nije jednostavno pri upotrebi računala, naročito ako je funkcija f složenijeg oblika.

Drugi način je **približno** određivanje λ_i nekom minimizacijskom metodom.

- koristimo metode za jednodimenzionalnu minimizaciju (npr. metodu zlatnog reza ili Brentovu metodu).

Točka x_{i+1} nije nužno točka (globalnog ili lokalnog) minimuma

u njoj ponavljamo cijeli postupak: izabiremo novi smjer silaska i korak minimizacije.

- korak minimizacije nije nužno izračunati egzaktno ili s prevelikom točnošću

- utrošeno vrijeme ne rezultira s odgovarajućim rezultatom.

Jedan od alternativnih pristupa je **neegzaktno pretraživanje po pravcu**.

- u konačnom broju koraka određuje se λ_i

- ne minimizira se $f(x_i + \lambda s_i)$

- dovoljno se smanjuje vrijednost $f(x_{i+1})$ u odnosu na $f(x_i)$ tako da minimizacijska metoda konvergira, tj. da niz $(x_i)_i$ konvergira točki minimuma funkcije f.

Gradijentna metoda

Smjer silaska je:

$$s = -\nabla f(x)$$

s je smjer silaska u točki x:

$$\langle s, \nabla f(x) \rangle = \langle -\nabla f(x), \nabla f(x) \rangle = -\|\nabla f(x)\|^2 < 0$$

za $\nabla f(x) \neq 0$.

$\nabla f(x) = 0$ - zadovoljen kriterij konvergencije (skup smjerova silaska je prazan skup).

- To ne treba značiti da smo došli do točke minimuma.

- ova metoda lokalno bira smjer najbržeg silaska (**metoda najbržeg silaska**, m. najstrmijeg silaska - 'steepest descent method')

- nije najbrža u traženju globalnog minimuma.

- za primjenu metode funkcija f treba biti glatka ($f \in C^1$)

Modificirana Newtonova metoda

Metoda se zasniva na Newton–Raphsonovoj metodi za rješavanje nelinearne jednadžbe

$$h(x) = 0.$$

Newton–Raphsonova metoda za $h : \mathbb{R}^n \to \mathbb{R}^n$:

$$x_{i+1} = x_i - [H_h(x_i)]^{-1} \, h(x_i).$$

H_h - Jacobijeva matrica - $\left(\dfrac{\partial h_i}{\partial y_j} \right)_{ij}$

Umjesto

$$f(x) \to \min$$

rješavamo

$$\nabla f(x) = 0.$$

- nužan uvjet minimuma.

- ako je f konveksna funkcija, tada je taj uvjet i dovoljan.

Newton-Raphsonova metoda generira niz točaka:

$$x_{i+1} = x_i - [\nabla^2 f(x_i)]^{-1} \nabla f(x_i), \tag{1}$$

$\nabla^2 f$ - Hessijan (ili Hessova matrica) funkcije f:

$$[\nabla^2 f]_{ij} = \frac{\partial^2 f}{\partial y_i \partial y_j}.$$

Da li vrijedi $f(x_{i+1}) < f(x_i)$?

Je li

$$-[\nabla^2 f(x_i)]^{-1} \nabla f(x_i) \in D(x_i)?$$

Neka je H pozitivno definitna matrica.

$(Hx, x) > 0$ za sve $x \in \mathbb{R}^n$ i $x \neq 0$.

$$\Rightarrow \quad \langle -H\nabla f(x), \nabla f(x)\rangle = -\langle H\nabla f(x), \nabla f(x)\rangle < 0$$

za $\nabla f(x) \neq 0$.

$$\Rightarrow \quad -H\nabla f(x) \in D(x).$$

x^* točka minimuma funkcije $f \quad \Rightarrow \quad \nabla^2 f(x^*)$ je pozitivno definitna.

$\Rightarrow$ $\nabla^2 f(x)$ pozitivno definitna za x iz neke okoline točke minimuma x^*

$\Rightarrow$ $[\nabla^2 f(x)]^{-1}$ pozitivno definitna

$\Rightarrow \ -[\nabla^2 f(x)]^{-1}\nabla f(x) \in D(x)$

za x u okolini točke minimuma.

Ovo ne garantira da generirani niz zadovoljava $f(x_{i+1}) < f(x_i)$.

Modificirana Newtonova metoda:

$$\begin{aligned} s_i &= -[\nabla^2 f(x_i)]^{-1}\nabla f(x_i) \\ x_{i+1} &= x_i - \lambda_i[\nabla^2 f(x_i)]^{-1}\nabla f(x_i). \end{aligned}$$

- korak minimizacije λ_i biramo na jedan od prije opisanih načina.

- modifikacija je uvođenje koraka minimizacije λ_i u Newtonovu metodu.

- nužno je da je funkcija $f \in C^2$.

- metoda je primjenjiva samo na području gdje je $\nabla^2 f$ pozitivno definitna.

- ako f nije konveksna funkcija, tada je metoda primjenjiva samo u okolini točke minimuma, a ne na cijelom području minimizacije.

- u svakom je koraku minimizacije nužno izračunati Hessijan, te riješiti sustav

$$\nabla^2 f(x_i)(x_{i+1} - x_i) = -\nabla f(x_i)$$

(ovdje nema potrebe za invertiranjem Hessijana).

- radi ubrzanja metode, često se Hessijan ne računa u svakoj iteraciji.

- prednost metode je brža konvergencija generiranog niza $(x_i)_i$ k točki minimuma.

Usporedba na kvadratnoj formi

Prednosti modificirane Newtonove metode u odnosu na gradijentnu metodu ilustrirat ćemo na minimizaciji kvadratne forme.

A pozitivno definitna matrica.

Minimizirat ćemo kvadratnu formu

$$f(y) = \frac{1}{2} \langle Ay, y \rangle + \langle b, y \rangle + c.$$

Gradijent funkcije f je dan s

$$\nabla f(y) = Ay + b,$$

Točka minimuma je:

$$\nabla f(y) = 0 \quad \Rightarrow \quad x^* = -A^{-1} b.$$

Modificirana Newtonova metoda:

$$x_{i+1} = x_i - \lambda_i [\nabla^2 f(x_i)]^{-1} \nabla f(x_i),$$

uz egzaktni izbor koraka

$$\lambda_i = \arg\min_{\lambda > 0} f(x_i - \lambda [\nabla^2 f(x_i)]^{-1} \nabla f(x_i))$$

definira iteracije ($\nabla^2 f(x_i) = A$)

$$x_{i+1} = x_i - \lambda_i A^{-1}(Ax_i + b) = (1 - \lambda_i)x_i - \lambda_i A^{-1} b.$$

Egzaktni izbor koraka daje $\lambda_0 = 1$.

Za bilo koji izbor x_0 modificirana Newtonova metoda daje egzaktno rješenje u jednom koraku.

Gradijentna metoda na kvadratnoj formi općenito ne dolazi do minimuma u konačnom broju koraka.

Izaberimo matricu A oblika

$$A = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix},$$

gdje je $0 < \alpha_1 < \alpha_2$. Stavimo $b = 0$ i $c = 0$.

Kvadratna forma je

$$f(y) = \frac{1}{2}\langle Ay, y\rangle = \frac{1}{2}(\alpha_1 y_1^2 + \alpha_2 y_2^2),$$

Za $\alpha_1 = \alpha_2 > 0$ gradijentna metoda u jednom koraku dolazi do točke minimuma

$$x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Neka je početni vektor

$$x_0 = \begin{bmatrix} x_{0,1} \\ x_{0,2} \end{bmatrix}$$

$x_{0,1} \neq 0$ i $x_{0,2} \neq 0$.

(Ako bi jedna komponenta bila 0, tada gradijentna metoda u jednom koraku došli do minimuma.)

Za

$$x_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \end{bmatrix}$$

egzaktni izbor koraka λ_i daje

$$x_{i+1,1} = \frac{\alpha_2^2(\alpha_2 - \alpha_1)x_{i,1}x_{i,2}^2}{\alpha_1^3 x_{i,1}^2 + \alpha_2^3 x_{i,2}^2}, \qquad x_{i+1,2} = \frac{\alpha_1^2(\alpha_1 - \alpha_2)x_{i,2}x_{i,1}^2}{\alpha_1^3 x_{i,1}^2 + \alpha_2^3 x_{i,2}^2}.$$

$$(x_{i,1} \neq 0 \text{ i } x_{i,2} \neq 0) \text{ i } (\alpha_1 > \alpha_2 > 0) \quad \Rightarrow \quad x_{i+1,1} \neq 0 \text{ i } x_{i+1,2} \neq 0$$

- indukcijom zaključujemo da niti u jednom koraku ne dolazimo do točke minimuma.

Kvazi-Newtonove metode

- Gradijentna metoda
  - potrebno samo računanje derivacija (gradijenta)
  - ne zahtjeva konveksnost
  - do minimuma kvadratne forme ne dolazi u konačnom broju koraka
- Newtonova metoda
  - potrebno računanje prvih derivacija (gradijenta) i drugih derivacija (Hessijan)
  - zahtjeva konveksnost
  - do minimuma kvadratne forme ne dolazi u jednom koraku
- Kvazi-Newtonove metode
  - potrebno samo računanje prvih derivacija (gradijenta)
  - ne zahtjeva konveksnost
  - do minimuma kvadratne forme dolaze u **?** koraka

Pokazali smo:
H pozitivno definitna matrica
$\Rightarrow \quad -H\nabla f(x)$ je smjer silaska.

Promatramo metode za koje je smjer silaska definiran s

$$s_i = -H_i \nabla f(x_i),$$

gdje je H_i pozitivno definitna matrica.

Ako je

$$H_i = [\nabla^2 f(x_i)]^{-1}$$

to je modificirana Newtonova metoda.

Ideja: matricu H_{i+1} relativno jednostavno izračunati iz prethodne matrice H_i.

Nećemo u svakom koraku rješavati sustav

$$\nabla^2 f(x_i)s_i = -\nabla f(x_i),$$

već u svakom koraku množimo

$$H_i \nabla f(x_i),$$

što je daleko brže.

U svakom koraku matricu H_{i+1} generirat ćemo tako da bude zadovoljeno

$$H_{i+1}(\nabla f(x_{i+1}) - \nabla f(x_i)) = x_{i+1} - x_i.$$

Ovaj uvjet je sličan uvjetu koji zadovoljava Hessijan (iz modificirane Newtonovove metode):

$$\nabla f(x_{i+1}) - \nabla f(x_i) \approx \nabla^2 f(x_{i+1})(x_{i+1} - x_i).$$

Metode koje za izbor smjera silaska koriste

$$s_i = -H_i \nabla f(x_i),$$

pri čemu je

$$H_{i+1}(\nabla f(x_{i+1}) - \nabla f(x_i)) = x_{i+1} - x_i.$$

kvazi-Newtonove metode.

(Koristi i naziv **metode promjenjive metrike**).

Za niz matrica koje zadovoljavaju gornji uvjet može se pokazati da vrijedi

$$\lim_{i\to\infty} H_i = [\nabla^2 f(x^*)]^{-1},$$

gdje je x^* točka minimuma.

Ovo dodatno opravdava naziv metode.

Sada ćemo vidjeti kako računati H_i da bi vrijedio uvjet.

Označimo

$$q_i = \nabla f(x_{i+1}) - \nabla f(x_i) \quad \text{i} \quad p_i = x_{i+1} - x_i.$$

Tražimo pozitivno definitnu matricu H_{i+1} koja zadovoljava

$$H_{i+1} q_i = p_i.$$

H_{i+1} bi trebala biti ‘lako izračunljiva’ iz matrice H_i.

Radi jednostavnosti, izbacimo indekse:

$$H = H_i, \quad p = p_i, \quad q = q_i, \quad H' = H_{i+1}.$$

Uvjet sada glasi

$$H' q = p.$$

'Lako izračunljiva':

$$H' = H + E.$$

gdje je E relativno jednostavna matrica.

Sada smo problem sveli na određivanje matrice E koja zadovoljava

$$(H + E)q = p,$$

odnosno

$$Eq = p - Hq.$$

H i H' su simetrične $\rightarrow$ E simetrična.

Najjednostavnija simetrična matrica M koja zadovoljava

$$Mx = y$$

za zadane x i y ima oblik

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

M zadovoljava:

$$Mx = \frac{yy^T}{\langle y, x\rangle} x = \frac{1}{\langle y, x\rangle} y(y^T x) = \frac{1}{\langle y, x\rangle} y \langle y, x\rangle = y.$$

Za

$$A = \frac{pp^T}{p^T q}$$

vrijedi $Aq = p$, a za

$$B = \frac{Hq(Hq)^T}{(Hq)^T q} = \frac{Hqq^T H}{q^T Hq}$$

vrijedi

$$Bq = \frac{Hq(Hq)^T q}{(Hq)^T q} = Hq.$$

Matrica E definirana s $A - B$:

Matrica E definirana s:

$$E = A - B = \frac{pp^T}{p^T q} - \frac{Hqq^T H}{q^T Hq}$$

zadovoljava $Eq = p - Hq$.

Dakle, H_{i+1} definirana s

$$H_{i+1} = H_i + \frac{p_i p_i^T}{p_i^T q_i} - \frac{H_i q_i q_i^T H_i}{q_i^T H_i q_i},$$

gdje je

$$q_i = \nabla f(x_{i+1}) - \nabla f(x_i),$$

a

$$p_i = x_{i+1} - x_i,$$

zadovoljava kvazi-Newtonovu jednadžbu.

Ova metoda poznata je pod nazivom '**DFP metoda**' (Davidon, Fletcher, Powel).

H_{i+1} je dobro definirana.

Nazivnici u definiciji H_{i+1} su različiti od 0:

$$p_i^T q_i > 0 \quad \text{i} \quad q_i^T H_i q_i > 0$$

ako je $\nabla f(x_i) \neq 0$.

Za egzaktni izbor koraka

$$f(x_i + \lambda_i s_i) = \min_{\lambda \geq 0} f(x_i + \lambda s_i)$$

vektori s_i i $\nabla f(x_{i+1})$ okomiti.

Jer je λ_i točka minimuma:

$$0 = \frac{d}{d\lambda} f(x_i + \lambda s_i)\bigg|_{\lambda=\lambda_i} = \langle \nabla f(x_i + \lambda_i s_i), s_i \rangle = \langle \nabla f(x_{i+1}), s_i \rangle.$$

Ako korak ne biramo egzaktno, već nekom numeričkom metodom, gornja jednakost neće vrijediti.

U tom slučaju je jednoznačno određen $\mu_i \neq 0$ za koji je

$$\langle \nabla f(x_{i+1}), s_i \rangle = \mu_i \langle \nabla f(x_i), s_i \rangle.$$

Za egzaktan izbor koraka je $\mu_i = 0$, dok za neegzaktan izbor očekujemo da je $\mu_i \approx 0$.

Primjenom gornjih oznaka dobivamo

$$\begin{aligned} p_i^T q_i &= \langle \nabla f(x_{i+1}) - \nabla f(x_i), x_{i+1} - x_i \rangle \\ &= \langle \nabla f(x_{i+1}) - \nabla f(x_i), \lambda_i s_i \rangle \\ &\vdots \\ &= -\lambda_i(\mu_i - 1)\langle \nabla f(x_i), H_i \nabla f(x_i) \rangle. \end{aligned}$$

H_i pozitivno definitna i $\nabla f(x_i) \neq 0 \ \rightarrow \ p_i^T q_i > 0$

Uz uvjet da je $\mu_i < 1$.

Ovo nije veliko ograničenje jer kod neegzaktnog izbora koraka je $\mu_i \approx 0$.

H_i pozitivno definitna $\rightarrow$

$$q_i^T H_i q_i \geq 0.$$

$= 0$ samo ako je i $q_i = 0$ odnosno

$$\nabla f(x_{i+1}) = \nabla f(x_i).$$

Tada je

$$\langle \nabla f(x_{i+1}), s_i \rangle = \langle \nabla f(x_i), s_i \rangle$$

te je $\mu_i = 1$.

Znači, $\mu_i < 1$ garantira da je

$$q_i^T H_i q_i > 0.$$

Ovime smo pokazali da je matrica H_{i+1} dobro definirana.

H_i je pozitivno definitna $\quad\rightarrow\quad$ simetrična.

Iz definicije matrice H_{i+1} jasno je da je i ona simetrična, no pozitivna definitnost nije tako očita.

Uzmimo proizvoljan vektor $y \in \mathbb{R}^n$, $y \neq 0$.
H_i pozitivno definitna $\quad\rightarrow$

$$H = LL^T,$$

gdje je L regularna donjetrokutasta matrica.

Uz oznake $u = L^T y$ i $v = L^T q_i$, izlazi

$$\langle H_{i+1} y, y \rangle = u^T u + \frac{(p^T y)^2}{p^T q} - \frac{(v^T u)^2}{v^T v}.$$

Kako je $p^T q > 0$, vrijedi

$$\frac{(p^T y)^2}{p^T q} \geq 0.$$

S druge strane vrijedi

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

Posljedica Schwartzove nejednakosti.

Ovime smo pokazali da je

$$\langle H_{i+1} y, y\rangle \geq 0.$$

Sada ćemo pokazati da je

$$\langle H_{i+1} y, y\rangle \neq 0.$$

Kad bi bilo $\langle H_{i+1} y, y\rangle = 0$, onda bi moralo vrijediti i

$$\|u\|^2\|v\|^2 - \langle u, v\rangle^2 = 0,$$

$\Rightarrow$ u i v kolinearni: $u = \alpha v$.

No, tada je $y = \alpha q$, a $\alpha \neq 0$ jer je $y \neq 0$.

Nadalje, zbog $\langle H_{i+1}y, y\rangle = 0$ treba vrijediti i

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

U kontradikciji s prije dokazanim:

$$0 < p^T q = \langle p, q\rangle.$$

Ovime smo dokazali sljedeći teorem.

Teorem

Neka je H_i pozitivno definitna matrica i $\nabla f(x_i) \neq 0$, te neka je korak minimizacije određen tako da je $\mu_i < 1$. Tada je matrica H_{i+1} dobro definirana i pozitivno definitna te zadovoljava kvazi-Newtonovu jednadžbu.

Generalizacija

DFP metoda može se generalizirati složenijim izborom matrice H_{i+1}.

Neka je

$$q_i = \nabla f(x_{i+1}) - \nabla f(x_i) \quad \text{i} \quad p_i = x_{i+1} - x_i.$$

Za parametre

$$\gamma_i > 0, \quad \theta_i \geq 0$$

definiramo rekurziju oblika

$$H_{i+1} = \Psi(\gamma_i, \theta_i, H_i, p_i, q_i),$$

gdje je funkcija Ψ zadana s

$$\begin{aligned}\Psi(\gamma, \theta, H, p, q) &= \gamma H + \left(1 + \gamma\theta \frac{q^T H q}{p^T q}\right) \frac{p p^T}{p^T q} \\ &\quad - \gamma \frac{(1-\theta)}{q^T H q} Hq \cdot q^T H \\ &\quad - \frac{\gamma\theta}{p^T q} (p q^T H + H q p^T).\end{aligned}$$

Ψ je definirana jedino ako je $p^T q \neq 0$ i $q^T H q \neq 0$.

H_{i+1} je dobivena iz H_i dodavanjem korekcije reda ne većeg od 2 matrici $\gamma_i H_i$:

$$\mathrm{rang}(H_{i+1} - H_i) \leq 2.$$

Stoga se kaže da se radi o metodi reda dva.

Metoda obuhvaća nekoliko poznatih metoda kao specijalne slučajeve:

(a) $\gamma_i = 1$, $\theta_i = 0$; prije opisana 'DFP metoda';

(b) $\gamma_i = 1$, $\theta_i = 1$; Broyden, Fletcher, Goldfarb

i Shannova metoda reda 2 ('BFGS metoda');

(c) $\gamma_i = 1$, $\theta_i = p_i^T q_i / (p_i^T q_i - q_i^T H_i q_i)$; Broydenova simetrična metoda reda jedan.

Za ove metode vrijedi teorem analogan teoremu za DFP metodu.

Teorem

Ako postoji $i \geq 0$ za koji je H_i pozitivno definitna matrica,

$$\nabla f(x_i) \neq 0$$

i ako je korak minimizacije određen tako da je $\mu_i < 1$, tada je za sve $\gamma_i > 0$ i $\theta_i \geq 0$ matrica $H_{i+1} = \Psi(\gamma_i, \theta_i, H_i, p_i, q_i)$ dobro definirana i pozitivno definitna te zadovoljava kvazi-Newtonovu jednadžbu.

Kvazi-Newtonove metode i kvadratna forma

Kvazi-Newtonove metode dolaze do točke minimuma kvadratne funkcije u najviše n koraka.

Uz uvjet da je korištena egzaktna minimizacija pri izboru koraka minimizacije.

Teorem

Neka je

$$f(x) = \frac{1}{2} \langle Ax, x \rangle + \langle b, x \rangle + c$$

kvadratna forma gdje je A $n \times n$ pozitivno definitna matrica. Nadalje, neka je $x_0 \in \mathbb{R}^n$ i H_0 $n \times n$ pozitivno definitna matrica. Ako je za minimizaciju $f(x)$ korištena kvazi-Newtonova metoda definirana uz egzaktan izbor koraka s početnim vrijednostima x_0 i H_0,

tada generirani nizovi
$(x_i)_i$, $(H_i)_i$, $(\nabla f(x_i))_i$, $(p_i)_i = (x_{i+1} - x_i)_i$ *i* $(q_i)_i = (\nabla f(x_{i+1}) - \nabla f(x_i))_i$
imaju sljedeća svojstva:

(a) *Postoji najmanji indeks* $m \leq n$ *za koji je* $x_m = x^* = -A^{-1}b$ *minimum od* f *i* $\nabla f(x_m) = 0$.

(b) $\langle p_i, q_k \rangle = \langle p_i, Ap_k \rangle = 0$ *za* $0 \leq i \neq k \leq m-1$, $\langle p_i, q_i \rangle > 0$ *za* $0 \leq i \leq m-1$. *(Vektori* p_i *su A-konjugirani.)*

(c) $\langle p_i, \nabla f(x_k) \rangle = 0$ *za sve* $0 \leq i < k \leq m$.

(d) $H_k q_i = \gamma_{i,k} p_i$ *za* $0 \leq i < k \leq m$, *gdje je*

$$\gamma_{i,k} = \begin{cases} \gamma_{i+1}\gamma_{i+2}\cdots\gamma_{k-1}, & \text{za } i < k-1, \\ 1 & \text{za } i = k-1 \end{cases}$$

(e) *Ako je $m = n$ tada dodatno vrijedi*

$$H_m = H_n = PDP^{-1}A^{-1},$$

gdje je $D = diag(\gamma_{0,n}, \gamma_{1,n}, \ldots, \gamma_{n-1,n})$, $P = (p_0, p_1, \ldots, p_{n-1})$.
Ako je i $\gamma_i = 1$ tada je $H_n = A^{-1}$.

Dokaz. Promatrajmo sljedeće uvjete za proizvoljan indeks $l \geq 0$:

$$\begin{aligned}
\langle p_i, q_k \rangle &= \langle p_i, Ap_k \rangle = 0 \quad \text{za} \quad 0 \leq i \neq k \leq l-1, \\
\langle p_i, q_i \rangle &> 0 \quad \text{za} \quad 0 \leq i \leq l-1, \\
H_l & \quad \text{je pozitivno definitna}, \\
\langle p_i, q_k \rangle &= 0 \quad \text{za sve} \quad 0 \leq i < k \leq l, \\
H_k q_i &= \gamma_{i,k} p_i \quad \text{za} \quad 0 \leq i < k \leq l.
\end{aligned}$$

Ako su ovi uvjeti zadovoljeni za l te ako je $\nabla f(x_l) \neq 0$ tada ćemo pokazati da su uvjeti zadovoljeni i za $l+1$.

⋮ $\square$

Konvergencija minimizacijskih metoda

Skup smjerova silaska

$$D(x) = \{s \mid \langle \nabla f(x), s \rangle < 0\}.$$

$s \in D(x)$ - vrijednost funkcije f smanjuje u smjeru s u nekoj okolini točke x.

Neka je $\| \ \|$ standardna Euklidska norma ($\| \ \| = \| \ \|_2$).

Za $\gamma \in \mathbb{R}$, $0 < \gamma \leq 1$, promatrajmo skup

$$D(\gamma, x) = \{s \in \mathbb{R}^n \mid \|s\| = 1, \ \langle \nabla f(x), s \rangle \leq -\gamma \|\nabla f(x)\|\}.$$

Uvjet

$$\langle \nabla f(x), s \rangle \leq -\gamma \|\nabla f(x)\|,$$

zbog $\|s\| = 1$, možemo zapisati kao

$$\frac{\langle \nabla f(x), -s \rangle}{\|\nabla f(x)\| \, \|s\|} \geq \gamma$$

ako je $\|\nabla f(x)\| \neq 0$.

Ako je $\|\nabla f(x)\| = 0$, uvjet je uvijek zadovoljen.

Označimo s Θ kut između vektora $\nabla f(x)$ i $-s$:

$$\langle \nabla f(x), -s \rangle = \cos \Theta \|\nabla f(x)\| \, \|s\|.$$

Uvjet glasi $\quad \cos \Theta \geq \gamma.$

$\rightarrow$ promatramo samo one smjerove silaska koji zatvaraju dovoljno mali kut s vektorom $-\nabla f(x)$.

Koliko je taj kut malen, definira koeficijent γ.

Sljedeća lema pokazuje pod kojim uvjetima postoji skalar λ i (smjer silaska) $s \in \mathbb{R}^n$ za koje je $f(x + \lambda s) < f(x)$.

Lema

Neka je $f : \mathbb{R}^n \to \mathbb{R}$ funkcija čiji je gradijent $\nabla f(x)$ definiran i neprekidan za sve $x \in V(\bar{x})$ u okolini $V(\bar{x})$ točke $\bar{x}$. Nadalje, pretpostavimo da je $\nabla f(\bar{x}) \neq 0$ i neka je $1 \geq \gamma > 0$. Tada postoji okolina $U(\bar{x}) \subseteq V(\bar{x})$ od $\bar{x}$ i broj $\mu > 0$ takav da je

$$f(x + \lambda s) \leq f(x) - \frac{\lambda\gamma}{4} \|\nabla f(\bar{x})\|$$

za sve $\lambda \in [0, \mu]$.

Promatrimo sljedeću metodu za minimizaciju diferencijabilne funkcije $f : \mathbb{R}^n \to \mathbb{R}$.

Metoda 1.

(a) Izaberi brojeve γ_i, σ_i, $i = 0, 1, 2, \ldots$ koji zadovoljavaju

$$\sup_i \gamma_i \leq 1, \quad \inf_i \gamma_i > 0, \quad \inf_i \sigma_i > 0,$$

i izaberi početnu točku $x_0 \in \mathbb{R}^n$.

(b) Za $i = 0, 1, 2, \ldots$ izaberi (proizvoljan) $s_i \in D(\gamma_i, x_i)$ i postavi

$$x_{i+1} := x_i + \lambda_i s_i,$$

gdje je $\lambda_i \in [0, \sigma_i \|\nabla f(x_i)\|]$ takav da je

$$f(x_{i+1}) = \min_{0 \leq \lambda \leq \sigma_i \|\nabla f(x_i)\|} f(x_i + \lambda s_i).$$

Metoda 1 poopćava velik broj metoda.

Npr., kod gradijentne metode je

$$s_i = -\frac{\nabla f(x_i)}{\|\nabla f(x_i)\|}.$$

Budući da je je

$$\langle \nabla f(x_i), s_i \rangle = \left\langle \nabla f(x_i), -\frac{\nabla f(x_i)}{\|\nabla f(x_i)\|} \right\rangle = -\|\nabla f(x_i)\|,$$

uvijek je zadovoljen uvjet

$$\langle \nabla f(x_i), -s_i \rangle \geq 1 \cdot \|\nabla f(x_i)\|.$$

Možemo uzeti $\gamma_i = 1$.

Parametar σ_i kontrolira interval po kojem minimiziramo funkciju

$$g_i(\lambda) := f(x_i + \lambda s_i).$$

- Traženje minimuma po neograničenom intervalu ($\lambda \in [0, \infty)$) je često zahtjevno, ponekad i nemoguće.
- Jednostavniije: na segmentu $[0, \sigma_i \|\nabla f(x_i)\|]$
- σ_i definiramo proizvoljno.
- Veći σ_i - manja (bolja) vrijednost funkcije $f(x_i + \lambda s_i)$ ali i veći interval(dulji postupak - bilo analitički, bilo numerički).
- Manji σ_i - manji interval i brža jednodimenzionalnu minimizaciju u i-tom koraku, ali veća vrijednost funkcije f u novoj točki x_{i+1}.

Konvergenciju niza $(x_i)_i$ dobivenovog metodom 1 opisuje sljedeći teorem.

Teorem

Neka je $f : \mathbb{R}^n \to \mathbb{R}$ *i neka je* $x_0 \in \mathbb{R}^n$ *izabran da vrijedi*

(a) $K := \{x \mid f(x) \leq f(x_0)\}$ *je kompaktan,*

(b) f *je neprekidno diferencijabilna na nekom otvorenom skupu koji sadrži* K*.*

Tada za svaki niz $(x_i)_i$ *definiran Metodom 1 vrijedi*

(a) $x_i \in K$ *za sve* $i = 0, 1, 2, \ldots$

i $(x_i)_i$ *ima barem jedno gomilište* $\bar{x}$ *u* K*.*

(b) *Svako gomilište niza* $(x_i)_i$ *je stacionarna točka od* f*;* $\nabla f(\bar{x}) = 0$*.*

Dokaz.

1. Gomilište. Iz definicije niza $(x_i)_i$ direktno slijedi da je niz $(f(x_i))_i$ monoton:

$$\cdots \leq f(x_{i+1}) \leq f(x_i) \leq \cdots \leq f(x_1) \leq f(x_0),$$

te posebno vrijedi $f(x_i) \leq f(x_0)$, tj. $x_i \in K$.

K kompaktan $\Rightarrow$ $(x_i)_i$ ima barem jedno gomilište u K.

Prva tvrdnja dokazana.

2. Stacionarna točka. Neka je $(x_i)_i$ niz u K i $\bar{x}$ njegovo gomilište.

BSO $\bar{x} = \lim_i x_i$ (tj. uzimamo konvergentan podniz).

Pretpostavit ćemo suprotno:

$$\nabla f(\bar{x}) \neq 0.$$

Skica. Neka je

$$\gamma := \inf_i \gamma_i > 0 \quad \text{i} \quad \sigma := \inf_i \sigma_i > 0.$$

Prema Lemi, postoji okolina $U(\bar{x})$ od $\bar{x}$ u kojoj je

$$f(x + \lambda s) \leq f(x) - \lambda \frac{\gamma}{4} \|\nabla f(x)\|$$

(lema) i

$$\|\nabla f(x_i)\| \geq \frac{1}{2} \|\nabla f(\bar{x})\| > 0.$$

(neprekidnost gradijenta) za dovoljno veliki i.

Definira se M

$$M := \min \left\{ \mu, \frac{1}{2} \sigma \|\nabla f(\bar{x})\| \right\}.$$

Sada je

$$
\begin{aligned}
f(x_{i+1}) &= \min_{0\le\lambda\le\mu_i} f(x_i+\lambda s_i) \\
&\le \min_{0\le\lambda\le M} f(x_i+\lambda s_i) \\
&\le f(x_i+M s_i) \\
&\le f(x_i) - M\frac{\gamma}{4}\,\|\nabla f(x_i)\| \\
&\le f(x_i) - M\frac{\gamma}{8}\,\|\nabla f(\bar{x})\|.
\end{aligned}
$$

Ako označimo

$$\varepsilon := M\,\frac{\gamma}{8}\,\|\nabla f(\bar{x})\| > 0.$$

vrijedi $\quad f(x_{i+1}) \le f(x_i) - \varepsilon \quad$ za sve dovoljno velike k.

$\Rightarrow \quad \lim_i f(x_i) = -\infty.$

Kontradikcija s činjenicom da je $(f(x_i))_i$ odozdo ograničen s $f(\bar{x})$. $\square$

- Korak (b) u Metodi 1 je minimizacija funkcije

$$g(\lambda) := f(x_i + \lambda s_i)$$

  na segmentu $[0, \sigma_i \|\nabla f(x_i)\|]$.

- **pretraživanje po pravcu** (engl. 'line search').
- Uvjeti Teorema zahtijevaju pronalaženje egzaktnog minimuma.
- Težak a (često nemoguć) zadatak.
- Numerička jednodimenzionalna minimizacija (npr. metoda zlatnog reza, Brentova metoda) može dati dosta točnu aproksimaciju minimuma.
- Značajnu količina računanja.
- Drugi pristup: 'neegzaktno pretraživanje po pravcu' - konačni postupak minimizacije.

Metoda 2.

(a) Izaberu se brojevi γ_i, σ_i, $i = 0, 1, 2, \ldots$ koji zadovoljavaju

$$\sup_i \gamma_i \leq 1, \quad \inf_i \gamma_i > 0, \quad \inf_i \sigma_i > 0,$$

i početna točka $x_0 \in \mathbb{R}^n$.

(b) Za $i = 0, 1, 2, \ldots$ izračuna se x_{i+1} iz x_i na sljedeći način:

(α) Izabere se proizvoljan $s_i \in D(\gamma_i, x_i)$, definira

$$\rho_i := \sigma_i \|\nabla f(x_i)\|, \quad \phi_i(\lambda) := f(x_i + \lambda s_i),$$

i odredi najmanji cijeli broj $j \geq 0$ takav da je

$$\phi_i(\rho_i 2^{-j}) \leq \phi_i(0) - \rho_i 2^{-j} \, \frac{\gamma_i}{4} \, \|\nabla f(x_i)\|.$$

(β) Odredi se $\bar{i} \in \{0, 1, \ldots, j\}$ takav da je $\phi_i(\rho_i 2^{-\bar{i}})$ minimalan i definira

$$x_{i+1} := x_i + \lambda_i s_i$$

gdje je $\lambda_i = \rho_i 2^{-\bar{i}}$.

- Uočimo da j iz Metode 2 postoji.
- Metoda 2 je konvergentna kao i Metoda 1:

Teorem

Pod pretpostavkom Teorema za Metodu 1, svaki niz $(x_i)_i$ *dobiven Metodom 2 zadovoljava tvrdnje Teorema za Metodu 1.*

Konvergencija modificirane Newtonove metode

U modificiranoj Newtonovoj metodi smjer silaska s_i zadan je s

$$\bar{s}_i = -[\nabla^2 f(x_i)]^{-1} \nabla f(x_i), \quad s_i := \frac{\bar{s}_i}{\|\bar{s}_i\|}.$$

Radi jednostavnosti, označimo

$$H := [\nabla^2 f(x_i)]^{-1} \quad \text{i} \quad g := \nabla f(x_i),$$

pa je

$$s_i = -\frac{Hg}{\|Hg\|}.$$

Pokušajmo odrediti γ_i za koji je

$$\langle \nabla f(x_i), s_i \rangle \geq \gamma_i \|\nabla f(x_i)\|,$$

tj.

$$\langle g, Hg \rangle \geq \gamma_i \|g\| \, \|Hg\|.$$

Ako je $f \in C^2$ tada je $\nabla^2 f(x_i)$ simetrična matrica, te vrijedi

$$\lambda_{\max}\|x\|^2 \geq \langle \nabla^2 f(x_i)x, x\rangle \geq \lambda_{\min}\|x\|^2, \quad \forall x \in \mathbb{R}^n,$$

$\lambda_{\max}$ - najveća
$\lambda_{\min}$ - najmanja svojstvena vrijednost.

Za euklidsku normu $\| \ \|$ vrijedi:

$$\|\nabla^2 f(x_i)\| = \max\{|\lambda_{\max}|, |\lambda_{\min}|\}.$$

Ako je matrica $\nabla^2 f(x_i)$ pozitivno definitna, onda je

$$\lambda_{\max} > \lambda_{\min} > 0,$$

te je

$$\|\nabla^2 f(x_i)\| = \lambda_{\max}.$$

U okolini minimuma je $\nabla^2 f(x_i)$ pozitivno definitna.

λ svojstvena vrijednost od $\nabla^2 f(x_i)$
$\Rightarrow 1/\lambda$ svojstvena vrijednost od $[\nabla^2 f(x_i)]^{-1}$.

Sada je

$$\|[\nabla^2 f(x_i)]^{-1}\| = \frac{1}{\lambda_{\min}}.$$

$\rightarrow \quad H = [\nabla^2 f(x_i)]^{-1}$ zadovoljava:

$$\frac{1}{\lambda_{\min}} \|x\|^2 \geq \langle Hx, x\rangle \geq \frac{1}{\lambda_{\max}} \|x\|^2, \quad \forall x \in \mathbb{R}^n.$$

Nadalje vrijedi

$$\|Hx\| \leq \|H\| \, \|x\|.$$

Iskoristivši ove nejednakosti, dobivamo

$$\langle g, Hg\rangle \geq \frac{1}{\lambda_{\max}} \|g\|^2 \quad \text{i} \quad \gamma_i \|g\| \, \|Hg\| \leq \gamma_i \frac{1}{\lambda_{\min}} \|g\|^2.$$

Ukoliko je zadovoljeno

$$\frac{1}{\lambda_{\max}} \geq \gamma_i \frac{1}{\lambda_{\min}},$$

tada je

$$\langle g, Hg\rangle \geq \frac{1}{\lambda_{\max}} \|g\|^2 \geq \gamma_i \frac{1}{\lambda_{\min}} \|g\|^2 \geq \gamma_i \|g\| \, \|Hg\|.$$

Prvi uvjet možemo zapisati u obliku

$$\gamma_i \leq \frac{\lambda_{\min}}{\lambda_{\max}} = \frac{1}{\|\nabla^2 f(x_i)\| \, \|[\nabla^2 f(x_i)]^{-1}\|} = \frac{1}{\mathrm{cond}(\nabla^2 f(x_i))}.$$

Ako je matrica singularna, tada je

$$\mathrm{cond}(\nabla^2 f(x_i)) = \infty \quad \text{i} \quad \gamma_i = 0$$

te nemamo zadovoljen uvjet konvergencije.

Ako je $\nabla^2 f$ regularna u minimumu $\bar{x}$
- (onda je automatski i pozitivno definitna),
tada možemo staviti

$$\gamma_i = \frac{1}{\mathrm{cond}(\nabla^2 f(x_i))}$$

te je zbog regularnosti

$$\inf_i \gamma_i = \frac{1}{\sup_i \mathrm{cond}(\nabla^2 f(x_i))} > 0.$$

- Modificirana Newtonova metoda je konvergentna ako je f konveksna ($\nabla^2 f$ pozitivno definitna).
- Jači uvjet od npr. uvjeta konvergencije gradijentne metode.
- Metoda brže konvergira ukoliko su ispunjeni uvjeti konvergencije.

Primjer.

Ilustracija DFP metode, BFGS metode i gradijentne metode.

minimiziramo

$$f(x,y) = 100\left[y^3(3-x) - x^2(3+x)\right]^2 + \frac{(2+x)^2}{1+(2+x)?^2}.$$

Minimum se postiže u

$$\bar{x} = 2 \quad \text{i} \quad \bar{y} = 0.8942719099\ldots, \qquad f(\bar{x},\bar{y}) = 0,$$

Startne vrijednosti:

$$x_0 = 0.1, \qquad y_0 = 4.2$$

i

$$H_0 = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Za DFP i BFGS metodu.

Korištena je ista metoda za pretraživanje po pravcu i isti kriterij zaustavljanja eps $= 10^{-11}$.

	BFGS	DFP	Gradijentna m.
N	54	47	201
F	374	568	1248
ε	$\leq 10^{-11}$	$\leq 10^{-11}$	0.7

N - broj iteracija
F - broj poziva funkcije f.
$\varepsilon = \|\nabla f(x_N), y_N)\|$