

Univerzitet u Beogradu
Matematički fakultet

Master rad

Iterativne nelinearne metode najmanjih kvadrata

Autor : Jagoda Ljuboja-Rađen

Mentor : Prof. dr Đorđe Dugošija

Beograd 2013.

Uvodna reč

Problem najmanjih kvadrata je računski problem, prvo bitno proistekao iz potrebe da se matematički model uklopi u dobijena zapažanja. Što je više zapažanja dostupno, to je moguće tačnije odrediti parametre modela. Može se dokazati, da je rešenje koje minimizira određenu sumu kvadrata rezidula, na neki način optimalno. Smatra se da je Gauss (Carl Friedrich Gauss) otkrio metodu najmanjih kvadrata 1795. godine, u svojoj 18. godini. Razvoj modernih numeričkih metoda za rešavanje problema najmanjih kvadrata odigrao se u drugoj polovini šezdesetih godina prošlog veka. Razvoj ove grupe problema je pratio i razvoj QR dekompozicije iz Householerove transformacije (Alston Scott Householder, 1965. godine). Implicitni QR algoritam za izračunavanje dekompozicije singularnih vrednosti (SVD) su razvili Kahan, Golub i Wilkinson otprilike u isto vreme, a završna verzija algoritma je objavljena 1970. (Gene Golub, William Kahan, James Wilkinson) godine. Od tada su ove matrične dekompozicije razvijene i dovedene do visokog nivoa sofisticiranosti. Osamdesetih i devedesetih godina prošlog veka je došlo do velikog napretka kod metoda za opšte i modifikovane probleme najmanjih kvadrata, kao i kod metoda za rešavanje problema retkih matrica (matrice uglavnom ispunjene nulama, eng. "large sparse problems"). Metode za rešavanje potpunih problema najmanjih kvadrata su takođe sistematski razvijane.

Primenu metoda najmanjih kvadrata, koja je od izuzetne važnosti, nalazimo u oblastima statistike, geodezije, optike, obrade signala i druge (kako praktičnu tako i israživačku). Veliki problemi najmanjih kvadrata se danas rutinski rešavaju i to uglavnom zbog velikog povećanja mogućnosti za automatsko prikupljanje podataka.

Rad pred Vama, obrađuje najpoznatije, odnosno najviše korišćene iterativne nelinearne metode najmanjih kvadrata. Sastoji se iz pet poglavlja. U prvom poglavlju je predstavljen problem koji se rešava metodama najmanjih kvadrata. U drugom poglavlju su date definicije osnovnih pojmoveva optimizacije, a u trećem linearan pristup metodama najmanjih kvadrata. U četvrtom poglavlju su predstavljene metode za rešavanje nelinearnih problema najmanjih kvadrata iterativnim putem. Ovde su korišćena dva pristupa. Jedan je pristup linijskog pretraživanja i to u Newtonovoj (Isaac Newton) i Gauss-Newtonovoj metodi. A drugi je pristup oblasti poverenja i to u Levenberg-Marquardtovoj metodi (Kenneth Levenberg, Donald Marquardt). Prilikom predstavljanja metoda, tvrđenja koja direktno učestvuju u dokazivanju konvergencije metoda su dokazana. Pojedina tvrđenja su bez dokaza a razlog tome je to što su u pitanju ili osnovna tvrđenja, čiji se dokazi mogu pronaći u svakoj literaturi koja obrađuje optimizacione probleme ili što bi se dokazivanjem tih tvrđenja mnogo udaljili od glavne teme. Kod tih tvrđenja je data informacija gde se dokaz može pronaći. U petom poglavlju je prikazana praktična primena prikazanih metoda i to na primeru razvoja veštačkih neuronskih mreža.

Ovom prilikom se zahvaljujem mentoru profesoru dr Đorđu Dugošiji na smernicama i sugestijama koje su mi pomogle u izradi ovog rada, kao i kolegama u Republičkom zavodu za statistiku i svojoj porodici na podršci i razumevanju.

Sadržaj:

1. Prikaz problema
1.1 Primer
2. Osnovni pojmovi
2.1 Bezuslovna optimizacija (neophodni i dovoljni uslovi)
2.2 Lagrangeova funkcija
2.3 Osnovni principi optimizacionih metoda
3. Aproksimacija funkcija diskretnom metodom najmanjih kvadrata
3.1 Opšti problem aproksimacije
3.2 Diskretna metoda najmanjih kvadrata
 3.2.1 Linearni problemi i linearizacija
 3.2.2 Matrična formulacija linearnog problema najmanjih kvadrata
 3.2.3 Karakterizacija rešenja
 3.2.4 Numeričko rešavanje problema
4. Iterativne metode nelinearnih problema najmanjih kvadrata
4.1 Postavka problema
4.2 Newtonova i modifikovana Newtonova metoda
4.3 Gauss-Newtonova metoda
4.4 Levenberg-Marquartova metoda
 4.4.1 Metode oblasti poverenja
 4.4.2 Analiza Levenberg-Marquardtovе metode
5. Primena metoda najmanjih kvadrata
5.1 Osnove neuronskih mreža
 5.1.1 Biološke osnove neuronskih mreža
 5.1.2 Klasifikacija neuronskih mreža
5.2 Algoritmi učenja neuronskih mreža
 5.2.1 Osnovni pojmovi
 5.2.2 Nerekurzivni algoritmi učenja neuronskih mreža
 5.2.3 Newtonove metode
 5.2.4 Levenberg-Marquardtova metoda
 5.2.5 Primena neuronskih mreža u OCR tehnologiji
6. Literatura

1.Prikaz problema

1.1. Primer

Hteli bismo na primer da proučavamo efekat lekova na pacijenta. Označimo vreme uzimanja uzoraka krvi, nakon uzimanja određene doze leka i merimo njihovu koncentraciju u svakom uzorku. Označimo sa t_i vremenske promenljive, a y_i koncentraciju leka u svakom uzorku.

Hteli bismo da konstruišemo model koji ukazuje na koncentraciju leka, kao funkciju po vremenu, birajući parametre modela, tako da se vrednosti funkcije što više poklapaju sa izmerenim vrednostima uzoraka krvi.

Izaberemo sledeću funkciju, kao model :

$$(1.1.1) \quad \phi(x, t) = x_1 + tx_2 + e^{-x_3 t}$$

Ovde su x_1, x_2, x_3 i t realni brojevi, ordinata t predstavlja vreme, dok su x_i parametri modela. Predviđena koncentracija u vremenu t je data modelom $\phi(x, t)$. Razlika između predviđajućeg modela i praktično dobijenih vrednosti su „povezani“ sledećom funkcijom najmanjih kvadrata :

$$(1.1.2) \quad \frac{1}{2} \sum_{i=1}^m [y_i - \phi(x, t_i)]^2 = \frac{1}{2} \sum_{i=1}^m r_i^2(x)$$

gde su

$$(1.1.3) \quad r_i(x) = y_i - \phi(x, t_i), \quad i = 1, \dots, m.$$

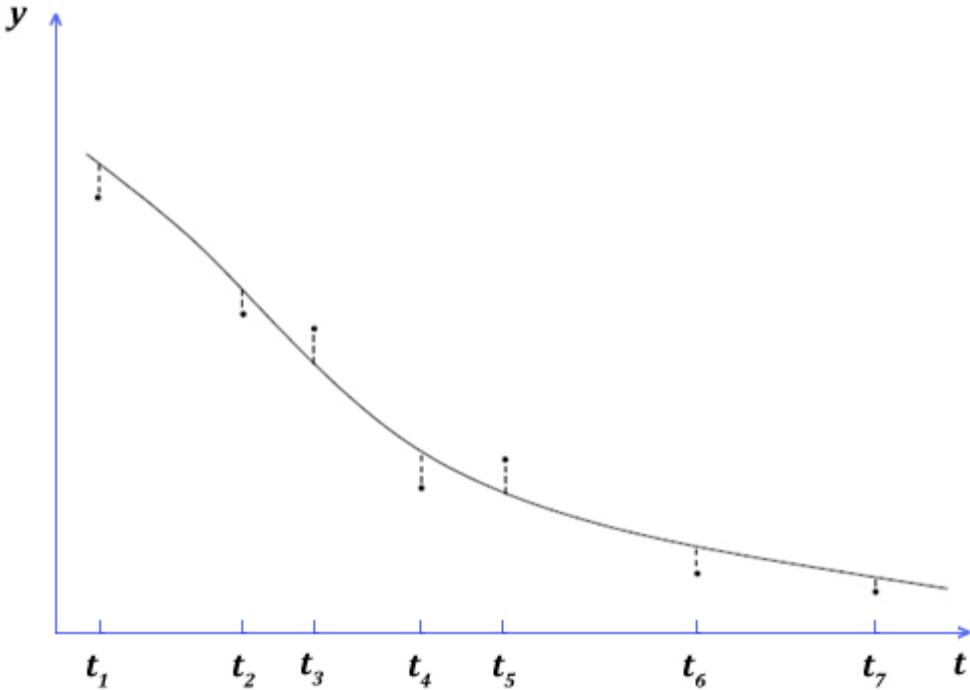
Grafički, svaki član u (1.1.2) predstavlja kvadrat vertikalnog razmaka između krive $\phi(x, t)$, nactrana kao funkcija po t , i tačke (t_i, y_i) . Prikazano na slici (1.1).

Biramo minimum x^* problema najmanjih kvadrata, kao najbolju procenu parametara i koristimo $\phi(x^*, t)$ za procenu koncentracije preostale u krvotoku pacijenta u bilo kom trenutku.

Predpostavlja se da je vreme, koje je tu prikazano, dato sa visokom tačnošću, dok koncentracije y_i , mogu sadržati manje-više nasumične greške, usled ograničenih mogućnosti opreme (ili laboratorijskih tehničara).

Generalno, u problemima aproksimacije, definisanih na ovakav način, , ordinata t , u modelu $\phi(x, t)$, može biti vektor, umesto skalara.(U gornjem primeru bi t moglo sadržati jednu dimenziju za svakog pacijenta.)

Funkcija suma kvadrata (1.1.2) nije jedini način merenja razlike između modela i izmerenih podataka. Umesto toga, mogli bi koristiti na primer šesti stepen sume $\sum_{i=1}^m [y_i - \phi(x, t_i)]^6$ ili suma apsolutnih vrednosti $\sum_{i=1}^m |y_i - \phi(x, t_i)|$ ili neku komplikovaniju funkciju. U stvari, postoje konkretni *statistički razlozi* zbog čega se baš bira određeni kriterijum aproksimacije.



Slika 1.1

Odstupanje modela (1.1.2), glatke krive, od izmerenih vrednosti je prikazano isprekidanim vertikalnim linijama.

2. Uvodni pojmovi

2.1 Bezuslovna optimizacija (neophodni i dovoljni uslovi)

Problem najmanjih kvadrata je deo šire klase problema optimizacije a to je problem bezuslovne optimizacije

$$(2.1.1) \quad \min f(x), \quad x \in R^n.$$

Zbog toga u ovom odeljku navodimo neophodne i dovoljne uslove postojanja minimuma.

Definicija 2.1.1 Tačka x^* se naziva lokalnim minimumom ako postoji $\delta > 0$ takvo da za svako $x \in R^n$ koje ispunjava $f(x^*) \leq f(x)$, važi da je $\|x - x^*\| < \delta$.

Tačka x^* se naziva strogim lokalnim minimumom ako postoji $\delta > 0$ takvo da za svako $x \in R^n$ koje ispunjava $f(x^*) < f(x)$, važi da je $x \neq x^*$ i $\|x - x^*\| < \delta$.

Definicija 2.1.2 Tačka x^* se naziva globalnim minimumom ako je $f(x^*) \leq f(x)$, za svako $x \in R^n$. Tačka x^* se naziva strogim globalnim minimumom ako je $f(x^*) < f(x)$, za svako $x \in R^n$ i $x \neq x^*$.

Definicija 2.1.3 Neka je $f : R^n \rightarrow R$ diferencijabilna za $x \in R^n$ i neka je $\langle \cdot, \cdot \rangle$ skalarni proizvod definisan na R^n . Ako postoji vektor $d \in R^n$ takav da

$$(2.1.2) \quad \langle \nabla f(x), d \rangle < 0,$$

tada d nazivamo opadajućim smerom f u x .

Iz Tejlorovog razvoja

$$f(x_k + td) = f(x_k) + t \nabla f(x_k)^T d + o(t),$$

primećujemo da postoji $\delta > 0$ takvo da je $f(x_k + td) < f(x_k)$, za svako $t \in (0, \delta)$ ako i samo ako je d opadajući smer f u x_k .

Teorema 2.1.4 (Neophodan uslov prvog reda)

Neka je $f : D \subset R^n \rightarrow R$ neprekidno diferencijabilna na otvorenom skupu D . Ako je $x^* \in D$ lokalni minimum problema (2.1.1), tada je

$$(2.1.3) \quad \nabla f(x^*) = 0.$$

Dokaz se može naći u knjizi [1].

Teorema 2.1.5 (Neophodan uslov drugog reda)

Neka je $f : D \subset R^n \rightarrow R$ dva puta neprekidno diferencijabilna na otvorenom skupu D . Ako je $x^* \in D$ lokalni minimum problema (2.1.1), tada je $\nabla f(x^*) = 0$ i $\nabla^2 f(x^*)$ je pozitivno semidefinitna.

Dokaz se može naći u knjizi [1].

Teorema 2.1.6 (Dovoljan uslov drugog reda)

Neka je $f : D \subset R^n \rightarrow R$ dva puta neprekidno diferencijabilna na otvorenom skupu D . Ako je $\nabla f(x^*) = 0$ i $\nabla^2 f(x^*)$ pozitivno definitna, tada je $x^* \in D$ strogi lokalni minimum.

Dokaz se može naći u knjizi [1].

Definicija 2.1.7 Za tačku $x \in R^n$ kažemo da je *stacionarna* (ili kritična) tačka za diferencijabilnu funkciju f ako je $\nabla f(x^*) = 0$.

Teorema (2.1.4) kaže da ako je x^* lokalni minimum, ona je stacionarna tačka. Ali obrnuto ne važi. Ako je x^* stacionarna tačka, onda je moguće da ona bude lokalni minimum ili maksimum, a takođe je moguće da i ne bude ekstremna tačka. Ako stacionarna tačka, nije ni minimum ni maksimum, onda kažemo da je *sedlasta* tačka. Znači, stacionarna tačka ne mora da bude lokalni minimum. Ali ako je posmatrana funkcija, koja je diferencijabilna, konveksna, njene stacionarne tačke jesu lokalni minimumi i globalni minimumi.

Definicija 2.1.8 (*Lipšic neprekidna funkcija*) $F : D \subset R^n \rightarrow R^m$ je Helder neprekidna funkcija na otvorenom skupu D ako postoji konstanta $\gamma \geq 0$ i $p \in (0,1]$ takvi da za svaki $x, y \in D$,

$$\|F(y) - F(x)\| \leq \gamma \|y - x\|^p.$$

Ako je $p = 1$, onda se F naziva Lipšic neprekidnom na D , a γ Lipšicovom konstantom.

Teorema 2.1.9 Neka je $F : R^n \rightarrow R^m$ neprekidno diferencijabilna na otvorenom konveksnom skupu $D \subset R^n$. Neka je F' Lipšic neprekidna u $x \in D$. Tada za svako $x + d \in D$ važi da je

$$\|F(x + d) - F(x) - F'(x)d\| \leq \frac{\gamma}{2} \|d\|^2.$$

Dokaz se može naći u knjizi [1].

Definicija 2.1.10 (*Armijovo i Goldsteinovo pravilo*) Neka je $f : R^n \rightarrow R$ glatka funkcija i $g_k = \nabla f(x_k)$. Tada za date $\beta \in (0,1), \rho \in (0, \frac{1}{2}), \tau > 0$, postoji najmanje nenegativan ceo m_k , takav da je

$$f(x_k) - f(x_k + \beta^m \tau d_k) \geq -\rho \beta^m \tau g_k^T d_k.$$

Ako uzmemo interval $J = \{\alpha > 0 \mid f(x_k + \alpha d_k) < f(x_k)\}$. Da bi posmatrana funkcija sigurno bila dovoljno opadajuća, želimo da odaberemo takvo α koje je daleko od krajnjih tačaka intervala J . Dva uslova za to su:

$$\begin{aligned} f(x_k + \alpha d_k) &\leq f(x_k) + \rho \alpha g_k^T d_k, \\ f(x_k + \alpha d_k) &\geq f(x_k) + (1 - \rho) \alpha g_k^T d_k \end{aligned}$$

i njima se isljučuju tačke u blizini desnog i levog kraja intervala. Ovi uslovi se nazivaju Goldsteinovo pravilo.

2.2 Lagrangeova funkcija

Teorija nelinearnog programiranja, podrazumeva i funkcije sa ograničenjima. Ponekad dokazivanje trvđenja u oblasti bezuslovne optimizacije zahteva i upotrebu tvrđenja i algoritama iz oblasti optimizacije funkcija sa ograničenjima. Takav slučaj je i sa Lagrangeovom funkcijom i trvđenjima koji je prate.

Definišimo problem:

$$(2.2.1) \quad \begin{aligned} &\min f(x), \quad x \in R^n \\ &\text{pri ograničenjima } c_i(x) = 0, \quad i \in E \\ &\quad c_i(x) \geq 0, \quad i \in I, \end{aligned}$$

pri čemu su f i c_i glatke na R^n i čije vrednosti pripadaju nekom podskupu na R^n , a E i I su konačni skupovi.

Možemo definisati i dopustiv skup tačaka

$$(2.2.2) \quad \Omega = \{x \mid c_i(x) = 0, i \in E; c_i(x) \geq 0, i \in I\},$$

pa problem (2.2.1) možemo elegantnije zapisati kao

$$(2.2.3) \quad \min_{x \in \Omega} f(x).$$

Definicija 2.2.1

Za skup $A(x)$ kažemo da je *aktivan* za svaku dopustivu tačku x , ako se sastoji od indeksa ograničenja jednakosti iz skupa E , zajedno sa indeksima ograničenja nejednakosti i , za koje je $c_i(x) = 0$ tj.

$$A(x) = E \cup \{i \in I \mid c_i(x) = 0\}.$$

U dopustivoj tački x , za ograničenje nejednakosti $i \in I$, se kaže da je aktivno ako je $c_i(x) = 0$ i neaktivno ako je ispunjena stroga nejednakost $c_i(x) > 0$.

Definicija 2.2.2 (uslov linearne nezavisnosti)

Neka je x data tačka i $A(x)$ aktivni skup. Kažemo da važi uslov linearne nezavisne ograničenosti, ako je skup izvoda aktivnih ograničenja $\{\nabla c_i(x), i \in A(x)\}$ linearno nezavisan.

Za problem (2.2.1) definišemo Lagrangeovu funkciju

$$(2.2.4) \quad L(x, \lambda) = f(x) - \sum_{i \in E \cup I} \lambda_i c_i(x).$$

Neophodni uslovi definisani u sledećoj teoremi su takozvani uslovi prvog reda (odnose se na osobine prvih izvoda funkcije cilja i funkcija ograničenja). Ovi uslovi su osnove mnogih algoritama.

Teorema 2.2.1 (Neophodni uslovi prvog reda)

Neka je x^* lokalno rešenje problema (2.2.1) i f i c_i u (2.2.1) neprekidno diferencijabilne i neka važe uslovi iz definicije (2.2.2) za x^* . Tada postoji Lagrangeov množilac, vektor λ^* , čiji su članovi λ_i^* , $i \in E \cup I$, takav da su sledeći uslovi ispunjeni za (x^*, λ^*)

$$(2.2.5) \quad \Delta_x L(x^*, \lambda^*) = 0$$

$$(2.2.6) \quad c_i(x^*) = 0, \quad \forall i \in E$$

$$(2.2.7) \quad c_i(x^*) \geq 0, \quad \forall i \in I$$

$$(2.2.8) \quad \lambda_i^* \geq 0, \quad \forall i \in I$$

$$(2.2.9) \quad \lambda_i^* c_i(x^*) = 0, \quad \forall i \in E \cup I.$$

Dokaz se može naći u knjizi [3].

2.3 Osnovni principi optimizacionih metoda

Optimizacione metode koje se ovde razmatraju su iterativne i cilj im je naći minimum optimizacionog problema. Osnovna ideja je, da je data početna tačka $x_0 \in R^n$ i generiše se iterativni niz $\{x_k\}$, po nekim iterativnim pravilima, takav da kada je $\{x_k\}$ konačan niz, poslednja tačka je optimalno rešenje; a kada je $\{x_k\}$ beskonačan, onda on ima graničnu tačku koja je optimalno rešenje. Uglavnom se za prihvatljiv algoritam smatra onaj kod koga se iteracija x_k kreće polako ka okolini minuma x^* , a onda brzo kovergira ka tački x^* . Kada je ispunjeno pravilo konvergencije, iteracija se prekida. Generalno, najprirodnije pravilo prekida je

$$(2.3.1) \quad \|\nabla f(x_k)\| \leq \delta ,$$

gde je δ propisana tolerancija. Ako je (2.3.1) ispunjeno, to ukazuje da gradijent vektor $\nabla f(x_k)$ teži nuli, a iterativni niz $\{x_k\}$ konvergira ka stacionarnoj tački.

Neka je x_k k-ta iteracija, d_k k-ti smer kretanja i α_k k-ta dužina koraka. Tada je k-ta iteracija

$$(2.3.2) \quad x_{k+1} = x_k + \alpha_k d_k .$$

Odavde se vidi da različite α_k i d_k formiraju različite metode. Najčešće optimizacione metode su tzv. opadajuće metode, u smislu da f u svakoj iteraciji ispunjava

$$(2.3.3) \quad f(x_{k+1}) = f(x_k + \alpha_k d_k) < f(x_k) ,$$

pri čemu je d_k opadajući smer određen definicijom (2.1.3).

Osnovna šema optimizacionih metoda:

Algoritam 2.3.1

Korak 0. (Početni korak) Zadati početnu tačku $x_0 \in R^n$ i toleranciju $\varepsilon > 0$.

Korak 1. (Kriterijum prekida) Ako je $\|\nabla f(x_k)\| < \varepsilon$, stati.

Korak 2. (Naći smer) Prema nekoj iterativnoj šemi, pronaći opadajući smer d_k .

Korak 3. (Linijsko pretraživanje) Odrediti dužinu koraka α_k , takvu da vrednosti posmatrane funkcije opadaju, tj.

$$f(x_k + \alpha_k d_k) < f(x_k)$$

Korak 4.(Petlja) Staviti da je $x_{k+1} = x_k + \alpha_k d_k$, $k := k + 1$ i vratiti se na korak 1.

Brzina konvergencije je jedna od najvažnijih karakteristika metoda optimizacije.

Definicija 2.3.2

Neka je $p \in N$ i neka niz $\{x_k\}$ konvergira ka x^* , kada $k \rightarrow \infty$.

- (1) Ukoliko za dovoljno veliko k važi $\|x_{k+p} - x^*\| \leq q \|x_k - x^*\|$, $0 \leq q < 1$, konvergencija se naziva p-linearnom. Za $p = 1$, kovregencija se naziva linarnom.
- (2) Ukoliko važi $\|x_{k+p} - x^*\| \leq q_k \|x_k - x^*\|$, gde $q_k \rightarrow 0$, kada $k \rightarrow \infty$, konvergencija se naziva p-superlinearnom. Za $p = 1$, kovregencija se naziva superlinarnom.
- (3) Ukoliko važi $\|x_{k+1} - x^*\| \leq C \|x_k - x^*\|^2$, konvergencija se naziva kvadratnom.

Teorema 2.3.3

Neka je $\nabla f(x)$ ravnomerno neprekidna na skupu $L = \{x \in R^n \mid f(x) \leq f(x_0)\}$ i neka je ugao θ_k između $-\nabla f(x)$ i smera d_k generisanog u algoritmu 2.3.1, ravnomerno ograničen daleko od 90° tj. ispunjava

$$(2.3.7) \quad \theta_k \leq \frac{\pi}{2} - \mu, \text{ za neko } \mu > 0.$$

Tada je $\nabla f(x_k) = 0$, za neko k ili $f(x_k) \rightarrow -\infty$ ili $\nabla f(x_k) \rightarrow 0$.

Dokaz se može naći u knjizi [1].

3. Aproksimacija funkcija diskretnom metodom najmanjih kvadrata

3.1. Opšti problem aproksimacije

Problem je sledeći: Imamo poznate neke informacije o funkciji f , definisanoj na nekom skupu $X \subseteq R$ i na osnovu tih informacija želimo f zameniti nekom drugom funkcijom φ na istom skupu, tako da su f i φ bliske u nekom smislu. Skup X je najčešće interval oblika $[a, b]$ (može i neograničen) ili diskretan skup tačaka.

Postoje dva oblika problema aproksimacije:

(a) Poznata je funkcija f (npr. analitički), ali je njena forma prekomplikovana za računanje. U tom slučaju odaberemo neke informacije o f i po nekom kriterijumu odredimo aproksimacijsku funkciju φ . U tom slučaju možemo birati informacije o f koje ćemo koristiti i tada možemo i oceniti grešku dobijene aproksimacije.

(b) Funkcija f nije poznata ali su poznate samo neke informacije o njoj, na primer, vrednosti na nekom skupu tačaka. Aproksimirajuća funkcija φ se određuje iz raspoloživih informacija, koje osim samih podataka, uključuju i očekivani oblik ponašanja podataka tj. funkcije φ . U ovom slučaju se ne može napraviti ocena greške bez dodatnih informacija o nepoznatoj funkciji f .

Varijanta (b) je mnogo češća u praksi. Najčešće se javlja kod merenja raznih veličina, jer osim izmerenih podataka, pokušavamo aproksimirati i podatke koji se nalaze „između“ izmerenih tačaka. Prilikom merenja se javljaju i greške merenja, pa postoje posebne tehnike za ublažavanje tako nastalih grešaka.

Funkcija φ se bira prema prirodi modela, ali tako da bude relativno jednostavna za računanje. Ona obično zavisi od parametara a_k , $k = 0, \dots, m$, koje treba odrediti po nekom kriterijumu,

$$\varphi(x) = \varphi(x; a_0, \dots, a_m).$$

Kad smo zapisali funkciju u ovom obliku, onda kažemo da smo odabrali opšti oblik aproksimacijske funkcije. Oblike aproksimacijskih funkcija možemo (grubo) podeliti u dve grupe: linearne i nelinearne aproksimacijske funkcije.

Opšti oblik linearne aproksimacijske funkcije je

$$\varphi(x) = a_0\varphi_0(x) + a_1\varphi_1(x) + \dots + a_m\varphi_m(x),$$

gde su $\varphi_0, \dots, \varphi_m$ poznate funkcije koje znamo računati. Linearost se ne odnosi na oblik funkcije φ , već na njenu zavisnost od parametara a_k , koje treba odrediti. Prednost ovog oblika aproksimacijske funkcije je u tome što se parametri a_k određuju pomoću sistema linearnih jednačina. Načešće korišćeni oblici linearnih aproksimacijskih funkcija su algebarski polinomi, trigonometrijski polinomi i po delovima polinomi tj. splajn funkcije. Načešće korišćeni oblici nelinearnih aproksimacijskih funkcija su eksponencijalne i racionalne aproksimacije.

Aproksimacijske funkcije se biraju tako da „najbolje“ zadovolje uslove koji im se postavljaju. Najčešći su zahtevi da grafik aproksimacijske funkcije prolazi određenim tačkama tj. da *interpolira* funkciju u tim tačkama ili da je odstupanje aproksimacijske funkcije od polazne u nekom smislu minimalno.

3.2. Diskretna metoda najmanjih kvadrata

Neka je funkcija f zadata na diskretnom skupu tačaka x_0, \dots, x_n kojih je mnogo više nego nepoznatih parametara aproksimacijske funkcije

$$\varphi(x; a_0, \dots, a_m).$$

Funkcija φ se određuje iz uslova da Euklidska norma vektora grešaka u čvorovima aproksimacije bude najmanja moguća tj. tako da minimizujemo S ,

$$S = \sum_{k=0}^n (f(x_k) - \varphi(x_k))^2 \rightarrow \min.$$

Ovu funkciju S (kvadrat euklidske norme vektora greške) interpretiramo kao funkciju nepoznatih parametara $S = S(a_0, \dots, a_m)$. Očigledno je da je uvek $S \geq 0$, bez obzira kakvi su parametri. Dakle, zadatak je minimizovati funkciju S , kao funkciju više promenljivih a_0, \dots, a_m . Ako je S dovoljno glatka, a ovako definisana jeste, neophodan uslov ekstrema je

$$\frac{\partial S}{\partial a_k} = 0, \quad k = 0, \dots, m.$$

To je tzv. *sistem normalnih jednačina*.

3.2.1. Linearni problemi i linearizacija

Ilustrujmo problem na najjednostavnijem primeru, kad je aproksimacijska funkcija pravac.

Primer 3.2.1 Zadate su tačke $(x_0, f_0), \dots, (x_n, f_n)$ koje po diskretnoj metodi najmanjih kvadrata aproksimiramo pravcem

$$\varphi(x) = a_0 + a_1 x.$$

Greška aproksimacije u čvorovima koju minimizujemo je

$$S = S(a_0, a_1) = \sum_{k=0}^n (f_k - \varphi(x_k))^2 = \sum_{k=0}^n (f_k - a_0 - a_1 x_k)^2 \rightarrow \min.$$

Nađimo parcijalne izvode po parametrima a_0 i a_1 :

$$0 = \frac{\partial S}{\partial a_0} = -2 \sum_{k=0}^n (f_k - a_0 - a_1 x_k),$$

$$0 = \frac{\partial S}{\partial a_1} = -2 \sum_{k=0}^n (f_k - a_0 - a_1 x_k) x_k.$$

Deljenjem sa -2 i sređivanjem po nepoznatim a_0 i a_1 , dobija se linearan sistem

$$a_0 (n+1) + a_1 \sum_{k=0}^n x_k = \sum_{k=0}^n f_k$$

$$a_0 \sum_{k=0}^n x_k + a_1 \sum_{k=0}^n x_k^2 = \sum_{k=0}^n f_k x_k.$$

Uvedemo li standardne skraćene označke

$$s_l = \sum_{k=0}^n x_k^l, \quad t_l = \sum_{k=0}^n f_k x_k^l, \quad l > 0,$$

linearni sistem možemo zapisati kao

$$(3.2.1) \quad \begin{aligned} s_0 a_0 + s_1 a_1 &= t_0, \\ s_1 a_0 + s_2 a_1 &= t_1. \end{aligned}$$

Može se pokazati da je matrica sistema regularna, jer je njena determinanta različita od nule. Determinanta matrice sistema je $s_0 s_2 - s_1^2$. Koši-Švarcova nejednakost primenjena na vektore

$$e = [1, 1, \dots, 1]^T \text{ i } x = [x_0, x_1, \dots, x_n]^T,$$

daje $s_0 s_2 \geq s_1^2$, pri čemu jednakost važi samo ako su vektori e i x linearno zavisni. Kako predpostavljamo da imamo barem dve različite apscise polaznih tačaka x_k (prirodan uslov za pravac koji nije paralelan sa ordinatom), nejednakost mora biti stroga tj. važi da je $s_0 s_2 > s_1^2$. Dakle, postoji jedinstveno rešenje sistema. Samo rešenje dobija se rešavanjem sistema (3.2.1).

Ostaje još pitanje da li smo dobili minimum, a to se može pokazati korišćenjem drugih parcijalnih izvoda (dovoljan uslov minimuma je pozitivna definitnost Hesijana). Međutim, može i na lakši način. Kako se radi o zbiru kvadrata, S predstavlja paraboloid s otvorom prema gore u promenljivama a_0 i a_1 , pa je jasno da takvi paraboloidi imaju minimum.

Za funkciju φ bismo mogli uzeti i polinom višeg stepena,

$$\varphi(x) = a_0 + a_1x + \dots + a_m x^m,$$

ali postoji opasnost da je za malo veće m ($m \approx 10$) dobijeni sistem vrlo loše uslovljen (matrica sistema vrlo blizu singularne matrice), pa dobijeni rezultati mogu biti veoma pogrešni. Zbog toga se za prikaz sa polinomima višeg stepena ni ne koristi prikaz polinoma u bazi potencija, već ako se koriste aproksimacije polinomima višeg stepena, koriste ortogonalni polinomi.

Linearni model diskretnih najmanjih kvadrata je potpuno primenljiv na opštu linearnu funkciju

$$\varphi(x) = a_0 \varphi_0(x) + \dots + a_m \varphi_m(x),$$

gde su $\varphi_0, \dots, \varphi_m$ poznate (zadate) funkcije. Sledi ilustracija na opštoj linearnej funkciji sa 2 parametra.

Primer 3.2.2 Zadate su tačke $(x_0, f_0), \dots, (x_n, f_n)$ koje po diskretnoj metodi najmanjih kvadrata aproksimiramo funkcijom oblika $\varphi(x) = a_0 \varphi_0(x) + a_1 \varphi_1(x)$.

Postupak je isti kao u prethodnom primeru. Minimizujemo kvadrat Euklidske norme vektora grešaka aproksimacije u čvorovima

$$S = S(a_0, a_1) = \sum_{k=0}^n (f_k - \varphi(x_k))^2 = \sum_{k=0}^n (f_k - a_0 \varphi_0(x_k) - a_1 \varphi_1(x_k))^2 \rightarrow \min.$$

Sređivanjem parcijalnih izvoda

$$0 = \frac{\partial S}{\partial a_0} = -2 \sum_{k=0}^n (f_k - a_0 \varphi_0(x_k) - a_1 \varphi_1(x_k)) \varphi_0(x_k),$$

$$0 = \frac{\partial S}{\partial a_1} = -2 \sum_{k=0}^n (f_k - a_0 \varphi_0(x_k) - a_1 \varphi_1(x_k)) \varphi_1(x_k),$$

po promenljivama a_0 i a_1 , uz dogovor da je

$$s_0 = \sum_{k=0}^n \varphi_0^2(x_k), \quad s_1 = \sum_{k=0}^n \varphi_0(x_k) \varphi_1(x_k), \quad s_2 = \sum_{k=0}^n \varphi_1^2(x_k),$$

$$t_0 = \sum_{k=0}^n f_k \varphi_0(x_k), \quad t_1 = \sum_{k=0}^n f_k \varphi_1(x_k),$$

dobija se isti oblik linearne sistema

$$s_0 a_0 + s_1 a_1 = t_0,$$

$$s_1 a_0 + s_2 a_1 = t_1.$$

A ovaj sistem ima ista svojstva kao i u prethodnom primeru.

Ako φ nelinearno zavisi od parametara, dobija se nelinearan sistem jednačina, koji se relativno teško rešava. To postaje ozbiljan problem optimizacije, koji se može rešavati metodama pretraživanja ili nekim drugim metodama posebno prilagođenim upravo za rešavanje nelinearnog problema najmanjih kvadrata (npr. Gauss-Newtonova metoda ili Levenberg-Marquartova metoda).

Sledi ilustracija *nelinearnog* problema najmanjih kvadrata na jednostavnom primeru.

Primer 3.2.3 Zadate su tačke $(x_0, f_0), \dots, (x_n, f_n)$ koje po diskretnoj metodi najmanjih kvadrata aproksimiramo funkcijom oblika $\varphi(x) = a_0 e^{a_1 x}$. Greška aproksimacije u čvorovima (funkcija koju minimizujemo) je

$$S = S(a_0, a_1) = \sum_{k=0}^n (f_k - \varphi(x_k))^2 = \sum_{k=0}^n (f_k - a_0 e^{a_1 x_k})^2 \rightarrow \min.$$

Parcijalni izvodi po promenljivama a_0 i a_1 su

$$\begin{aligned} 0 &= \frac{\partial S}{\partial a_0} = -2 \sum_{k=0}^n (f_k - a_0 e^{a_1 x_k}) e^{a_1 x_k}, \\ 0 &= \frac{\partial S}{\partial a_1} = -2 \sum_{k=0}^n (f_k - a_0 e^{a_1 x_k}) a_0 x_k e^{a_1 x_k}, \end{aligned}$$

što je nelinearan sistem jednačina.

S druge strane, ako logaritmujemo relaciju

$$\varphi(x) = a_0 e^{a_1 x},$$

dobija se

$$\ln \varphi(x) = \ln(a_0) + a_1 x.$$

Logaritmujemo još i vrednosti funkcije f u tačakama x_k , pa uz supsituicije

$$h(x) = \ln f(x), \quad h_k = h(x_k) = \ln f_k, \quad k = 0, \dots, n,$$

i

$$\psi(x) = \ln \varphi(x) = b_0 + b_1 x,$$

gde je

$$b_0 = \ln(a_0), \quad b_1 = a_1,$$

dobija se linearni problem najmanjih kvadrata

$$\tilde{S} = \tilde{S}(b_0, b_1) = \sum_{k=0}^n (h_k - \psi(x_k))^2 = \sum_{k=0}^n (h_k - b_0 - b_1 x_k)^2 \rightarrow \min.$$

Na kraju iz rešenja b_0 i b_1 ovog problema, lako očitamo a_0 i a_1

$$a_0 = e^{b_0}, \quad a_1 = b_1.$$

Može se uočiti da ovako dobijeno rešenje uvek daje pozitivan a_0 , tj. linearizacijom dobijena funkcija $\varphi(x)$ će uvek biti veća od 0. Naravno, to nije „pravo“ rešenje za sve početne tačke (x_k, f_k) i ne mogu se na ovakav način linearizovati svi početni podaci, jer mora biti $f_k > 0$, da bismo mogli logaritmovati.

Ipak i kad su neki $f_k \leq 0$, može se, korišćenjem translacije svih podataka dobiti $f_k + translacija > 0$, pa onda nastaviti postupak linearizacije.

Sledi nekoliko funkcija koje su često u upotrebi i njihovih standardnih linearizacija u problemu najmanjih kvadrata.

(a) Funkcija

$$\varphi(x) = a_0 e^{a_1 x}$$

se linearizuje logaritmovanjem

$$\psi(x) = \log \varphi(x) = \log(a_0) + a_1 \log x, \quad h_k = \log f_k, \quad k = 0, \dots, n.$$

Dobija se linearan problem najmanjih kvadrata

$$\tilde{S} = \tilde{S}(b_0, b_1) = \sum_{k=0}^n (h_k - b_0 - b_1 \log(x_k))^2 \rightarrow \min,$$

gde je,

$$b_0 = \log(a_0), \quad b_1 = a_1.$$

Da bi se izvela linearizacija moraju biti i $x_k > 0$ i $f_k > 0$.

(b) Funkcija

$$\varphi(x) = \frac{1}{a_0 + a_1 x}$$

se linearizuje na sledeći način

$$\psi(x) = \frac{1}{\varphi(x)} = a_0 \frac{1}{x} + a_1, \quad h_k = \frac{1}{f_k}, \quad k = 0, \dots, n.$$

Pripadajući linearan problem najmanjih kvadrata je

$$\tilde{S} = \tilde{S}(a_0, a_1) = \sum_{k=0}^n (h_k - a_0 - a_1 x_k)^2 \rightarrow \min.$$

(c) Funkcija

$$\varphi(x) = \frac{x}{a_0 + a_1 x}$$

se može linearizovati na više načina. Prvo, može se staviti da su

$$\psi(x) = \frac{1}{\varphi(x)} = a_0 \frac{1}{x} + a_1, \quad h_k = \frac{1}{f_k}, \quad k = 0, \dots, n.$$

Pripadajući linearan problem najmanjih kvadrata je

$$\tilde{S} = \tilde{S}(a_0, a_1) = \sum_{k=0}^n (h_k - a_0 \frac{1}{x_k} - a_1)^2 \rightarrow \min.$$

A može se koristiti i sledeći način

$$\psi(x) = \frac{x}{\varphi(x)} = a_0 + a_1 x, \quad h_k = \frac{x_k}{f_k}, \quad k = 0, \dots, n,$$

pa je pripadajući linearan problem najmanjih kvadrata

$$\tilde{S} = \tilde{S}(a_0, a_1) = \sum_{k=0}^n (h_k - a_0 - a_1 x_k)^2 \rightarrow \min.$$

(d) Funkcija

$$\varphi(x) = \frac{1}{a_0 + a_1 e^{-x}}$$

se linearizuje stavljanjem

$$\psi(x) = \frac{1}{\varphi(x)} = a_0 + a_1 e^{-x}, \quad h_k = \frac{1}{f_k}, \quad k = 0, \dots, n.$$

Pripadajući linearan problem najmanjih kvadrata je

$$\tilde{S} = \tilde{S}(a_0, a_1) = \sum_{k=0}^n (h_k - a_0 - a_1 e^{-x_k})^2 \rightarrow \min.$$

3.2.2. Matrična formulacija linearog problema najmanjih kvadrata

Da bi se formirao matrični zapis linearog problema najmanjih kvadrata, moraju se preimenovati nepoznate, kako bismo matricu, vektor desne strane i nepoznate u linearom sistemu pisali u uobičajenoj formi (standardno su nepoznate $x_1, x_2, \dots, , a$ ne $a_0, a_1, \dots,).$

Predpostavimo da imamo skup mernih podataka $(t_k, y_k), k = 1, \dots, n$ i želimo taj model aproksimirati funkcijom oblika $\varphi(t)$. Ako je $\varphi(t)$ linearna tj. ako je

$$\varphi(x) = x_1 \varphi_1(t) + \dots + x_m \varphi_m(t),$$

onda treba pronaći parametre x_j tako da merni podaci (t_k, y_k) zadovoljavaju

$$y_k = \sum_{j=1}^m x_j \varphi_j(t_k), \quad k = 1, \dots, n.$$

Ako označimo

$$a_{kj} = \varphi_j(x_k), \quad b_k = y_k,$$

onda prethodne jednačine možemo zapisati u obliku $Ax = b$.

Ako je mernih podataka više nego parametara, tj. ako je $n > m$, onda sistem jednačina ima više jednačina nego nepoznatih, pa je predimenzionisan.

Kao što je već rečeno, postoji mnogo načina da se odredi „najbolje“ rešenje, ali iz statističkih razloga to je često metoda najmanjih kvadrata, tj. određujemo x tako da minimizira grešku $r = Ax - b$ (za r kažemo da je rezidual).

$$(3.2.2) \quad \min_x \|r\|_2 = \min_x \|Ax - b\|_2, \quad b \in R^n, \quad x \in R^m.$$

Ako je $\text{rang}(A) < m$, onda rešenje x ovog problema nije jedinstveno, jer mu možemo dodati bilo koji vektor iz nul-prostora od A , a da se rezidual ne promeni. Sa druge strane, među svim rešenjima x problema najmanjih kvadrata uvek postoji jedinstveno rešenje x najmanje norme tj. koje još minimizira $\|x\|_2$.

3.2.3 Karakterizacija rešenja

Prvo, okarakterišimo skup svih rešenja problema najmanjih kvadrata.

Teorema 3.2.4 Skup svih rešenja problema najmanjih kvadrata (3.2.2) označimo sa

$$S = \{x \in \mathbb{R}^m \mid \|Ax - b\|_2 = \min\}.$$

Tada je $x \in S$ ako i samo ako važi sledeća relacija ortogonalnosti

$$(3.2.3) \quad A^T(b - Ax) = 0.$$

Dokaz: Predpostavimo da \hat{x} zadovoljava relaciju ortogonalnosti $A^T \hat{r} = 0$, $\hat{r} = b - A\hat{x}$. Tada za bilo koji $x \in \mathbb{R}^m$ imamo

$$r = b - Ax = \hat{r} + A\hat{x} - Ax = \hat{r} - A(x - \hat{x}).$$

Ako označimo sa $e = x - \hat{x}$, onda je

$$\begin{aligned} \|r\|_2^2 &= r^T r = (\hat{r} - Ae)^T (\hat{r} - Ae) \\ &= \hat{r}^T \hat{r} + \|Ae\|_2^2 = \|\hat{r}\|_2^2 + \|Ae\|_2^2. \end{aligned}$$

Kako je $\|Ae\|_2 \geq 0$, vidimo da je minimalna vrednost za $\|r\|_2$ jednaka $\|\hat{r}\|_2$, pa $x = \hat{x}$ minimizira $\|r\|_2$. Time je pokazano da je $\hat{x} \in S$.

Pokažimo obrnutu implikaciju kontradikcijom. Predpostavimo da je

$$A^T \hat{r} = z \neq 0 \text{ i uzmimo } x = \hat{x} + \varepsilon z.$$

Tada je

$$r = \hat{r} - \varepsilon Az$$

i

$$\|r\|_2^2 = r^T r = \hat{r}^T \hat{r} - 2\varepsilon z^T z + \varepsilon^2 (Az)^T (Az) < \hat{r}^T \hat{r},$$

za dovoljno mali ε , pa \hat{x} nije rešenje u smislu najmanjih kvadrata, što znači da nije u S .

□

Relacija (3.2.3) često se naziva *sistem normalnih jednačina* i piše se u obliku
(3.2.4) $A^T Ax = A^T b$.

Matrica $A^T A$ je simetrična i pozitivno semidefinitna, a sistem normalnih jednačina je uvek saglasan (tj. ima bar jedno rešenje), jer je

$$A^T b \in \mathcal{R}(A^T) = \mathcal{R}(A^T A).$$

Važi i sledeće tvrđenje:

Tvrđenje 3.2.5 Matrica $A^T A$ je pozitivno definitna ako i samo ako su kolone matrice A linearno nezavisne tj. $\text{rang}(A) = m$.

Dokaz: Ako su kolone matrice $A = [a_1, \dots, a_m]$ linearne nezavisne, tada za svaki $x \neq 0$ važi (prema definiciji linearne nezavisnosti)

$$Ax = x_1 a_1 + \cdots + x_m a_m \neq 0,$$

pri čemu su x_j komponente od x . Za takav x je

$$x^T A^T Ax = \|Ax\|_2^2 > 0,$$

tj. $A^T A$ je pozitivno definitna.

Sa druge strane, ako su kolone linearne zavisne, onda postoji $x_0 \neq 0$ takav da je $Ax_0 = 0$, pa za takav x_0 važi $x_0^T A^T Ax_0 = 0$. Ako je x takav da je $Ax \neq 0$, onda je $x^T A^T Ax \geq 0$, pa je $A^T A$ pozitivno semidefinitna. Drugim rečima, $A^T A$ je uopšteno pozitivno semidefinitna, a pozitivnu definitnost obezbeđuje tek pun rang kolone matrice A .

□

Iz prethodnog sledi da uslov $\text{rang}(A) = m$, osigurava postojanje jedinstvenog rešenja problema najmanjih kvadrata. U tom slučaju rešenje i pripadni rezidual zadovoljavaju

$$x = (A^T A)^{-1} A^T b, \quad r = b - A(A^T A)^{-1} A^T b.$$

Ako je $S \subset \mathbb{R}^n$ potprostor, onda je $P_s \in \mathbb{R}^{n \times n}$ ortogonalni projektor na S , ako je $\mathcal{R}(P_s) = S$ i

$$P_s^2 = P_s, \quad P_s^T = P_s.$$

Dalje, važi i

$$(I - P_s)^2 = I - P_s, \quad (I - P_s)P_s = 0,$$

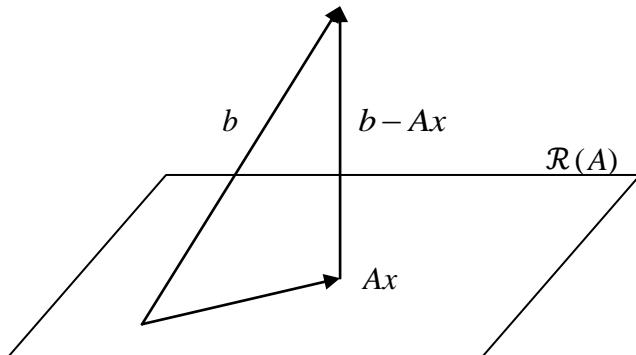
pa je $I - P_s$ projektor na ortogonalni komplement od S .

Tvrđimo da postoji jedinstveni ortogonalni projektor na S . Predpostavimo da postoje dva ortogonalna projektori P_1 i P_2 . Za sve $z \in \mathbb{R}^n$, tada važi

$$\begin{aligned} \|(P_1 - P_2)z\|_2^2 &= z^T (P_1 - P_2)^T (P_1 - P_2)z = z^T (P_1^T P_1 - P_2^T P_1 - P_1^T P_2 + P_2^T P_2)z \\ &= z^T (P_1 - P_2 P_1 - P_1 P_2 + P_2)z \\ &= z^T P_1 (I - P_2)z + z^T P_2 (I - P_1)z = 0. \end{aligned}$$

Odatle sledi da je $P_1 = P_2$, tj. ortogonalni projektor je jedinstven.

Iz geometrijske interpretacije problema najmanjih kvadrata se vidi da je Ax ortogonalna projekcija vektora b na $\mathcal{R}(A)$.



Takođe je $r = (I - P_{R(A)})b$ i u slučaju punog ranga matrice A važi

$$P_{R(A)} = A(A^T A)^{-1} A^T.$$

Ako je $\text{rang}(A) < m$, onda A ima netrivijalni nul-potprostor i rešenje problema najmanjih kvadrata nije jedinstveno. Neka je \hat{x} jedno od rešenja. Skup svih rešenja S onda možemo opisati kao

$$S = \{x = \hat{x} + z \mid z \in \mathcal{N}(A)\}.$$

Ako je $\hat{x} \perp \mathcal{N}(A)$, onda je

$$\|x\|_2^2 = \|\hat{x}\|_2^2 + \|z\|_2^2,$$

pa je \hat{x} jedinstveno rešenje problema najmanjih kvadrata koje ima minimalnu euklidsku normu.

3.2.4 Numeričko rešavanje problema najmanjih kvadrata

Postoji nekoliko načina rešavanja nelinearnih problema najmanjih kvadrata u praksi. Obično se koristi jedna od sledećih metoda:

1. Sistem normalnih jednačina
2. QR faktorizacija
3. Dekompozicija singularnih vrednosti
4. Transformacija u linearan sistem

Sistem normalnih jednačina

Ova metoda je u odnosu na ostale, najbrža ali najmanje tačna. Koristi se kad je $A^T A$ pozitivno definitna i kad je njena uslovljenošć mala. Matrica $A^T A$ se rastavi faktorizacijom Čoleskog, a zatim se reši linearan sistem

$$A^T A x = A^T b.$$

Ukupan broj aritmetičkih operacija za računanje $A^T A$, $A^T b$, a zatim i faktorizaciju Čoleskog je $nm^2 + \frac{1}{3}m^3 + O(m^2)$. Kako je $n \geq m$, onda je prvi član dominantan u ovom izrazu, a potiče od formiranja $A^T A$.

QR faktorizacija

Ponovo predpostavimo da je $A^T A$ pozitivno definitna. Polazimo od rešenja problema najmanjih kvadrata dobijenog iz sistema normalnih jednačina

$$x = (A^T A)^{-1} A^T b.$$

Zatim napišemo QR faktorizaciju matrice A

$$A = QR = Q_0 R_0,$$

gde je Q_0 ortogonalna matrica tipa (n, m) , a R_0 trouglasta tipa (m, m) i uvrstimo u rešenje. Dobija se

$$\begin{aligned} x &= (A^T A)^{-1} A^T b = (R_0^T Q_0^T Q_0 R_0)^{-1} R_0^T Q_0^T b \\ &= (R_0^T R_0)^{-1} R_0^T Q_0^T b = R_0^{-1} R_0^{-T} R_0^T Q_0^T b = R_0^{-1} Q_0^T b, \end{aligned}$$

tj. x se dobija primenom „invertovane“ skraćene QR faktorizacije matrice A na b (po analogiji sa rešavanjem linearog sistema, samo što A ne mora imati inverz)

Preciznije, da bi se našlo x , rešava se trouglasti linearan sistem

$$R_0 x = Q_0^T b.$$

Na ovakav način se najčešće rešavaju problemi najmanjih kvadrata. Može se pokazati da je cena računanja $2nm^2 - \frac{2}{3}m^3$, što je dvostruko više nego za sistem normalnih jednačina kad je $n \gg m$, a približno jednako za $n = m$.

QR faktorizacija se može koristiti i za problem najmanjih kvadrata kad matrica A nema pun rang kolone, ali tada se koristi QR faktorizacija sa pivotiranjem kolone (na prvo mesto se dovodi kolona čiji „radni deo“ ima najveću normu). Jer, ako matrica ima rang $r < m$, onda njena QR faktorizacija ima oblik

$$A = QR = Q \begin{bmatrix} R_{11} & R_{12} \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

gde je R_{11} nesingularna reda r , a R_{12} neka $r \times (m-r)$ matrica. Zbog grešaka zaokruživanja, umesto pravog R , izračunamo

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \\ 0 & 0 \end{bmatrix}.$$

Naravno, želeli bismo da je $\|R_{22}\|_2$ vrlo mala, reda veličine $\varepsilon \|A\|_2$, pa da je možemo „zaboraviti“ tj. staviti $R_{22} = 0$ i tako odrediti rang matrice A . Ali to nije uvek tako. Na primer, bidijagonalna matrica

$$A = \begin{bmatrix} \frac{1}{2} & 1 & & & \\ 2 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & & \frac{1}{2} \end{bmatrix}$$

je skoro singularna ($\det(A) = 2^{-2}$), njena QR faktorizacija je $Q = I$, $R = A$ i nema nijednog R_{22} koji bi bio po normi mali. Zbog toga koristimo pivotiranje, koje R_{11} pokušava držati što bolje usloveljnim, a R_{22} po normi što manjim.

Dekompozicija singularnih vrednosti

Smatra se da je jedna od najkorisnijih dekompozicija i sa teorijske strane (za dokazivanje činjenica) i sa praktične strane dekompozicija singularnih vrednosti ili skraćeno SVD (na engleskom „singular value decomposition“).

Teorema 3.2.6 Ako A ima pun rang, onda je rešenje problema najmanjih kvadrata

$$\min_x \|Ax - b\|_2$$

jednako

$$x = V \Sigma^{-1} U^T b,$$

tj. dobija se primenom „invertovanog“ skraćenog SVD-a od A na b .

Dokaz: Važi

$$\|Ax - b\|_2^2 = \|U \Sigma V^T x - b\|_2^2.$$

Kako je A punog ranga to je i Σ . Zbog unitarne ekvivalencije euklidske norme, važi

$$\begin{aligned} \|U \Sigma V^T x - b\|_2^2 &= \|\hat{U}^T (U \Sigma V^T x - b)\|_2^2 = \left\| \begin{bmatrix} U^T \\ U_0^T \end{bmatrix} (U \Sigma V^T x - b) \right\|_2^2 \\ &= \left\| \begin{bmatrix} \Sigma V^T x - Ub \\ -U_0^T b \end{bmatrix} \right\|_2^2 = \|\Sigma V^T x - U^T b\|_2^2 + \|U_0^T b\|_2^2. \end{aligned}$$

Prethodni izraz se minimizira ako je prvi član jednak 0 tj. ako je $x = V \Sigma^{-1} U^T b$.

Dobija se i da je vrednost minimuma $\min_x \|Ax - b\|_2 = \|U_0^T b\|_2$.

□

Primetimo da je u prethodnoj teoremi, rešenje dobijeno, kad je matrica A punog ranga. Uobičajeno je da se SVD primenjuje u metodi najmanjih kvadrata i kad matrica A nema pun rang kolone. Rešenja su istog oblika, samo što moramo znati „izračunati“ inverz matrice Σ kad ona nije regularna, tj. kad ima neke nule na dijagonali. Takav inverz zove se generalizovani inverz i označava se sa Σ^+ . U slučaju da je

$$\Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix}$$

pri čemu je Σ_1 regularna, onda je

$$\Sigma^+ = \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Još preciznije, za problem najmanjih kvadrata važi sledeće tvrđenje:

Tvrđenje 3.2.7 Neka matrica A ima rang $r < m$. Rešenje x koje minimizira $\|Ax - b\|_2$ može se okarakterisati na sledeći način. Neka je $A = U \Sigma V^T$ SVD matrice A i neka je

$$A = U \Sigma V^T = [U_1, U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} [V_1, V_2]^T = U_1 \Sigma_1 V_1^T,$$

gde je Σ_1 nesingularna, reda r , a matrice U_1 i V_1 imaju r kolona. Neka je

$$\sigma := \sigma_{\min}(\Sigma_1)$$

najmanja ne-nula singularna vrednost od A . Tada se sva rešenja problema najmanjih kvadrata mogu napisati u formi

$$x = V_1 \Sigma_1^{-1} U_1^T b + V_2 z,$$

gde je z proizvoljan vektor. Rešenje x koje ima minimalnu euklidsku normu je ono za koje je $z = 0$, tj.

$$x = V_1 \Sigma_1^{-1} U_1^T b$$

i važi ocena

$$\|x\|_2 \leq \frac{\|b\|_2}{\sigma}.$$

Dokaz: Dopunimo matricu $[U_1, U_2]$ kolonama matrice U_3 do ortogonalne matrice reda n i označimo je sa \hat{U} . Korišćenjem unitarne invarijantnosti euklidske norme, dobija se

$$\begin{aligned} \|Ax - b\|_2^2 &= \|\hat{U}(Ax - b)\|_2^2 = \left\| \begin{bmatrix} U_1^T \\ U_2^T \\ U_3^T \end{bmatrix} (U_1 \Sigma_1 V_1^T x - b) \right\|_2^2 \\ &= \left\| \begin{bmatrix} \Sigma_1 V_1^T x - U_1^T b \\ -U_2^T b \\ -U_3^T b \end{bmatrix} \right\|_2^2 = \|\Sigma_1 V_1^T x - U_1^T b\|_2^2 + \|U_2^T b\|_2^2 + \|U_3^T b\|_2^2. \end{aligned}$$

Izraz je minimiziran kad je prva od tri norme u poslednjem redu jednaka 0, tj. ako je

$$\Sigma_1 V_1^T x = U_1^T b$$

ili

$$x = V_1 \Sigma_1^{-1} U_1^T b.$$

Kolone matrica V_1 i V_2 su međusobno ortogonalne, pa je $V_1^T V_2 z = 0$ za sve vektore z . Odavde se vidi da x ostaje rešenje problema najmanjih kvadrata i kad mu dodamo $V_2 z$, za bilo koji z , tj. ako je

$$x = V_1 \Sigma_1^{-1} U_1^T b + V_2 z.$$

To su ujedno i sva rešenja, jer kolone matrice V_2 razapinju nul-potprostor $\mathcal{N}(A)$. Osim toga, zbog spomenute ortogonalnosti važi i

$$\|x\|_2^2 = \|V_1 \Sigma_1^{-1} U_1^T b\|_2^2 + \|V_2 z\|_2^2,$$

a to je minimalno za $z = 0$. Na kraju, za to minimalno rešenje važi ocena

$$\|x\|_2 = \|V_1 \Sigma_1^{-1} U_1^T b\|_2 = \|\Sigma_1^{-1} U_1^T b\|_2 \leq \frac{\|U_1^T b\|_2}{\sigma} = \frac{\|b\|_2}{\sigma}.$$

□

Rešenje problema najmanjih kvadrata korišćenjem SVD-a je najstabilnije, a može se pokazati da je za $n \gg m$, njegovo trajanje približno jednako kao i trajanje rešenja korišćenjem QR-a. Za manje n , trajanje je približno $4nm^2 + \frac{4}{3}m^3 + O(m^2)$.

Transformisanje problema najmanjih kvadrata u linearan sistem

Ako matrica A ima pun rang po kolonama, onda problem najmanjih kvadrata možemo transformisati i u linearan sistem različit od sistema normalnih jednačina.

Simetričan linearan sistem

$$\begin{bmatrix} I & A \\ A^T & 0 \end{bmatrix} \begin{bmatrix} r \\ x \end{bmatrix} = \begin{bmatrix} b \\ 0 \end{bmatrix},$$

ekvivalentan je sistemu normalnih jednačina. Ako napišemo prvu i drugu blok komponentu

$$r + Ax = b, \quad A^T r = 0,$$

onda uvrštavanjem r -a iz prve blok jednačine u drugu dobijamo sistem

$$A^T(b - Ax) = 0.$$

Prvi sistem ima bitno manji raspon elemenata od sistema normalnih jednačina. Osim toga, ako je matrica loše uslovljena, kod tog sistema možemo lakše koristiti iterativno profinjavanje rešenja.

4. Iterativne metode za rešavanje nelinearnih problema najmanjih kvadrata

4.1 Postavka problema

Potrebno je naći globalni minimum sume kvadrata m nelinearnih funkcija,

$$(4.1.1) \quad \min f(x), \quad f(x) = \frac{1}{2} \sum_{i=1}^m r_i^2(x) = \frac{1}{2} r(x)^T r(x), \quad m > n, \quad x \in R^n.$$

Ovde je svaka $r_i(x)$, $i = 1, \dots, m$ nelinearna funkcija na R^n . Naravno, važi da ako bi svi $r_i(x)$ bili linearni po x , tada bi (4.1.1) bio *linearan* problem najmanjih kvadrata. Problem se može predstaviti i kao rešavanje sistema m nelinearnih jednačina

$$(4.1.2) \quad r_i(x) = 0, \quad i = 1, \dots, m,$$

gde su $r_i(x)$ rezidualne funkcije.

Osnovne metode za rešavanje nelinearnih problema najmanjih kvadrata zahtevaju informacije o izvodima komponenti $r_i(x)$. U daljem tekstu predpostavljamo da su $r_i(x)$ dva puta neprekidno diferencijabilne.

Jakobijan rezidualnog vektora $r(x) = (r_1(x), \dots, r_m(x))^T$ je

$$(4.1.3) \quad J(x) = \frac{\partial r_i(x)}{\partial x_j}, \quad i = 1, \dots, m, \quad j = 1, \dots, n,$$

a Hesijan po $r_i(x)$ je

$$(4.1.4) \quad G_i(x) = \nabla^2 r_i(x) \in R^{n \times n}, \quad G_i(x)_{jk} = \frac{\partial^2 r_i(x)}{\partial x_j \partial x_k}, \quad i = 1, \dots, m.$$

Prvi i drugi izvod funkcije $f(x)$ su:

$$(4.1.5) \quad \nabla f(x) = \sum_{i=1}^m r_i(x) \nabla r_i(x) = J(x)^T r(x) = g(x)$$

$$(4.1.6) \quad \begin{aligned} \nabla^2 f(x) &= \sum_{i=1}^m (\nabla r_i(x) \nabla r_i(x)^T + r_i(x) \nabla^2 r_i(x)) \\ &= J(x)^T J(x) + S(x) = H(x), \end{aligned}$$

pri čemu je

$$(4.1.7) \quad S(x) = \sum_{i=1}^m r_i(x) \nabla^2 r_i(x).$$

4.2 Newtonova i modifikovana Newtonova metoda

Uzimamo da je kvadratni model, odnosno kvadratna aproksimacija $q^{(k)}(x)$, koja se dobija pomoću Tejlorovog razvoja funkcije $f(x)$ u okolini tačke x_k .

$$\begin{aligned} (4.2.1) \quad q^{(k)}(x) &= f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2} (x - x_k)^T H(x_k) (x - x_k) \\ &= \frac{1}{2} r(x_k)^T r(x_k) + (J(x_k)^T + r(x_k))^T (x - x_k) \\ &\quad + \frac{1}{2} (x - x_k)^T (J(x_k)^T J(x_k) + S(x_k)) (x - x_k). \end{aligned}$$

Diferenciranjem $q^{(k)}(x)$ i izjednačavanjem njenog prvog izvoda sa nulom, pod predpostavkom da je $H(x_k)$ pozitivno definitna matrica, dobićemo tačku x_{k+1} koja minimizira funkciju $q^{(k)}(x)$:

$$0 = \nabla q^{(k)}(x) = \nabla f(x_k) + H(x_k)(x - x_k).$$

A odavde imamo da je $x_{k+1} = x_k - H(x_k)^{-1} \nabla f(x_k)$, pa je

$$(4.2.2) \quad s_k = x_{k+1} - x_k = -H(x_k)^{-1} \nabla f(x_k).$$

Sada možemo da definišemo Newtonovu metodu za problem (4.1.1) tj. iterativni proces:

$$x_{k+1} = x_k - H(x_k)^{-1} \nabla f(x_k), \quad k = 0, 1, \dots$$

tj.

$$(4.2.3) \quad x_{k+1} = x_k - (J(x_k)^T J(x_k) + S(x_k))^{-1} J(x_k)^T r(x_k).$$

Za smer s_k (za koji kažemo da je Newtonov smer) važi da je opadajući, jer uz uslov da je $H(x_k)$ pozitivno definitna matrica, imamo da je

$$\langle \nabla f(x), s_k \rangle = s_k^T \nabla f(x_k) = -\nabla f(x_k)^T H(x_k)^{-1} \nabla f(x_k) < 0.$$

Algoritam 4.2.1 (Newtonova metoda)

Korak 0. Zadati $x_0 \in R^n$, $\varepsilon > 0$, $k := 0$

Korak 1. Ako je $\|\nabla f(x_k)\| < \varepsilon$ stati, inače preći na korak 2.

Korak 2. Rešiti $H(x_k)s_k = -\nabla f(x_k)$, po s_k .

Korak 3. Staviti da je $x_{k+1} = x_k + s_k$, $k := k + 1$ i vratiti se na korak 1.

Teorema 4.2.2 (Teorema o konvergenciji Newtonove metode)

Neka je $f \in C^2$ i neka je x_k dovoljno blizu rešenja minimizacijskog problema x^* , pri čemu je $\nabla f(x^*) = 0$. Ako je Hesijan $H(x_k)$ pozitivno definitan, a $H(x)$ ispunjava Lipšicov uslov, $|H_{ij}(x) - H_{ij}(y)| \leq \beta \|x - y\|$, za neko β i sve i, j , pri čemu su $H_{ij}(x)$ (i, j -ti element $H(x)$), onda je Newtonova iteracija $x_{k+1} = x_k - H(x_k)^{-1}\nabla f(x_k)$ dobro definisana. Generisani niz $\{x_k\}$ kvadratno konvergira ka x^* .

Dokaz:

Neka je $h_k = x_k - x^*$. Iz Tejlorove formule imamo da je

$$0 = \nabla f(x^*) = \nabla f(x_k) - H(x_k)h_k + O(\|h_k\|^2).$$

Kako je $f \in C^2$, x_k je dovoljno blizu x^* , a $H(x^*)$ je pozitivno definitan, logično je pretpostaviti da je x_k u okolini x^* , da je $H(x_k)$ pozitivno definitan i $H(x_k)^{-1}$ ograničen odozgo. Stoga k -ta Newtonova iteracija postoji. Množeći sa $H(x_k)^{-1}$ imamo da je

$$\begin{aligned} 0 &= H(x_k)^{-1}\nabla f(x_k) - h_k + O(\|h_k\|^2) \\ &= -s_k - h_k + O(\|h_k\|^2) \\ &= -h_{k+1} + O(\|h_k\|^2). \end{aligned}$$

Iz definicije $O(\cdot)$, imamo da postoji konstanta C , takva da je

$$(4.2.4) \quad \|h_{k+1}\| \leq C\|h_k\|^2.$$

Ako $x_k \in \Omega = \{x \mid \|h\| \leq \gamma/C, h = x - x^*, \gamma \in (0, 1)\}$, tada je

$$(4.2.5) \quad \|h_{k+1}\| \leq \gamma\|h_k\| \leq \gamma^2/C < \gamma/C.$$

Stoga je $x_{k+1} \in \Omega$, pa indukcijom po k imamo da je Newtonova iteracija dobro definisana, za sve k i $\|h_k\| \rightarrow 0$, kad $k \rightarrow \infty$, pa iterativni niz konvergira. Takođe (4.2.4) pokazuje da ovde imamo kvadratnu konvergenciju.

□

Primećuje se da je Newtonova metoda, lokalna metoda. Ako je početna tačka suviše daleko od rešenja, nije sigurno da je Hesijan pozitivno definitan, ni da je Newtonov smer opadajući, pa konvergencija nije zagarantovana.

Kako bi obezbedili konvergenciju, iteracijski postupak (4.2.2) modifikujemo, stavljajući da je

$$(4.2.6) \quad \alpha_k d_k = x_{k+1} - x_k = -\alpha_k H(x_k)^{-1} \nabla f(x_k),$$

gde je α_k dužina koraka.

Odatle sledi algoritam modifikovane Newtonove metode.

Algoritam 4.2.3 (Modifikovana Newtonova metoda)

Neka je f dva puta neprekidno diferencijabilna, sa pozitivno definitnim Hesijanom.

Korak 0. Zadati $x_0 \in R^n$, $\varepsilon > 0$, $k := 0$.

Korak 1. Izračunati $\nabla f(x_k)$. Ako je $\|\nabla f(x_k)\| < \varepsilon$ stati, inače ići na korak 2.

Korak 2. Izračunati $H(x_k)d_k = -\nabla f(x_k)$, po d_k .

Korak 3. Naći α_k , takvo da je $f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k)$.

Korak 4. Staviti da je $x_{k+1} = x_k + \alpha_k d_k$, $k := k + 1$ i vratiti se na korak 1.

Teorema 4.2.4 (Konvergencija modifikovane Newtonove metode)

Neka je $f \in R^n \rightarrow R$ dva puta neprekidno diferencijabilna na otvrenom konveksnom skupu $D \in R^n$. Predpostavimo da za bilo koju tačku $x_0 \in D$, postoji konstanta $m > 0$, takva da $f(x)$ ispunjava

$$(4.2.7) \quad u^T \nabla^2 f(x)u \geq m \|u\|^2, \text{ za svako } u \in R^n \text{ i } x \in L(x_0),$$

gde je $L(x_0) = \{x | f(x) \leq f(x_0)\}$ nivo skup.

Tada za niz $\{x_k\}$, generisan algoritmom 4.2.3 važi:

(1) Ako je $\{x_k\}$ konačan, za neko k će biti $\nabla f(x_k) = 0$.

(2) Ako je $\{x_k\}$ beskonačan, konvergiraće ka jedinstvenom minimumu x^* .

Dokaz:

Prvo, iz (4.2.7) imamo da je $f(x)$ strogo konveksna funkcija na R^n , pa je stacionarna tačka jedinstveni globalni minimum. Takođe, iz predpostavki imamo da je skup $L(x_0)$ ograničen, zatvoren, konveksan skup. Kako je niz $\{f(x_k)\}$ monotono opadajući, imamo da je $\{x_k\} \subset L(x_0)$ i $\{x_k\}$ je ograničen. Stoga, postoji granična tačka $\bar{x} \in L(x_0)$, takva da $x_k \rightarrow \bar{x}$ i $f(x_k) \rightarrow f(\bar{x})$. Takođe, kako je $f \in C^2(D)$, iz teoreme 2.3.3, imamo da $\nabla f(x_k) \rightarrow \nabla f(\bar{x}) = 0$. Na kraju, uz napomenu da je stacionarna tačka jedinstvena, imamo da ceo niz $\{x_k\}$ konvergira ka \bar{x} , koja je jedinstveni minimum.

□

Ali, bez obzira na modifikaciju, glavni nedostatak, gore navedene Newtonove metode je da je član $S(x)$, Hesijana $H(x)$ teško ili skupo izračunati. Zbog toga bi, sa ciljem smanjivanja obima računa, bilo razumno i efikasno, da ili zanemarimo $S(x)$ ili da koristimo prvi izvod za aproksimaciju $S(x)$.

Primetimo iz (4.2.4), da kada se $r_i(x)$ približi nuli ili se približi linearnej funkciji, tada se $\nabla^2 r_i(x)$ približava nuli, a $S(x)$ je malo i može se zanemariti. Time je definisan mali rezidualni problem. U suprotnom imamo veliki rezidualni problem.

4.3 Gauss-Newtonova metoda

Ova metoda se bazira na nizu linearnih aproksimacija reziduala $r(x)$. Ako u kvadratnom modelu (4.2.1), zanemarimo član $S(x)$ u $H(x)$, onda (4.2.1) postaje

$$(4.3.1) \quad \begin{aligned} \bar{q}^{(k)}(x) = & \frac{1}{2} r(x_k)^T r(x_k) + (J(x_k)^T + r(x_k))^T (x - x_k) \\ & + \frac{1}{2} (x - x_k)^T (J(x_k)^T J(x_k)) (x - x_k), \end{aligned}$$

a (4.2.2) postaje

$$(4.3.2) \quad x_{k+1} = x_k + s_k = x_k - (J(x_k)^T J(x_k))^{-1} J(x_k)^T r(x_k).$$

Algoritam (4.3.1) (Gauss-Newtonova metoda)

Korak 0: Dati su $x_0, \varepsilon > 0, k := 0$

Korak 1: Ako je $\|\nabla f(x)\| < \varepsilon$ stati, inače preći na korak 2.

Korak 2: Rešiti

$$(4.3.3) \quad J(x_k)^T J(x_k) s_k = -J(x_k)^T r(x_k) \text{ po } s_k.$$

Korak 3: Staviti

$$(4.3.4) \quad x_{k+1} = x_k + s_k ; \quad k := k + 1 \text{ i ići na korak 1.}$$

Očigledno, kad $J(x_k)$ ima pun rang i gradijent $\nabla f(x)$ je nenula, smer kretanja s_k je opadajući za f , zato što je

$$\begin{aligned} s_k^T \nabla f(x_k) &= s_k^T J(x_k)^T r(x_k) = -s_k^T J(x_k)^T J(x_k) s_k \\ &= -(J(x_k) s_k)^T J(x_k) s_k = -\|J(x_k) s_k\|^2 \leq 0. \end{aligned}$$

Poslednja nejednakost je uvek stroga, sem kada je $J(x_k) s_k = 0$, što je ekvivalentno sa $J(x_k)^T r(x_k) = \nabla f(x_k) = 0$.

Jednačina (4.3.2) se naziva Gauss-Newtonovom jednačinom. Poredeći (4.2.3) i (4.3.2), vidimo da je razlika između Gauss-Newtonove i Newtonove metode, u korišćenju $J(x_k)^T J(x_k)$ umesto Hesijana $H(x_k)$, odnosno Gauss-Newtonova metoda zahteva samo računanje prvog izvoda, dok Newtonova zahteva i drugi.

Primetimo da je drugi korak Algoritma 4.3.1 , samo analogan normalnoj jednačini linearogn problema najmanjih kvadrata (3.2.4). Pored toga, model tj. aproksimacija $\bar{q}^{(k)}(x)$ (4.3.1) je ekvivalentna , s obzirom na model $r(x)$ u blizini x_k ,

$$(4.3.5) \quad \bar{M}_k = r(x_k) + J(x_k)(x - x_k)$$

i rešava se linearan problem najmanjih kvadrata

$$(4.3.6) \quad \min \frac{1}{2} \|\bar{M}_k(x)\|^2.$$

Ova dva zapažanja daju da je Gauss-Newtonova metoda, ustvari linearizujuća metoda za nelinearan problem najmanjih kvadrata. Kako je Newtonova metoda, pod standardnim predpostavkama , lokalno i kvadratno konvergentna , uspeh Gauss-Newtonove metode će zavisiti od važnosti zanemarenog člana $S(x)$ u $H(x)$.

Sledeća teorema pokazuje da:

1. Ako je $S(x^*)=0$, onda je Gauss-Newtonova metoda kvadratno konvergentna;
2. Ako je $S(x^*)$ malo u odnosu na $J(x^*)^T J(x^*)$, onda je Gauss-Newtonova metoda Q-linearno konvergentna;
3. Ako je $S(x^*)$ previše veliko , Gauss-Newtonova metoda neće biti konvergentna.

Teorema 4.3.2

Neka je $f : R^n \rightarrow R$ i $f \in C^2$. Predpostavimo da je x^* lokalni minimum nelinearnog problema najmanjih kvadrata (4.1.1) i $J(x^*)^T J(x^*)$ pozitivno definitna. Predpostavimo i da je niz $\{x_k\}$ generisan algoritmom (4.4.1) i konvergira ka x^* . Tada ako je $(J(x)^T J(x))^{-1}$ Lipšic neprekidna u okolini x^* , postoje $K > 0$ i $\delta > 0$, takvi da za $x_k \in \mathcal{B}(\delta)$ važi

$$(4.3.7) \quad \|x_{k+1} - x^*\| \leq K (\|x_k - x^*\|^2 + \|r(x^*)\| \|x_k - x^*\|).$$

Dokaz:

Uzmimo δ dovoljno malo, takvo da za $\|x - x^*\| < \delta$, $J(x)^T J(x)$ je regularna. Neka je γ Lipšic konstanta za J . Iz (4.3.3) imamo da važi:

$$s_k = -(J(x_k)^T J(x_k))^{-1} J(x_k)^T r(x_k),$$

pa je

$$(4.3.8) \quad \begin{aligned} x_{k+1} - x^* &= x_k - x^* - (J(x_k)^T J(x_k))^{-1} J(x_k)^T r(x_k) \\ &= (J(x_k)^T J(x_k))^{-1} J(x_k)^T (J(x_k)(x_k - x^*) - r(x_k)). \end{aligned}$$

Primetimo da je

$$\begin{aligned} J(x_k)(x_k - x^*) - r(x_k) &= J(x_k)(x_k - x^*) - r(x^*) + r(x^*) - r(x_k) \\ &= -r(x^*) + (J(x_k)(x_k - x^*) + r(x^*) - r(x_k)), \end{aligned}$$

a iz Teoreme 2.1.9 imamo da je

$$\|J(x_k)(x_k - x^*) + r(x^*) - r(x_k)\| \leq \gamma \frac{\|x_k - x^*\|}{2}.$$

A kako je $J(x^*)^T r(x^*) = 0$, onda je $-J(x_k)^T r(x^*) = (J(x^*) - J(x_k))^T r(x^*) = 0$.

Sada, normirajući dobijamo

$$\begin{aligned} \|x_{k+1} - x^*\| &\leq \|(J(x_k)^T J(x_k))^{-1}\| \|(J(x^*) - J(x_k))^T r(x^*)\| \\ (4.3.9) \quad &+ \frac{\|(J(x_k)^T J(x_k))^{-1}\| \|J(x_k)^T\| \gamma \|x_k - x^*\|^2}{2} \\ &\leq \|(J(x_k)^T J(x_k))^{-1}\| \gamma \|x_k - x^*\| \left(\frac{\|r(x^*)\| + \|J(x_k)^T\| \|x_k - x^*\|}{2} \right). \end{aligned}$$

Stavljujući da je

$$K = \gamma \max_{x \in B(\delta)} \left(\|(J(x)^T J(x))^{-1}\| \left(\frac{1 + \|J(x)^T\|}{2} \right) \right)$$

dobijamo (4.3.7).

□

Posledica ove teoreme je da je za nula rezidual probleme lokalna konvergencija q-kvadratna jer član $\|r(x^*)\| \|x_k - x^*\|$ na desnoj strani (4.3.7) nestaje. Za ostale probleme, ni q-linearna konvergencija nije zagarantovana. U stvari, ako je $x_k \in B(\delta)$ onda (4.3.8) implicira $\|x_{k+1} - x^*\| \leq r \|x_k - x^*\|$, za $0 < r < 1$ ako je

$$(4.3.10) \quad K(\delta + \|J(x^*)\|) \leq r.$$

Stoga će q-faktor biti $K \|J(x^*)\|$. Pa imamo da za male rezidualne probleme i precizne početne podatke, konvergencija Gauss-Newtonove metode će biti brza, ali možda neće konvergirati za velike rezidualne probleme.

Iz izraza (4.3.9) dobijamo još jedan problem, kada se $(J(x^*) - J(x_k))^T r(x^*)$ posmatra kao ceo, a ne kao aproksimacija sa $\gamma \|x_k - x^*\| \|r(x^*)\|$. Koristeći Tejlorovu teoremu i neophodne uslove minimuma ($J(x^*)^T J(x^*) = 0$) dobija se:

$$\begin{aligned} J(x_k)^T r(x^*) &= \left[J(x^*)^T + \nabla J(x^*)(x_k - x^*) + O(\|x_k - x^*\|^2) \right]^T r(x^*) \\ &= (x_k - x^*)^T \nabla J(x^*)^T r(x^*) + O(\|x_k - x^*\|^2). \end{aligned}$$

Podsetimo se da je

$$\nabla J(x^*)^T r(x^*) = \nabla^2 f(x^*) - J(x^*)^T r(x^*),$$

pa važi

$$(4.3.11) \quad \|(J(x^*) - J(x_k))^T r(x^*)\| \leq \|\nabla^2 f(x) - J(x^*)^T\| \|r(x^*)\| + O(\|x_k - x^*\|^2).$$

Odatle imamo, da čak i za velike rezidualne probleme, konvergencija može biti brza, ako problem nije jako nelinearan (mali ∇J).

U specijalnom slučaju linearnih problema najmanjih kvadrata (gde je $\nabla J = 0$) Gauss-Newton postaje rešenje normalnih jednačina i konvergira u jednoj iteraciji.

Primer 4.3.3:

Neka su $r_1(x) = x + 1$, $r_2(x) = \lambda x^2 + x - 1$. Razmotrimo

$$\min f(x), f(x) = \frac{1}{2} \sum_{i=1}^2 r_i^2(x) = \frac{1}{2}(x+1)^2 + \frac{1}{2}(\lambda x^2 + x - 1)^2$$

gde su $n = 1$, $m = 2$.

Imamo da je $\nabla f(x) = 2\lambda^2 x^3 + 3\lambda x^2 - 2(\lambda - 1)x$, pa je $x = 0$ stacionarna tačka.

Imamo i da je $\nabla^2 f(x) = 6\lambda^2 x^2 + 6\lambda x - 2(\lambda - 1)$, što pokazuje da ako je $\lambda < 1$, onda je $\nabla^2 f(0) > 0$, pa će $x = 0$ biti lokalni minimum (u ovom slučaju biće i globalni). Jakobijan matrica je $J(x) = \begin{bmatrix} 1 \\ 2\lambda x + 1 \end{bmatrix}$, i u tom slučaju, iz (4.4.2) imamo da je Gauss-Newtonova iteracija

$$x_{k+1} = \frac{2\lambda^2 x_k^3 + 3\lambda x_k^2 - 2(\lambda - 1)x_k}{2 + 4\lambda x_k + 4\lambda^2 x_k^2}.$$

Za $\lambda = 0.1$, Gauss-Newtonova iteracija ima sledeći rezultat :

k	1	2	3	4	...	20	21
x_k	1.000000	0.868852	0.758504	0.664814	...	0.103807	0.931008

k	22	...	99	100	101	...	106
x_k	0.083529	...	0.000024	0.000022	0.000019	...	0.000017

Može se videti da, kada je $\lambda = 0.1$ stepen nelinearnosti u $r(x)$ je mali i Gauss-Newtonova metoda radi dobro, a kada je $\lambda = 0$, pri čemu je $r(x)$ linearna, tada je $x_1 = 0 = x^*$. Ovo ukazuje da Gauss-Newtonova metoda dobija minimum u jednoj iteraciji. Kada je $\lambda \neq 0$, a x_k u blizini nule imamo

$$x_{k+1} = x_k + (\lambda - 1)x_k + O(x_k^2) = \lambda x_k + O(x_k^2).$$

Kada je λ dovoljno malo, konvergencija je linearna. Kada je $|\lambda| > 1$, Gauss-Newtonova metoda je nekonvergentna. Ovaj primer pokazuje da je Gauss-Newtonova metoda dobra samo kad važi da je x_0 blizu x^* i kad je matrica $S(x^*)$ mala.

4.4. Levenberg - Marquardtova metoda

4.4.1 Metode oblasti poverenja

Problemi koji se rešavaju Newtonovom i Gauss-Newtonovom metodom koriste princip linijskog pretraživanja tj. aproksimira se posmatrana funkcija kvadratnim modelom, kreira se iterativni korak i nađe pogodna veličina koraka duž pravca. Ove metode jesu efikasne, ali se ne mogu koristiti ako Hesijan nije pozitivno definitna matrica tj. ako je početna tačka daleko od tačke u kojoj se dostiže lokalni minimum. Razvijen je i drugačiji princip rešavanja ovakvih problema. To je princip oblasti poverenja.

U metodama oblasti poverenja, prvo se definiše skup

$$\Omega_k = \{x \mid \|x - x_k\| \leq \Delta_k\},$$

gde je Δ_k prečnik Ω_k za koji smatramo da je adekvatan posmatranoj funkciji (funkciji cilja). Zatim se bira korak koji će približno minimizovati kvadratni model u oblasti poverenja tj. takav da je $x_k + s_k$ približno najbolja tačka na sferi

$$\{x_k + s \mid \|s\| \leq \Delta_k\},$$

sa centrom x_k i prečnikom Δ_k . Ukoliko je korak neprihvatljiv, smanji se veličina oblasti poverenja a zatim traži novi minimum. Ove metode zadržavaju brzu lokalnu konvergenciju Newtonovih metoda ali imaju i dobru globalnu konvergenciju.

Model potproblema metode oblasti poverenja je

$$(4.4.1) \quad \begin{aligned} \min q^{(k)}(s) &= f(x_k) + g_k^T s + \frac{1}{2} s^T B_k s \\ \text{pri ograničenju } &\|s\| \leq \Delta_k, \end{aligned}$$

gde je $\Delta_k > 0$ prečnik oblasti poverenja, $g_k = \nabla f(x_k)$, B_k simetrična matrica i aproksimira Hesijan H_k . Uglavnom se koristi Euklidska norma, tako da s_k minimizira $q^{(k)}(s)$ u lopti prečnika Δ_k . I ostale norme mogu biti korišćene, s tim što različite norme određuju različite oblasti poverenja.

Ako u (4.1.1) stavimo da je $B_k = H_k$, dobija se Newtonova metoda kroz oblast poverenja. Ključno pitanje je kako izabrati Δ_k za svaku iteraciju. Generalno važi da, ako postoji dobra usaglašenost između modela $q^{(k)}(s)$ i vrednosti funkcije $f(x_k + s)$, treba izabrati što je moguće veći Δ_k .

Neka je

$$(4.4.2) \quad Ared_k = f(x_k) - f(x_k + s_k).$$

I neka je

$$(4.4.3) \quad Pred_k = q^{(k)}(0) - q^{(k)}(s_k).$$

Za $Ared_k$ kažemo da je *stvarno smanjenje*, a za $Pred_k$ da je *predviđeno smanjenje*.

Definišimo odnos

$$(4.4.4) \quad r_k = \frac{\text{Ar ed}_k}{\text{Pr ed}_k},$$

koji meri usaglašenost između modela $q^{(k)}$ i funkcije f . Odnos r_k igra važnu ulogu u odabiru nove iteracije x_{k+1} , kao i u ažuriranju prečnika oblasti poverenja Δ_k . Ako je r_k blizu 1, onda imamo dobru usaglašenost i možemo proširiti oblast poverenja u sledećoj iteraciji. Ako je r_k blizu 0 ili negativan, sužavamo oblast poverenja.

Algoritam 4.4.1 (Algoritam oblasti poverenja)

Korak 0. Zadati početne tačke $x_0, \bar{\Delta}, \Delta_0 \in (0, \bar{\Delta}), \varepsilon \geq 0, 0 < \eta_1 < \eta_2 < 1,$
 $0 < \gamma_1 < 1 < \gamma_2$ i $k := 0$

Korak 1. Ako je $\|\nabla f(x_k)\| < \varepsilon$ stati, inače preći na korak 2.

Korak 2. Približno rešiti potproblem (4.4.1), po s_k .

Korak 3. Izračunati $f(x_k + s_k)$ i r_k . Staviti da je

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{za } r_k \geq \eta_1 \\ x_k, & \text{inače} \end{cases}$$

Korak 4. Ako je $r_k < \eta_1$, tada $\Delta_{k+1} \in (0, \gamma_1 \Delta_k]$.

Ako je $r_k \in [\eta_1, \eta_2)$, tada $\Delta_{k+1} \in [\gamma_1 \Delta_k, \Delta_k]$.

Ako je $r_k \geq \eta_2$ i $\|s_k\| = \Delta_k$, tada $\Delta_{k+1} \in [\Delta_k, \min\{\gamma_2 \Delta_k, \bar{\Delta}\}]$.

Korak 5. Generisati B_{k+1} , ažurirati $q^{(k)}$, staviti $k := k + 1$ i preći na korak 1.

Ovde imamo da je $\bar{\Delta}$ globalna granica za svako Δ_k . Za iteraciju u kojoj je $r_k \geq \eta_2$ i prema tome u kojoj je $\Delta_{k+1} \geq \Delta_k$, kažemo da je veoma uspešna. Za iteraciju u kojoj je $r_k \geq \eta_1$ i stoga $x_{k+1} = x_k + s_k$, kažemo da je uspešna, a za iteraciju u kojoj je $r_k < \eta_1$ i stoga i $x_{k+1} = x_k$, kažemo da je neuspešna.

Sledi karakterizacija potproblema (4.4.1).

Teorema 4.4.2

Vektor s^* je globalno rešenje problema oblasti poverenja

$$(4.4.5) \quad \min f + g^T s + \frac{1}{2} s^T B s, \\ \text{pri ograničenju } \|s\| \leq \Delta,$$

ako i samo ako je s^* dopustivo (zadovoljava sva postavljena ograničenja) i postoji skalar $\lambda \geq 0$, takav da su sledeći uslovi ispunjeni:

$$(4.4.6) \quad (B + \lambda I)s^* = -g,$$

$$(4.4.7) \quad \lambda(\Delta - \|s^*\|) = 0,$$

$$(4.4.8) \quad (B + \lambda I) \text{ je pozitivno semidefinitna.}$$

Da bi dokazali ovu teoremu, potrebno je izložiti još jedno tvrđenje.

Tvrđenje 4.4.3

Neka je kvadratna funkcija

$$(4.4.9) \quad m(s) = g^T s + \frac{1}{2} s^T B s,$$

gde je B neka simetrična matrica. Tada su sledeća tvrđenja tačna.

(i) m dostiže minimum ako i samo ako je B pozitivno semidefinitna i g je u rangu B . Ako je B pozitivno semidefinitna, tada svaki s koji ispunjava $Bs = -g$ je globalni minimum m .

(ii) m ima globalni minimum ako i samo ako je B pozitivno definitna.

Dokaz:

(i) (\Leftarrow) Ako je g u rangu B , onda postoji s takvo da je $Bs = -g$. Za svako $w \in R^n$, imamo da je

$$\begin{aligned} m(s+w) &= g^T(s+w) + \frac{1}{2}(s+w)^T B(s+w) \\ &= (g^T s + \frac{1}{2} s^T B s) + g^T w + (Bs)^T w + \frac{1}{2} w^T B w \\ &= m(s) + \frac{1}{2} w^T B w \\ &\geq m(s), \end{aligned} \tag{4.4.10}$$

jer je B pozitivno semidefinitna. Zbog toga je s minimum m .

(\Rightarrow) Uzmimo da je s minimum m . Kako je $\nabla m(p) = Bs + g = 0$, imamo da je $\nabla^2 m(p) = B$ pozitivno semidefinitna, što daje dokaz.

(ii) (\Leftarrow) Isto kao kod (i), uz dodatak da je $w^T B w > 0$ kad god je $w \neq 0$.

(\Leftarrow) Kao i u delu (i), zaključujemo da je B pozitivno semidefinitna. A ako B nije pozitivno definitna, tada za vektor $w \neq 0$ imamo da je $Bw = 0$. Stoga je, iz (4.4.10) $m(s+w) = m(s)$, pa minimum nije jedinstven, što daje kontradikciju.

□

Radi ilustracije slučaja (i), predpostavimo da je

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

koja ima sopstvene vrednosti 0,1,2 i zbog toga je singularna. Ako je g bilo koji vektor koji ima drugu komponentu 0, tada bi g bila u rangu B i kvadrat bi dostigao minimum. Ali ako druga komponenta nije 0, možemo sniziti $m(\cdot)$ beskonačno mnogo puta, pomeranjem duž pravca $(0, -g_2, 0)^T$.

Sada možemo dokazati teoremu 4.4.2.

Dokaz teoreme 4.4.2:

(\Leftarrow) Predpostavimo da postoji $\lambda \geq 0$, takvo da su uslovi (4.4.6), (4.4.7) i (4.4.8) ispunjeni. Tvrđenje (4.4.3)(i) ukazuje da je s^* globalni minimum kvadratne funkcije

$$(4.4.11) \quad \hat{m}(s) = g^T s + \frac{1}{2} s^T (B + \lambda I) s = m(s) + \frac{\lambda}{2} s^T s.$$

Kako je $\hat{m}(s) \geq \hat{m}(s^*)$, imamo da je

$$(4.4.12) \quad m(s) \geq m(s^*) + \frac{\lambda}{2} ((s^*)^T s^* - s^T s).$$

Kako je $\lambda(\Delta - \|s\|) = 0$, imamo da je $\lambda(\Delta^2 - (s^*)^T s^*) = 0$, pa je

$$m(s) \geq m(s^*) + \frac{\lambda}{2} (\Delta^2 - s^T s).$$

Kako smo uzeli da je $\lambda \geq 0$, iz prethodne nejednakosti imamo da je $m(s) \geq m(s^*)$, za svako $\|s\| \leq \Delta$. Stoga je s^* globalni minimum problema (4.4.5).

(\Rightarrow) Predpostavimo da je s^* globalni minimum problema (4.4.5) i pokažimo da postoji $\lambda \geq 0$, koje ispunjava (4.4.6)-(4.4.8). U slučaju da je $\|s^*\| < \Delta$, s^* je neograničen minimum funkcije m i važi:

$$\nabla m(s^*) = Bs^* + g = 0, \quad \nabla^2 m(s^*) = B,$$

gde je B je pozitivno semidefinitna, pa uslovi (4.4.6)-(4.4.8) važe za $\lambda = 0$.

Predpostavimo da je $\|s^*\| = \Delta$. Tada je (4.4.7) odmah ispunjeno, a s^* takođe rešava problem sa ograničenjima

$$\min m(s), \text{ pri } \|s\| = \Delta.$$

Prihvatajući optimalne uslove za optimizaciju sa ograničenjima, za ovaj problem (videti odeljak 2.2), imamo da postoji λ , takav da je Lagrangeova funkcija definisana kao:

$$L(s, \lambda) = m(s) + \frac{\lambda}{2} (s^T s - \Delta^2),$$

ima stacionarnu tačku u s^* . Stavljujući da je $\nabla_s L(s^*, \lambda) = 0$, dobija se

$$(4.4.13) \quad Bs^* + g + \lambda s^* = 0 \Rightarrow (B + \lambda I)s^* = -g,$$

pa važi (4.4.6).

Kako je $m(s) \geq m(s^*)$ za bilo koje s , za koje je $s^T s = (s^*)^T s^* = \Delta^2$, imamo takve vektore s da važi

$$m(s) \geq m(s^*) + \frac{\lambda}{2} ((s^*)^T s^* - s^T s).$$

Ako zamenimo izraz za g u (4.4.13), u ovaj izraz, posle sređivanja dobija se da je

$$(4.4.14) \quad \frac{1}{2} (s - s^*)(B + \lambda I)(s - s^*) \geq 0.$$

Kako je skup smerova:

$$\left\{ w : w = \pm \frac{s - s^*}{\|s - s^*\|}, \text{ za neko } s \text{ za koje je } \|s\| = \Delta \right\}$$

gust na jedinici sfere, (4.4.14) je dovoljan za dokaz (4.4.8).

Ostalo je da se pokaže da je $\lambda \geq 0$. Pošto važe (4.4.6) i (4.4.8) po s^* , iz tvrđenja (4.4.3)(i) sledi da je s^* minizuje \hat{m} , pa (4.4.12) važi.

Predpostavimo da samo negativne vrednosti λ ispunjavaju (4.4.6) i (4.4.8). Tada iz (4.4.12) imamo da je $m(s) \geq m(s^*)$ kad god je $\|s\| \geq \|s^*\| = \Delta$. Kako već znamo da s^* minimizuje m , za $\|s\| \leq \Delta$, iz toga sledi da je s^* neograničeni minimum funkcije m . Iz tvrđenja (4.4.3)(i) sledi da je $Bs = -g$ i B je pozitivno semidefinitna. A zbog toga su uslovi (4.4.6) i (4.4.8) ispunjeni za $\lambda = 0$, što je kontradikcija sa predpostavkom da samo negativne vrednosti λ ispunjavaju uslove. Zaključujemo da je $\lambda \geq 0$, čime je dokaz završen.

□

4.4.2 Analiza Levenberg-Marquardtove metode

U praksi se uglavnom koristi Gauss-Newtonova metoda sa linijskim pretraživanjem. Međutim, ako $J(x)$ nema pun rang, ili Gauss-Newtonova metoda neće raditi kako treba ili će algoritam konvergirati ka nestacionarnoj tački (neophodan uslov za minimum je da bude stacionarna tačka).

Da bi prevazišli ovaj problem, koristi se tehnika oblasti poverenja. Nova ideja je da stavimo ograničenje u vidu oblasti poverenja i razmatramo sledeći model:

$$(4.4.15) \quad \min \frac{1}{2} \|J(x_k)(x - x_k) + r(x_k)\|^2$$

gde je

$$(4.4.16) \quad \|(x - x_k)\| \leq \Delta_k,$$

gde je $\Delta_k > 0$ prečnik oblasti poverenja. Model (4.4.15)-(4.4.16) se može zapisati i kao

$$(4.4.17) \quad \min q_k(x) = \frac{1}{2} \|r_k\|^2 + r_k^T J(x_k)(x - x_k) + \frac{1}{2} (x - x_k)^T J(x_k)^T J(x_k)(x - x_k)$$

gde je $\|(x - x_k)\|_2 \leq \Delta_k$.

Stavimo da je $s = x - x_k$.

Rešenje potproblema (4.4.15)-(4.4.16) je okarakterisan rešenjem sistema

$$(4.4.18) \quad (J(x_k)^T J(x_k) + \mu_k I)s = -J(x_k)^T r(x_k)$$

Dakle,

$$(4.4.19) \quad x_{k+1} = x_k - (J(x_k)^T J(x_k) + \mu_k I)^{-1} J(x_k)^T r(x_k).$$

Kada je $\|(J(x_k)^T J(x_k))^{-1} J(x_k)^T r(x_k)\| \leq \Delta_k$, tada je $\mu_k = 0$ i potproblem je rešen po s_k . U ostalim slučajevima postoji $\mu_k > 0$ takvo da rešenje s_k ispunjava $\|s_k\| = \Delta_k$ i

$$(4.4.20) \quad (J(x_k)^T J(x_k) + \mu_k I)s_k = -J(x_k)^T r(x_k).$$

Kako je $(J(x_k)^T J(x_k) + \mu_k I)$ pozitivno definitna, smer s izведен iz (4.4.18) je opadajući. Ova metoda se naziva **Levenberg-Marquardtova metoda**.

Ovaj metoda može i drugačije da se posmatra. Na primer, kao zamena pravila između Gauss-Nutnove metode i Metode najbržeg spusta (gradijentna metoda u kojoj se veličina koraka α_k bira tako da se u svakoj iteraciji metode postiže maksimalno opadanje vrednosti funkcije cilja). Ovo podrazumeva da ova metoda dozvoljava biranje bilo kog od ova smera, za smer kretanja.

Kada je $\mu_k = 0$, to se svodi na Gauss-Newtonov smer.

Kada je μ_k veoma velik, (4.4.18) se aproksimira sa

$$(4.4.21) \quad \mu_k Is = -J(x_k)^T r(x_k).$$

Dobijeni smer je blizu smera najbržeg spusta. Štaviše ako umesto I , ubacimo neku pozitivno definitnu i dijagonalnu matricu D_k , tada (4.4.18) postaje

$$(4.4.22) \quad (J(x_k)^T J(x_k) + \mu_k D_k)s = -J(x_k)^T r(x_k).$$

U ovom slučaju je dobijeni smer kombinacija Gauss-Newtonovog smera i smera najbržeg spusta uzimajući u obzir matricu D_k .

Razmotrimo sada neke osobine Levenberg-Marquardtove metode.

Neka je $s = s(\mu)$ rešenje

$$(4.4.23) \quad (J^T J + \mu I)s = -J^T r$$

gde je $J = J(x)$, $r = r(x)$ i $g = g(x) = J^T r$.

Teorema 4.4.4

Neka su $\mu_k > 0$ i s_k rešenje sistema (4.4.18). Tada je s_k globalno rešenje problema

$$(4.4.24) \quad \min q^{(k)}(s) = \frac{1}{2} \|J_k s + r_k\|^2$$

gde je

$$(4.4.25) \quad \|s\| \leq \|s_k\|.$$

Dokaz:

Kako je s_k rešenje problema (4.4.18), tada je

$$\begin{aligned} q^{(k)}(s_k) &= \frac{1}{2} r_k^T r_k + r_k^T J_k s_k + \frac{1}{2} s_k^T J_k^T J_k s_k \\ &= \frac{1}{2} r_k^T r_k - s_k^T (J_k^T J_k + \mu_k I)s_k + \frac{1}{2} s_k^T J_k^T J_k s_k \\ (4.4.26) \quad &= \frac{1}{2} r_k^T r_k - \mu_k s_k^T s_k - \frac{1}{2} s_k^T J_k^T J_k s_k. \end{aligned}$$

Sa druge strane , za bilo koje s , imamo

$$\begin{aligned}
 q^{(k)}(s) &= \frac{1}{2} r_k^T r_k + s^T J_k^T r_k + \frac{1}{2} s^T J_k^T J_k s \\
 &= \frac{1}{2} r_k^T r_k - s^T (J_k^T J_k + \mu I) s_k + \frac{1}{2} s^T J_k^T J_k s \\
 (4.4.27) \quad &= \frac{1}{2} r_k^T r_k - \mu_k s^T s_k - s^T J_k^T J_k s_k + \frac{1}{2} s^T J_k^T J_k s .
 \end{aligned}$$

Tada, za bilo koje s koje ispunjava uslov $\|s\| \leq \|s_k\|$, zaključujemo da

$$\begin{aligned}
 q^{(k)}(s) - q^{(k)}(s_k) &= \frac{1}{2} (s_k - s)^T J_k^T J_k (s_k - s) + \mu_k (s_k^T s_k - s^T s_k) \\
 &\geq \frac{1}{2} (s_k - s)^T J_k^T J_k (s_k - s) + \mu_k \|s_k\| (\|s_k\| - \|s\|) \\
 (4.4.28) \quad &\geq 0 ,
 \end{aligned}$$

što pokazuje da je s_k globalno optimalno rešenje problema (4.4.24)-(4.4.25).

□

Teorema 4.4.5

Vektor s_k je rešenje problema (4.4.15)-(4.4.16), to jest

$$(4.4.29) \quad \min \frac{1}{2} \|J_k s + r_k\|^2$$

gde je

$$(4.4.30) \quad \|s\| \leq \Delta_k ,$$

za neko $\Delta_k > 0$, ako i samo ako postoji $\mu \geq 0$, takvo da važi

$$(4.4.31) \quad (J_k^T J_k + \mu I) s_k = -J_k^T r_k ,$$

$$(4.4.32) \quad \mu(\Delta_k - \|s_k\|) = 0 ,$$

$$(4.4.33) \quad \|s_k\| \leq \Delta_k .$$

Dokaz:

U teoremi 4.4.2, uslov (4.4.8) je odmah ispunjen, jer je $J^T J$ pozitivno semidefinitna i $\lambda \geq 0$. Uslovi (4.4.31) i (4.4.32) slede iz (4.4.6) i (4.4.7) redom.

□

Najčešće je Levenberg-Marquardtova metoda okarakterisana jednakošću

$$(4.4.34) \quad (J(x_k)^T J(x_k) + \mu_k D(x_k)) s = -J(x_k)^T r(x_k) ,$$

gde je $D(x_k)$ dijagonalana i pozitivno definitna matrica.

Ako uvedemo dužinu koraka α_k , pri čemu je α_k minimum funkcije $\min_{\alpha \geq 0} (f(x_k) - \alpha (J(x_k)^T J(x_k))^{-1} J(x_k)^T r(x_k))$, tada iterativni niz $x_{k+1} = x_k - \alpha_k (J(x_k)^T J(x_k) + \mu_k I)^{-1} J(x_k)^T r(x_k)$ opada. Ako α_k ispunjava Armijovo i Goldsteinovo pravilo (definicija 2.1.10):

$$(4.4.35) \quad f(x_k + \alpha_k s_k) \leq f(x_k) + \sigma \alpha_k g_k^T s_k , \quad \sigma \in (0, \frac{1}{2}) ,$$

tada α_k obezbeđuje zadovoljavajuće opadanje posmatrane funkcije. Upotrebljava se kada je strogo linijsko pretraživanje neefikasno, na primer kada je tačka dobijena iterativnim postupkom daleko od rešenja. Izbegavanjem strogo liniskog pretraživanja skraćuje se vreme računanja.

Teorema 4.4.6

U (4.4.34), uslovni broj za $J(x)^T J(x) + \mu D(x)$ je nerastuća funkcija po μ .

Dokaz:

Neka su β_1 i β_n najveća i najmanja sopstvena vrednost matrice $D(x)$, respektivno. Neka su λ_1 i λ_n najveća i najmanja sopstvena vrednost $J(x)^T J(x) + \mu D(x)$, respektivno. Neka su $\mu_1 > \mu_2 \geq 0$. Kako je rang normalne matrice u koveksnom omotaču njegovog spektra, imamo da je

$$\begin{aligned} \frac{\lambda_1(\mu_1)}{\lambda_n(\mu_1)} &\leq \frac{\lambda_1(\mu_2) + (\mu_1 - \mu_2)\beta_1}{\lambda_n(\mu_2) + (\mu_1 - \mu_2)\beta_n} \\ &\leq \frac{\lambda_1(\mu_2) + (\mu_1 - \mu_2)(1 + \mu_2)^{-1}\lambda_1(\mu_2)}{\lambda_n(\mu_2) + (\mu_1 - \mu_2)(1 + \mu_2)^{-1}\lambda_n(\mu_2)} \\ &= \frac{\lambda_1(\mu_2)}{\lambda_n(\mu_2)}. \end{aligned}$$

Odatle sledi zaključak.

□

Konvergencija Levenberg-Marquardtove metode

Teorema 4.4.7

Neka je $\{x_k\}$ niz dobijen Levenberg-Marquardtovom metodom (4.4.34). Predpostavimo da je dužina koraka α_k određena Armijovim pravilom (4.4.35). Ako postoji podniz $\{x_{k_i}\}$ koji konvergira ka x^* i ako odgovarajući podniz $\{J_{k_i}^T J_{k_i} + \mu_{k_i} D_{k_i}\}$ konvergira ka nekoj pozitivno definitnoj matrici P , gde $J_{k_i} = J(x_{k_i})$ i $D_{k_i} = D(x_{k_i})$ označavaju pozitivno definitnu matricu, tada je $g(x^*) = 0$.

Dokaz:

Izvodi se kontradikcijom. Predpostavimo da je $g(x^*) \neq 0$. Neka su

$$\begin{aligned} s_{k_i} &= -(J_{k_i}^T J_{k_i} + \mu_{k_i} D_{k_i})^{-1} J_{k_i}^T r_{k_i}, \\ s^* &= \lim s_{k_i} = -P^{-1} J(x^*)^T r(x^*), \end{aligned}$$

gde je $r_{k_i} = r(x_{k_i})$. Važi da je $g(x^*)^T s^* < 0$. Neka su $\beta \in (0,1)$, $\sigma \in (0, \frac{1}{2})$ i neka je

m^* najmanje celobrojno ne negativno m , takvo da je

$$f(x^* + \beta^m s^*) \leq f(x^*) + \sigma \beta^m g(x^*)^T s(x^*).$$

Za veoma veliko k , imamo da je

$$f(x_{k_i} + \beta^{m^*} s_{k_i}) \leq f(x_{k_i}) + \sigma \beta^{m^*} g(x_{k_i})^T s_{k_i}.$$

Dakle važi:

$$(4.4.36) \quad f(x_{k_i+1}) = f(x_{k_i} + \beta^{m_{k_i}} s_{k_i}) \leq f(x_{k_i}) + \sigma \beta^{m^*} g(x_{k_i})^T s_{k_i}.$$

Kako je metoda monotono opadajuća imamo da je

$$\lim f(x_{k_i+1}) = \lim f(x_{k_i}) = f(x^*).$$

Uzimajući granične vrednosti na obe strane u (4.4.36), dobija se da je

$$f(x^*) \leq f(x^*) + \sigma \beta^{m^*} g(x^*)^T s^* < 0,$$

a to je nemoguće jer je $\sigma \beta^{m^*} g(x^*)^T s^* < 0$.

□

Gornja teorema formuliše konvergenciju podniza. U sledećoj dajemo konvergenciju celog niza.

Teorema 4.4.8

Predpostavimo da sledeće predpostavke važe:

- (a) Nivo skup $L(\bar{x}) = \{x \mid f(x) \leq f(\bar{x})\}$ je ograničen i zatvoren za bilo koje $\bar{x} \in R^n$.
- (b) Broj stacionarnih tačaka, u kojima je vrednost funkcije $f(x)$ ista, je konačan.
- (c) $J(x)^T J(x)$ je pozitivno definitna za svako x .
- (d) $\mu_k \leq M < \infty$ za svako k , to jest M je gornje ograničenje μ_k .

Tada za svaku početnu tačku x_0 , niz $\{x_k\}$ generisan Levenberg-Marquardtovom metodom konvergira ka stacionarnoj tački funkcije $f(x)$.

Dokaz:

Iz (a) i monotonosti iterativne funkcije znamo da je niz $\{x_k\}$ u kompaktnom skupu $L(\bar{x})$. To pokazuje da $\{x_k\}$ mora imati tačke nagomilavanja. Da bi dokazali teoremu, potrebno je dokazati da su tačke nagomilavanja jedinstvene. Iz (c) i (d) i Teoreme 4.4.7 imamo da je svaka tačka nagomilavanja niza $\{x_k\}$ jedinstvena. Kako je $\{f(x)\}$ monotono opadajući niz, $f(x)$ ima iste vrednosti kao i tačke nagomilavanja niza $\{x_k\}$. Takođe iz (b) sledi da je broj stacionarnih tačaka funkcije f na skupu $L(\bar{x})$ konačan. Odatle sledi da ima samo konačno mnogo tačaka nagomilavanja. Primetimo da za neke podnizove $\{x_{k_i}\}$ imamo da $x_{k_i} \rightarrow \hat{x}_k$ i $\lim_{k \rightarrow \infty} g(x_{k_i}) = g(\hat{x}_k) = 0$, kao i da je $s(\mu_{k_i}) = -(J(x_{k_i})^T J(x_{k_i}) + \mu_{k_i} D(x_{k_i}))^{-1} g(x_{k_i})$.

Tada iz (c) i (d) sledi da $s(\mu_{k_i}) \rightarrow 0$. Odatle za niz $\{s(\mu_k)\}$ imamo da $s(\mu_k) \rightarrow 0$.

Predpostavimo, na trenutak da imamo više od jedne tačke nagomilavanja, niza $\{x_k\}$. Neka je ε^* najmanji razmak između bilo koje dve tačke nagomilavanja. Kako je $\{x_k\}$ kompaktan skup, onda postoji pozitivan celobrojan broj N , takav da za svako $k \geq N$, x_k je sadržan u zatvorenoj lopti sa nekom tačkom nagomilavanja kao centrom i $\varepsilon^*/4$ kao prečnikom.

Sa druge strane, postoji celobrojni $N' \geq N$ takav da $\|s(\mu_k)\| < \frac{\varepsilon^*}{4}$, za svako $k \geq N'$. Odatle, kada je $k \geq N'$, svi x_k su u zatvorenoj lopti, gore navedenoj, sa tačkom nagomilavanja kao centrom i $\varepsilon^*/4$ kao prečnikom. Tada imamo uslove koji dokazuju teoremu. \square

Prethodno navedena teorema, uspostavlja globalnu konvergenciju Levenberg-Marquardtove metode. Razmotrimo još i brzinu ove metode.

Teorema 4.4.9

Predpostavimo da niz x_k generisan Levenberg-Marquardtovom metodom konvergira ka stacionarnoj tački x^* . Neka je l najmanja sopstvena vrednost matrice $J(x)^T J(x)$, a M maksimum aposlutnih vrednosti iz skupa sopstvenih vrednosti od $S(x^*) = \sum_{i=1}^m r_i(x^*) \nabla^2 r_i(x^*)$. Ako stavimo da je

$$(4.4.37) \quad \tau = M/l < 1, \quad 0 < \beta < (1-\tau)/2, \quad \mu_k \rightarrow 0,$$

tada za svako dovoljno veliko k , dužina koraka je $\alpha_k = 1$,

$$(4.4.38) \quad \limsup \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \tau,$$

i x^* je strogo lokalni minum funkcije $f(x)$.

Dokaz:

Prvo dokažimo da je $\alpha_k = 1$ za dovoljno veliko k . Primetimo da je

$$(4.4.39) \quad f(x_k + s_k) - f(x_k) = g_k^T s_k + \frac{1}{2} s_k^T H(x_k + \theta s_k) s_k,$$

gde je $\theta \in (0,1)$. U skladu sa Armijovim pravilom (4.4.35), da bi dokazali da je $\alpha_k = 1$, za dovoljno veliko k , treba dokazati da je

$$(4.4.40) \quad \beta g_k^T s_k - [f(x_k + s_k) - f(x_k)] \geq 0.$$

Koristeći $g_k = -(J_k^T J_k + \mu_k D_k)s_k$ i (4.4.39), leva strana (4.4.40) se može zapisati kao

$$\begin{aligned} & (1-\beta) s_k^T (J_k^T J_k + \mu_k D_k) s_k - \frac{1}{2} s_k^T H(x_k + \theta s_k) s_k \\ &= s_k^T \left[(1-\beta) J_k^T J_k - \frac{1}{2} H(x_k) + (1-\beta) \mu_k D_k s_k - \frac{1}{2} (H(x_k + \theta s_k) - H(x_k)) \right] s_k \end{aligned}$$

$$= s_k^T \left[\left(\frac{1}{2} - \beta \right) J_k^T J_k - \frac{1}{2} S(x_k) + V_k \right] s_k,$$

gde je $V_k = (1-\beta) \mu_k D_k - \frac{1}{2} (H(x_k - \theta s_k) - H(x_k))$, a $S(x)$ je definisan u (4.1.6).

Kako $V_k \rightarrow 0$, da bi dokazali da (4.4.40) važi za dovoljno veliko k , treba pokazati da $(1-\beta) J_k^T J_k - \frac{1}{2} S(x_k)$ konvergira ka pozitivno definitnoj matrici.

Primetimo da je najmanja sopstvena vrednost od $(1-\beta)J(x^*)^T J(x^*) - \frac{1}{2}S(x^*)$ ograničena odozdo i da je gornja granica

$$\left(\frac{1}{2}-\beta\right)l - \frac{1}{2}M = l\left[\frac{1}{2}-\beta-\frac{1}{2}\tau\right] > 0,$$

što je tačno jer se druga nejednakost u (4.4.37) odnosi na β . Tako dobijamo da je $\alpha_k = 1$ za dovoljno veliko k .

Dalje dokazujemo (4.4.38). Iz (4.4.34) i (4.1.7) imamo da je

$$(4.4.41) \quad \begin{aligned} x_{k+1} - x^* &= x_k - x^* - (J_k^T J_k + \mu_k D_k)^{-1} g_k \\ &= x_k - x^* - (J_k^T J_k + \mu_k D_k)^{-1} [H_k(x_k - x^*) + g_k + H_k(x^* - x_k)] \\ &= -(J_k^T J_k + \mu_k D_k)^{-1} [S(x_k)(x_k - x^*) - \mu_k D_k(x_k - x^*) + g_k + H_k(x^* - x_k)]. \end{aligned}$$

Uzimanjem normi važi

(4.4.42)

$$\|x_{k+1} - x^*\| \leq \| (J_k^T J_k)^{-1} \| (\|S(x_k)\| \|x_k - x^*\| + \mu_k \|D_k\| \|x_k - x^*\| + \|g_k + H_k(x^* - x_k)\|).$$

Kako je

$$(4.4.43) \quad \|g_k + H_k(x^* - x_k)\| = \|g_k - g(x^*) - H_k(x_k - x^*)\| \leq \varepsilon_k \|x_k - x^*\|,$$

gde $\varepsilon_k \rightarrow 0$, deleći obe strane (4.4.42) sa $\|x_k - x^*\|$ dobija se

$$(4.4.44) \quad \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \| (J_k^T J_k)^{-1} \| (\|S(x_k)\| + \mu_k \|D_k\| + \varepsilon_k).$$

Primetimo da $\mu_k \rightarrow 0$ i $\varepsilon_k \rightarrow 0$ i odatle imamo da

$$\limsup \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \frac{M}{l} = \tau$$

što dokazuje (4.4.38).

Konačno kako $g(x^*) = 0$ i $H(x^*) = J(x^*)^T J(x^*) + S(x^*)$ sa donjom granicom $l - M > 0$ najmanjih sopstvenih vrednosti, tada je $H(x^*)$ pozitivno definitna. Zbog toga je x^* strogo lokalni minimum funkcije $f(x)$.

□

Kao što je gore pomenuto, Levenberg-Marquardtova metoda se može opisati i analizirati koristeći princip metoda oblasti poverenja (4.4.15) i (4.4.16) ili (4.4.17). Da bi predstavili Levenberg-Marquardtov algoritam kroz oblast poverenja, potrebna su nam sledeća tvrđenja:

Trvđenje 4.4.10

Neka je s_k rešenje potproblema (4.4.1) i neka je $\|\cdot\| = \|\cdot\|_2$. Tada važi da je

$$(4.4.45) \quad \begin{aligned} Pred_k &= q^{(k)}(0) - q^{(k)}(s_k) \\ &\geq \frac{1}{2} \|g_k\| \min \left\{ \Delta_k, \frac{\|g_k\|}{\|B_k\|} \right\}. \end{aligned}$$

Dokaz se može naći u knjizi [1].

Predpostavimo da je norma aproksimacije Hesijana B_k ravnomerno ograničena i da je skup $\{x \mid f(x) \leq f(x_0)\}$ ograničen. Neka je i funkcija $f: R^n \rightarrow R$, na njemu neprekidno diferencijabilna. Zbog generalizacije, dopuštamo i da dužina aproksimativnog rešenja s_k potproblema (4.4.1) prelazi granicu oblasti poverenja, pod pretpostavkom da ostane u okviru fiksnog množioca granice tj. $\|s_k\| \leq \tilde{\eta} \Delta_k$, pri čemu je $\tilde{\eta}$ pozitivna konstanta. Ove predpostavke nazivamo pretpostavke A_0 .

Teorema 4.4.11

Neka važe pretpostavke A_0 . Ako algoritam 4.4.1 ima beskonačno mnogo uspešnih iteracija, onda niz algoritma 4.4.1 ispunjava

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

Dokaz se može naći u knjizi [1].

Tako prateći razmatranja u odeljku 4.4.1, dobijamo sledeći algoritam i teoremu, koji su direktna posledica Algoritma 4.4.1 i Teoreme 4.4.11.

Algoritam 4.4.12 (Levenberg-Marquardtov algoritam kroz oblast poverenja)

Korak 0. Data je početna tačka x_0 , $\bar{\Delta}$, $\Delta_0 \in (0, \bar{\Delta})$, $\varepsilon \geq 0$, $0 \leq \eta_1 < \eta_2 < 1$ i $0 < \gamma_1 < 1 < \gamma_2$, $k := 0$

Korak 1. Ako je $\|g_k\| = \|J_k^T r_k\| \leq \varepsilon$ stati; inače preći na korak 2.

Korak 2. Aproksimativno rešiti problem (4.4.15)-(4.4.16) po s_k .

Korak 3. Izračunati

$$\begin{aligned} \text{Pred}_k &= q^{(k)}(x_k) - q^{(k)}(s_k), \\ \text{Ared}_k &= f(x_k) - f(x_k + s_k), \\ r_k &= \frac{\text{Ared}_k}{\text{Pred}_k}. \end{aligned}$$

Korak 4. Ako je $r_k \leq \eta_1$, staviti da je $\Delta_k = \gamma_k \Delta_k$ i ići na korak 2.

Korak 5. Staviti da je $x_{k+1} = x_k + s_k$ i

$$\Delta_{k+1} = \begin{cases} \min\{\gamma_2 \Delta_k, \bar{\Delta}\}, & \text{ako je } r_k \geq \eta_2 \text{ i } \|s_k\| = \Delta_k, \\ \Delta_k, & \text{inače} \end{cases}$$

Korak 6. Staviti $k := k + 1$ i ići na korak 1.

Od koraka 2, u gore navedenom algoritmu s_k je približno rešenje problema (4.4.15) - (4.4.16).

To sledi iz tvrđenja 4.4.10, gde je

$$(4.4.46) \quad q^{(k)}(0) - q^{(k)}(s_k) \geq c_1 \|J_k^T r_k\| \min\left(\Delta_k, \frac{\|J_k^T r_k\|}{\|J_k^T J_k\|}\right)$$

za neku konstantu $c_1 > 0$.

Sada možemo zaključiti konvergenciju, koja je direktna posledica teoreme 4.4.11.

Teorema 4.5.10

Predpostavimo da je funkcija $f(x) = \frac{1}{2} \sum_{i=1}^m r_i^2(x)$ dva puta neprekidno diferencijabilna. Neka je skup $L(x_0) = \{x \mid f(x) < f(x_0)\}$ ograničen i neka postoje konstante $M_1 > 0, M_2 > 0$, takve da je

$$\begin{aligned} \|\nabla^2 f(x)\| &\leq M_1, \forall x \in L(x_0) \\ \|J(x)^T J(x)\| &\leq M_2, \forall x \in L(x_0) \end{aligned}$$

Tada važi :

$$(4.4.47) \quad \lim_{k \rightarrow \infty} \nabla f(x_k) = \lim_{k \rightarrow \infty} J_k^T r_k = 0.$$

5. Primena metoda najmanjih kvadrata

Primenu metoda najmanjih kvadrata možemo naći u mnogim oblastima, a ovde će biti prikazana primena u metodama veštačke inteligencije, konkretno u veštačkim neuronskim mrežama.

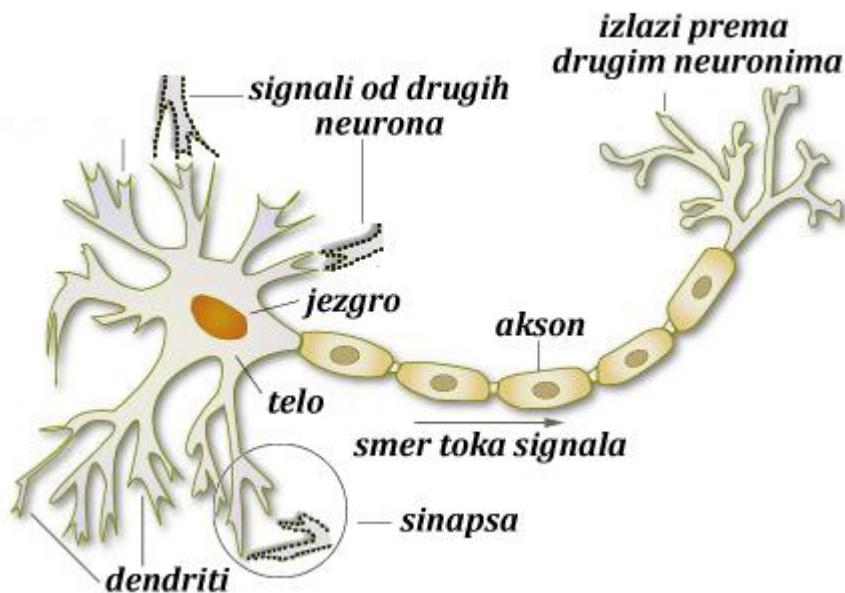
5.1 Osnove neuronskih mreža

Veštačka neuronska mreža ili samo neuronska mreža je jedan oblik implementacije sistema veštačke inteligencije. Neuronske mreže koriste principe ljudskog mozga i njegove strukture, kako bi razvile strategiju obrade podataka.

5.1.1 Biološke osnove neuronskih mreža

Neuron kao osnovna jedinica nerevnog sistema se sastoji iz četiri osnovna dela:

- ulazni deo ćelije – čine ga skup razgranatih niti nazvanih dendriti
- telo ćelije – obrađuje signale koje dobija od dendrita, dobijajući tako izlazni impuls (soma)
- izlazni deo ćelije – razgranata nit preko koje se izlazni impuls prosleđuje (aksonom)
- sinapsa – mesto gde se akson dodiruje sa dendritima neke druge ćelije i na taj način se impulsi prenose od jedne do druge nervne ćelije.



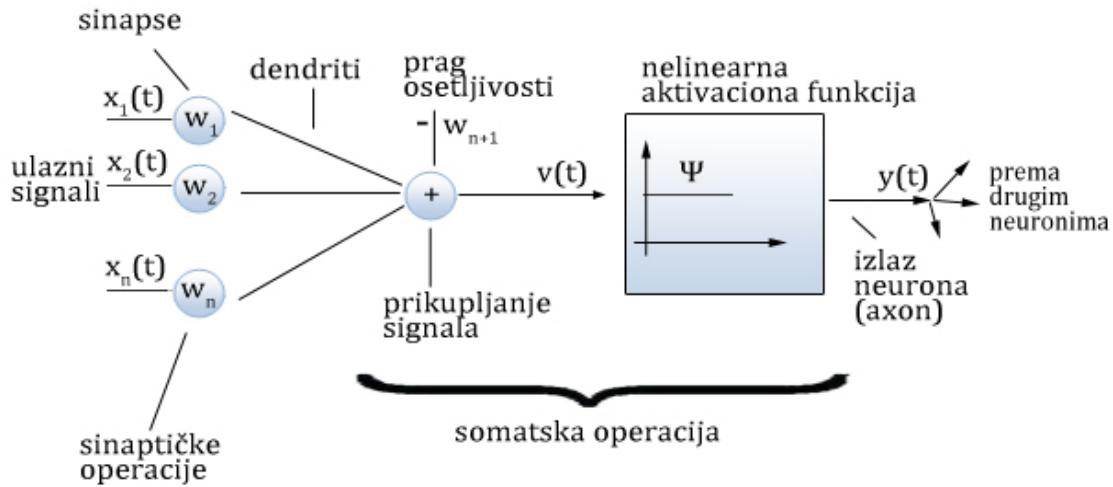
5.1 Šematski prikaz biološkog neurona

Posmatrajući kroz obradu signala, rad neurona se uprošćeno može opisati na sledeći način: signali se od sinapsi prosleđuju do tela neurona , gde se prikupljaju i obrađuju. Ti signali za telo neurona mogu biti pobuđujući ili smirujući. Matematički gledano, oni imaju suprotan predznak. Ako je njihova kumulativna vrednost tokom kratkog vremenskog intervala, veća od praga osetljivosti neurona, telo neurona generiše impulse, koji se šalju duž aksona prema drugim neuronima, a ako je manja, neuron ostaje nepobuđen i ne generiše impulse.

Odavde imamo da se obrada signala u neuronu odvija kroz dve operacije:

- *sinaptička operacija* – daje određeni značaj (težinu) svakom ulaznom signalu u neuron
- *somatska operacija* – skuplja sve „otežane“ ulazne signale i ako je njihova kumulativna vrednost veća od praga osetljivosti neurona, generiše impulse koje šalje prema drugim neuronima.

Po uzoru na biološki neuron, kreiran je veštački neuron. Prvi modeli veštačkih neurona su takođe obrađivali signale kroz sinaptičku i somatsku operaciju. Ovakav model neurona zove se *perceptron*.



5.2 Šematski prikaz perceptron-a

Matematički se perceptron može prikazati sledećim izrazima:

$$(5.1.1) \quad v(t) = \sum_{i=1}^n w_i(t) \cdot x_i(t) - w_{n+1}(t),$$

$$(5.1.2) \quad y(t) = \psi(v(t)),$$

gde su:

$x_u(t) = [x_1(t), \dots, x_n(t)]^T$ - vektor ulaznih signala neurona (pobudni vektor),

$w_s(t) = [w_1(t), \dots, w_n(t)]^T$ - vektor sinaptičkih težinskih koeficijenata,

w_{n+1} - prag osetljivosti neurona,

$v(t)$ - izlaz operacije konfluencije (mera sličnosti ulaznih signala sa sinaptičkim težinskim koeficijentima),

$\psi(t)$ - nelinearna aktivaciona funkcija,

$y(t)$ - izlaz neurona.

Ako se vektor ulaza proširi članom $x_{n+1}=1$, onda izraz (5.1.1) možemo napisati na sledeći način:

$$(5.1.3) \quad v(t) = \sum_{i=1}^{n+1} w_i(t) \cdot x_i(t) = w(t)^T x(t),$$

gde su

$x(t) = [x_1(t), \dots, x_n(t), x_{n+1}]^T$ - prošireni vektor ulaznih signala neurona,

$w(t) = [w_1(t), \dots, w_n(t), w_{n+1}]^T$ - vektor težinskih koeficijenata (vektor sinaptičkih težinskih koeficijenata proširen pragom osetljivosti neurona).

Izraz (5.1.3) opisuje sinaptičku operaciju i prve dve somatske operacije (prikljupljanje otežanih ulaznih signala i poređenje njihovog zbiru sa pragom osetljivosti). Ove tri operacije zajedno čine *operaciju konfluencije*, a izraz (5.1.2) opisuje nelinearnu *aktivacionu funkciju*.

Sa stanovišta primene nelinearnih metoda najmanjih kvadrata u neuronskim mrežama, najvažniji deo u teoriji neuronskih mreža su algoritmi učenja. Da bi bolje razumeli algoritme učenja, osvrnućemo se ukratko na klasifikaciju neuronskih mreža.

5.1.2 Klasifikacija neuronskih mreža

Neuronske mreže istražuje veliki broj naučnika u različitim naučnim disciplinama. Kao rezultat toga je veliki broj različitih vrsta neuronskih mreža koje se međusobno razlikuju i/ili po modelima veštačkih neurona od koji su sastavljene i/ili po načinu organizacije neurona u mreži i/ili po применjenom algoritmu učenja.

U odnosu na strukturu, neuronske mreže se dele na statičke i dinamičke, zavisno od modela neurona od kojeg su sastavljene i od načina prostiranja signala kroz mrežu. Neuroni se u okviru mreže najčešće organizuju u slojeve, pa se razlikuju jednoslojne i višeslojne mreže. Moguće ih je podeliti i po načinu prostiranja sinaptičkih veza, pa se razlikuje: samo unapredno (statičke neuronske mreže), samo laterarno (aditivne, Hopfieldove i shuntirajuće mreže), topološki određeno (LVQ mreže), unapredno/povratno (BAM i ART mreže) ili mešovito (celularne, time-delay i counterpropagation mreže) prostiranje sinaptičkih veza.

Statičke neuronske mreže su najčešće korišćene. Njihov osnovni gradivni element je statički neuron, a način organizovanja neurona je unapredni, što znači da svaki neuron može biti povezan s ulazima u mrežu i/ili drugim neuronima, ali tako da se pri povezivanju ne formiraju povratne veze tj. ne sadrže dinamičke članove, što ih čini strukturno stabilnima. Najčešće se koriste višeslojne perceptronske mreže (MLP mreže) i mreže bazirane na funkcijama sa kružnom osnovicom (RBF mreže).

Dinamičke neuronske mreže u svojoj strukturi, osim unaprednog imaju i povratno delovanje. To je korisno u slučajevima kada je potrebno da neuronska mreža sačuva neku informaciju neodređeni vremenski period. Nedostak dinamičkih neuronskih mreža je da stabilnost nije zagarantovana, kao i da je proces učenja znatno složeniji, a samim tim i sporiji. Najčešće korišćene su Hopfieldove, Elmanove i NARX dinamičke neuronske mreže.

Većina neuronskih mreža zahteva učenje tj. primenu algoritama koji podešavaju vrednosti sinaptičkih težinskih koeficijenata. Ciljevi učenja mreže, kao i izbor algoritma učenja zavisi od njene primene. Algoritme učenja neuronske mreže je po načinu učenja moguće podeliti na algoritme učenja bazirane na grešci (error-based learning algorithms), algoritme učenja baziranih na izlazu mreže (output-based learning algorithms) i algoritme učenja sa ojačanjem (reinforcement learning algorithms)

Neka je $f(x)$ neprekidna, nelinearna funkcija više promenljivih definisana na skupu $D \subseteq \mathbb{R}^{n(x)}$, koju je potrebno aproksimirati funkcijom $q_N(\theta, x)$, gde je $\theta \in \mathbb{R}^{n(\theta)}$ vektor parametara aproksimacijske funkcije koji se podešavaju. Zadatak je odrediti optimalne vrednosti parametara $\theta = \theta^*$, uz koje je q_N najbolja aproksimacija funkcije f .

Problem se može zapisati kao

$$(5.1.4) \quad \theta^* = \arg \min_{\theta \in \mathbb{R}^{n(\theta)}} \rho[q_N(\theta, x), f(x)],$$

pri čemu je ρ funkcija mere kvaliteta aproksimacije (uzima u obzir uticaj svih $x \in D$ na kvalitet aproksimacije). Za ρ se u najvećem broju slučajeva uzima funkcija udaljenosti definisana L_p normom (najčešće je $p = 2$), pa imamo da je:

$$(5.1.5) \quad \rho[q_N(\theta, x), f(x)] = \|f(x) - q_N(\theta, x)\|_p.$$

Kako se f aproksimira na osnovu konačnog skupa mernih ulazno-izlaznih vrednosti, prethodni izraz se može zapisati kao:

$$(5.1.6) \quad \rho[q_N(\theta, x), f(x)] = \sum_{v=1}^N [f(x(v)) - q_N(\theta, x(v))]^2,$$

pri čemu je v redni broj vektora mernih podataka, a N ukupan broj vektora mernih podataka. Optimalne vrednosti parametara aproksimacijske funkcije θ^* se mogu definisati kao argument koji minimizuje izraz (5.1.6):

$$(5.1.7) \quad \theta^* = \arg \min_{\theta} \sum_{v=1}^N [f(x(v)) - q_N(\theta, x(v))]^2 = \arg \min_{\theta} \sum_{v=1}^N r^T(v, \theta) \cdot r(v, \theta),$$

gde je $r(v, \theta)$ vektor greške na v -tom mernom uzorku, dimenzije $n(L)$.

Za vektor parametara θ^* se kaže da je najbolje rešenje problema aproksimacije.

Nelinearno preslikavanje iz ulaznog u izlazni prostor statičkom neuronskom mrežom se može opisati nelinearnom funkcijom

$$(5.1.8) \quad y_L = q_N(x, \theta),$$

gde je θ vektor parametara koji se podešavaju.

Tako na primer, vektor parametara MLP neuronske mreže sadrži sinaptičke težinske koeficijente mreže $w_{l,i,j}$ se može zapisati kao

$$(5.1.9) \quad \theta = [\theta_1, \dots, \theta_{n(\theta)}]^T = [w_{1,1,1}, \dots, w_{1,n(1),n(0)+1}, \dots, w_{L,1,1}, \dots, w_{L,n(L),n(L-1)+1}]^T,$$

pri čemu je $n(\theta) = \sum_{l=1}^L n(l) \cdot [n(l-1)+1]$ ukupan broj parametara MLP mreže.

5.2 Algoritmi učenja neuronskih mreža

5.2.1 Osnovni pojmovi

Algoritmi učenja podešavaju parametre neuronske mreže sa ciljem postizanja željenog ponašanja. U identifikaciji i upravljanju nelinearnim dinamičkim procesima najčešće je poznato željeno ponašanje, pa se za njeno učenje primenjuju algoritmi bazirani na grešci. Za meru greške uzima se neka kriterijumska funkcija $q(\theta)$, koja može biti bilo koja pozitivna skalarna funkcija zavisna od parametara neuronske mreže θ . Algoritam učenja podešava parametre mreže sve dok kriterijumska funkcija ne dobije minimalnu vrednost, odnosno vrednost manju od unapred zadate vrednosti. Drugim rečima, problem podešavanja parametara neuronske mreže se svodi na problem nelinearne optimizacije sa kriterijumskom funkcijom $q(\theta)$, kao funkcijom cilja.

Minimum funkcije $q(\theta)$ se može definisati njenim razvojem u Tejlorov red u okolini tačke minimuma θ^* , do reda dva.

$$(5.2.1) \quad q(\theta) \approx q(\theta^*) + \nabla q(\theta)|_{\theta=\theta^*} \cdot \Delta\theta + \frac{1}{2} \Delta\theta^T \cdot H(\theta)|_{\theta=\theta^*} \cdot \Delta\theta,$$

pri čemu je $\Delta\theta = \theta - \theta^*$. Sa $\nabla q(\theta)$ označavamo gradijentni vektor kriterijumske funkcije,

$$(5.2.2) \quad \nabla q(\theta) = \left[\frac{\partial q(\theta)}{\partial \theta_1} \dots \frac{\partial q(\theta)}{\partial \theta_{n(\theta)}} \right]^T,$$

a sa $H(\theta) = \nabla^2 q(\theta)$ Hesijan matricu kriterijumske funkcije:

$$(5.2.3) \quad H(\theta) = \nabla^2 q(\theta) = \begin{bmatrix} \frac{\partial^2 q(\theta)}{\partial \theta_1^2} & \dots & \frac{\partial^2 q(\theta)}{\partial \theta_1 \partial \theta_{n(\theta)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 q(\theta)}{\partial \theta_{n(\theta)} \partial \theta_1} & \dots & \frac{\partial^2 q(\theta)}{\partial \theta_{n(\theta)}^2} \end{bmatrix}.$$

Za kriterijumsku funkciju najčešće se uzima

$$(5.2.4) \quad q(\theta) = \frac{1}{2} \sum_{v=1}^N r^T(v, \theta) \cdot r(v, \theta) = \frac{1}{2} \sum_{v=1}^N \sum_{i=1}^{n(L)} r_i^2(v, \theta) = \frac{1}{2} r^*(\theta) \cdot r^*(\theta),$$

pri čemu je $r^*(\theta)$ vektor grešaka na celom skupu mernih podataka dimenzije $N_r = N \cdot n(L)$. Za ovako definisanu kriterijumsku funkciju, gradijentni vektor i Hesijan matrica su oblika:

$$(5.2.5) \quad \nabla q(\theta) = J^T(\theta) \cdot r^*(\theta),$$

$$(5.2.6) \quad H(\theta) = \nabla^2 q(\theta) = J^T(\theta) J(\theta) + \sum_{i=1}^{N_r} r_i^*(\theta) \nabla^2 r_i^*(\theta),$$

pri čemu je $J(\theta)$ Jakobijan matrica

$$(5.2.7) \quad J(\theta) = \frac{\partial r^*(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial r_1^*(\theta)}{\partial \theta_1} & \dots & \frac{\partial r_1^*(\theta)}{\partial \theta_{n(\theta)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial r_{n(\theta)}^*(\theta)}{\partial \theta_1} & \dots & \frac{\partial r_{n(\theta)}^*(\theta)}{\partial \theta_{n(\theta)}} \end{bmatrix}.$$

U tački $\theta = \theta^*$ će biti minimum funkcije, ako su sledeći uslovi ispunjeni:

$$(5.2.8) \quad \nabla q(\theta)|_{\theta=\theta^*} = 0,$$

$$(5.2.9) \quad \Delta\theta^T \cdot H \cdot \Delta\theta > 0.$$

Uslov (5.2.8) je uslov prvog reda i ukazuje da je u tački θ^* ekstrem funkcije $q(\theta)$, ali ne i da li je u pitanju minimum, maksimum ili sedlo. Ako je $q(\theta^*)$ minimum funkcije $q(\theta)$, tada će mala promena $\Delta\theta$ u okolini θ^* izazvati pozitivan prirast

funkcije $q(\theta)$, bez obzira u kom smeru ide pomak, iz čega proizilazi potreba za uslovom (5.2.9). Ovaj uslov će biti uvek ispunjen ako je Hesijan matrica pozitivno definitna. U suprotnom se dobija da je θ^* ili tačka maksimuma ili sedlo.

Postoje dva osnovna načina minimiziranja funkcije $q(\theta)$ na kojima se baziraju algoritmi učenja neuronski mreža: nerekurzivni i rekurzivni.

Nerekurzivan način podrazumeva da se funkcija $q(\theta)$ minimizira tako da se promene parametara mreže akumuliraju preko svih N vektora mernih podataka (preko bloka podataka) i tek posle toga se stvarno menjaju parametri mreže. Algoritmi učenja neuronskih mreža bazirani na ovom principu nazivaju se nerekurzivnim algoritmima učenja.

Rekurzivan način podrazumeva da se funkcija $q(\theta)$ minimizira na osnovu lokalne kriterijumske funkcije $q_v(\theta)$, to jest parametri mreže se menjaju posle svakog vektora mernih podataka. Algoritmi učenja neuronskih mreža bazirani na ovom principu nazivaju se rekurzivnim algoritmima učenja.

Dešavaju se i situacije kada se mogu primeniti i nerekurzivni i rekurzivni algoritmi, a to je kada su svi merni podaci funkcije koja se minimizira poznati pre početka postupka učenja mreže.

U cilju prikaza primene metoda navedenih u poglavlju 4, u daljem izlaganju biće detaljnije predstavljeni pojedini nerekurzivni algoritmi učenja neuronskih mreža.

5.2.2 Nerekurzivni algoritmi učenja neuronskih mreža

Većina ovih algoritama se bazira na klasičnim postupcima nelinearne optimizacije. Funkcija cilja je kriterijumska funkcija $q(\theta)$ i algoritmi učenja minimizuju funkciju $q(\theta)$ podešavanjem vrednosti parametara mreže θ .

Najčešće korišćeni algoritmi učenja baziraju se na iterativnom postupku
(5.2.10)
$$\theta_{k+1} = \theta_k + \Delta\theta_k = \theta_k + \alpha_k s_k,$$

pri čemu je s_k smer traženja minimuma u k -toj iteraciji, a α_k koeficijent učenja u k -toj iteraciji. Zavisno od informacija na kojima se bazira određivanje smera traženja minimuma, razlikuju se i metode minimizacije. Tako imamo metode neposrednog traženja minimuma (koriste samo informacije o vrednosti funkcije $q(\theta)$) i gradijentne metode (osim informacija o vrednostima funkcije, koriste i informacije o vrednostima parcijalnih izvoda po parametrima mreže koji se podešavaju). Svaka od metoda ima svoje prednosti i mane. Uopšteno gledajući, metode neposrednog traženja minimuma sporije konvergiraju od gradijentnih, ali je ponekad analitičko izračunavanje izvoda i Hesijan matrice kriterijumske funkcije ili nemoguće ili je računski veoma zahtevno, zbog velikog broja parametara, pa je tada sasvim opravdano koristiti metode neposrednog traženja minimuma.

Sa druge strane postoji mogućnost da se izvodi i Hesijan ne izračunavaju analitički, već numerički i da se na taj način dobiju odgovarajuće aproksimacije. Međutim i tu treba biti oprezan, jer dobijene aproksimacije mogu biti loše i na taj način značajno narušiti svojstva algoritma.

Najčešće korišćene metode nelinearne optimizacije su gradijentne metode. Primenuju se uz uslov da izvod tj. gradijentni vektor kriterijumske funkcije postoji, kao i njen Hesijan, ali i da se mogu analitički izračunati. Postoje četiri osnovne grupe gradijentnih metoda (podela je prema načinu određivanja smera traženja minimuma):

1. Metode najbržeg spusta - $s_k := -\nabla q(\theta_k)$
2. Konjugovane metode - $s_k := -\nabla q(\theta_k) + \beta_k \cdot s_{k-1}$, pri čemu je β_k skalarni parametar koji osigurava konjugovanost
3. Newtonove metode - $s_k := -[\nabla^2 q(\theta_k)]^{-1} \nabla q(\theta_k)$
4. Kvazi-Newtonove metode - $s_k := -S_k \nabla q(\theta_k)$, pri čemu je $S_k \equiv [\nabla^2 q(\theta_k)]^{-1}$.

5.2.3 Newtonove metode

Newtonove metode koriste informacije, ne samo o prvim parcijalnim izvodima po parametrima mreže, već i o drugim parcijalnim izvodima. Baziraju se na kvadratnoj aproksimaciji kriterijumske funkcije u okolini tačke θ_k , koja se dobija razvojem funkcije $q(\theta)$ u Tejlorov red u okolini tačke θ_k , do reda dva.

$$(5.2.11) \quad q(\theta) \approx q(\theta_k) + \nabla^T q(\theta_k)(\theta - \theta_k) + \frac{1}{2}(\theta - \theta_k)^T \nabla^2 q(\theta_k)(\theta - \theta_k).$$

Vrednost funkcije q u tački minimuma je nula, pa tu tačku dobijamo tražeći parcijalni izvod po vektoru parametra θ i izjednačavanjem sa nula.

$$(5.2.12) \quad \frac{\partial q}{\partial \theta} = \nabla q(\theta_k) + \nabla^2 q(\theta_k)[\theta - \theta_k] = 0.$$

Ako se u jednakost stavi da je $\theta = \theta_{k+1}$ i predpostavi da postoji inverz Hesijan matrice, dobija se iterativni korak za izračunavanje vrednosti parametara mreže:

$$(5.2.13) \quad \theta_{k+1} = \theta_k - [\nabla^2 q(\theta_k)]^{-1} \nabla q(\theta_k).$$

Iz (5.2.10) i (5.2.13) dobijamo da je smer traženja minimuma

$$(5.2.14) \quad s_k = -[\nabla^2 q(\theta_k)]^{-1} \nabla q(\theta_k) = -H^{-1}(\theta_k) \nabla q(\theta_k),$$

a to je upravo ono što je opisano u odeljku 4.

Ako je funkcija $q(\theta)$ kvadratnog oblika, Newtonova metoda definisana izrazom (5.2.13), izračunava njen minimum u jednoj iteraciji. Newtonova metoda ima kvadratnu konvergenciju, kao i konjugovane metode, ali je konvergencija znatno brža (umesto $n(\theta)$, potrebna je jedna). Uopšteno funkcija $q(\theta)$ je složenijeg oblika od kvadratnog, pa je potrebno više iteracija, ali ih je ipak manje nego kod metode najbržeg spusta i konjugovane gradijentne metode. Ali iz uslova (5.2.10) imamo da će smer traženja minimuma biti usmeren ka minimumu funkcije $q(\theta)$, samo ako je Hesijan matrica pozitivno definitna, a ona je pozitivno definitna samo za strogo konveksne funkcije. Pa ako kriterijumska funkcija nije strogo konveksna, metoda može divergirati. Zbog toga je prilikom upotrebe ove metode, veoma važno izabrati početne vrednosti parametara θ_0 , dovoljno blizu tačke minimuma, kako bi se osigurala pozitivna definitnost

Hesijan matrice. Za razliku od ove metode, metoda najbržeg spusta konvergira nezavisno od izbora početne tačke (uz uslov da je Jakobijan matrica svuda nesingularna), ali sa linearnom konvergentnošću.

Osim što ne osigurava pozitivnu definitnost Hesijana, u području daleko od minimuma, izraz (5.2.6) je i računski veoma zahtevan (potrebno je u svakoj iteraciji izračunati sve članove matrice $\nabla^2 r^*(\theta_k)$), pa bi ga iz tog razloga bilo dobro modifikovati. Pri tome, naravno, treba voditi računa da se ne naruše osobine kvadratne konvergentnosti, odnosno treba pronaći matricu koja dobro aproksimira Hesijan ali je i pozitivno definitna u oblasti vrednosti funkcije $q(\theta)$. U tom smislu se dobro pokazala Levenberg-Marquardtova metoda.

5.2.4 Levenberg-Marquardtova metoda

Zasniva se na predpostavci da se greška $r^*(\theta)$ u okolini tačke θ_k , može zadovoljavajuće aproksimirati s prva dva člana Tejlorovog reda:

$$(5.2.15) \quad r^*(\theta) \approx \tilde{r}^*(\theta) = r^*(\theta_k) + \nabla r^*(\theta_k) \cdot (\theta - \theta_k),$$

a umesto minimizacije funkcije $q(\theta)$, minimizuje se njena aproksimacija

$$(5.2.16) \quad \tilde{q}(\theta) \approx \frac{1}{2} \tilde{r}^{*T}(\theta) \cdot \tilde{r}^*(\theta).$$

Iz neophodnog uslova prvog reda ($\nabla \tilde{q}(\theta) = 0$), dobija se

$$(5.2.17) \quad J^T(\theta_k) \cdot J(\theta_k) \cdot (\theta - \theta_k) + J^T(\theta_k) \cdot r^*(\theta_k) = 0.$$

Uvrštavanjem (5.2.5) u (5.2.17) i dodavanjem koeficijenta učenja α_k i stavljanjem da je $\theta = \theta_{k+1}$, dobija se iterativni korak:

$$(5.2.18) \quad \theta_{k+1} = \theta_k - \alpha_k [J^T(\theta_k) \cdot J(\theta_k)]^{-1} J^T(\theta_k) \cdot r^*(\theta_k).$$

U literaturi se za $\alpha_k = 1$, ova metoda naziva Gauss-Newtonova, a za $\alpha_k < 1$ prigušena Gauss-Newtonova metoda. Upoređujući (5.2.13) i (5.2.18) vidimo da je Hesijan zamenjen matricom

$$(5.2.19) \quad \tilde{H}(\theta_k) = J^T(\theta_k) J(\theta_k).$$

Matrica \tilde{H} je pozitivno semidefinitna, čime je obezbeđeno da (5.2.18) ne divergira. Međutim ako merni podaci nisu dovoljno informativni ili neuronska mreža ima prevelik broj neurona, matrica \tilde{H} može postati loše uslovljena odnosno skoro singularna ili singularna, što daje kao posledicu numeričku nestabilnost metode. Kako bi se dobila pozitivno definitna matrica u čitavoj oblasti vrednosti kriterijumske funkcije, potrebno je matricu \tilde{H} , dodatno promeniti.

Tu je i još jedan problem koji je potrebno prevazići. Pri velikim promenama apsolutnih vrednosti parametara $\Delta\theta_k = \theta - \theta_k$, aproksimacija greške Tejlorovim redom prema izrazu (5.2.15), ne zadovoljava, pa smanjenje vrednosti funkcije $\tilde{q}(\theta)$ ne mora dovesti i do smanjenja vrednosti funkcije $q(\theta)$. U cilju poboljšanja uvedene Tejlorove aproksimacije greške odnosno povećanja vrednosti smanjenja

kriterijumske funkcije $q(\theta)$ smanjenjem aproksimacijske funkcije $\tilde{q}(\theta)$, Levenberg je 1944. godine predložio ograničavanje apsolutnih vrednosti promene parametara, te minimizaciju aproksimacijske funkcije uz uvedena ograničenja. Apsolutna vrednost promena parametara kao i vrednost greške, ograničavaju se primenom postupka najmanjih kvadrata za minimizaciju sledeće funkcije:

$$(5.2.20) \quad \check{q}(\theta) = \xi \cdot \tilde{q}(\theta) + \Delta\theta^T \cdot \begin{bmatrix} \zeta_1 & 0 & \cdots & 0 \\ 0 & \zeta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \zeta_{n(x)} \end{bmatrix} \cdot \Delta\theta,$$

pri čemu je ζ_i (pozitivan skalar) težinski koeficijent koji izražava relativan značaj ograničenja promene i -tog parametra, a ξ (takođe pozitivan skalar) izražava relativan značaj i greške i promena parametara u postupku minimizacije.

Iz uslova $\nabla\check{q}(\theta)=0$, dobija se izraz koji minimizira $\check{q}(\theta)$:

$$(5.2.21) \quad \check{H}(\theta_k) \cdot (\theta - \theta_k) + J^T(\theta_k) \cdot r^*(\theta_k) = 0,$$

pri čemu je $\check{H}(\theta_k)$ Levenbergova aproksimacijska matrica Hesijana

$$(5.2.22) \quad \check{H}(\theta_k) = J^T(\theta_k)J(\theta_k) + \frac{1}{\xi} \begin{bmatrix} \zeta_1 & 0 & \cdots & 0 \\ 0 & \zeta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \zeta_{n(\theta)} \end{bmatrix}.$$

Matrica $\check{H}(\theta_k)$ ima isti oblik kao i matrica $\tilde{H}(\theta_k)$, a razlikuju se samo članovi na glavnoj dijagonali, koji su uvećani proporcionalno vrednostima težinskih koeficijenata ζ_i . Aproksimacijska matrica $\check{H}(\theta_k)$ je simetrična, kao i Hesijan $H(\theta_k)$, a odgovarajućim izborom koeficijenata ζ_i i ξ se osigurava njena dobra uslvoljenost, odnosno pozitivna definitnost. Ova svojstva matrice $\check{H}(\theta_k)$, omogućuju primenu pojednostavljenih postupaka rešavanja sistema jednačina (5.2.21) bez neposrednog izračunavanja inverzne matrice. Rešavanjem sistema (5.2.21), dobija se vektor parametara koji minimizuje funkciju $\check{q}(\theta)$, a može se dokazati da odgovarajućim izborom koeficijenata ζ_i i ξ , isti vektor minimizuje i kriterijumsku funkciju $q(\theta)$.

Ako koeficijent $\xi \rightarrow \infty$, izraz (5.2.21) prelazi u Gauss-Newtonovu metodu (5.2.18), koji se prema tome može smatrati specijalnim slučajem Levenbergove metode.

Rešavanjem sistema jednačina (5.2.21) za ξ u okolini nule i traženjem izvoda dobijenog vektora parametara po koeficijentu ξ , uz $\xi=0$, dobija se

$$(5.2.23) \quad \frac{d\theta}{d\xi} \Big|_{\xi=0} = - \begin{bmatrix} \frac{1}{\zeta_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\zeta_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\zeta_{n(\theta)}} \end{bmatrix} J^T(\theta) r^*(\theta).$$

Iz (5.2.5) i (5.2.23) dobija se sledeći izraz za izračunavanje izvoda kriterijumske funkcije $q(\theta)$ po parametru ξ :

$$(5.2.24) \quad \frac{dq}{d\xi} \Big|_{\xi=0} = \frac{\partial q}{\partial \theta} \frac{d\theta}{d\xi} = - \left[J^T(\theta) r^*(\theta) \right]^T \begin{bmatrix} \frac{1}{\zeta_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\zeta_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\zeta_{n(\theta)}} \end{bmatrix} J^T(\theta) r^*(\theta).$$

Iz ovog izraza vidi se da je izvod funkcije q po koeficijentu ξ negativan (za $\xi=0$), tj. da se njena vrednost smanjuje. Prema tome postoji vrednost koeficijenta $\xi \in (0, \infty)$ uz koji će se vrednost kriterijumske funkcije smanjivati.

Teorijski, se najbolja vrednost koeficijenta ξ dobija rešavanjem jednačine:

$$(5.2.25) \quad \frac{dq(\theta_k, \xi)}{d\xi} = 0$$

Zbog složenosti rešavanja ove jednačine, u praksi se često koriste jednostavniji postupci za izračunavanje vrednosti koeficijenta ξ . Može se zadati i kao konstantna vrednost. Tada njegova vrednost mora biti dovoljno velika da se osigura pozitivna definitnost Levenbergove matrice u svakoj tački kriterijumske funkcije. Međutim, u tom slučaju se narušava kvadratna konvergentnost metode, što značajno povećava broj potrebnih iteracija. Zbog toga je koeficijent ξ najčešće skalarna funkcija, koja zavisi od lokalnih svojstava kriterijumske funkcije u tekućoj iteraciji.

Levenberg je predložio izračunavanje koeficijenta ξ , na osnovu sledećeg izraza

$$(5.2.26) \quad q(\theta_{k+1}, \xi) \approx q(\theta_k) + \xi \cdot \frac{dq(\theta_k, \xi)}{d\xi} \Big|_{\xi=0}$$

Ako predpostavimo da su početne vrednosti parametara θ_0 dobro odabrane, tako da je funkcija $q(\theta)$ približno kvadratnog oblika, tada će $q(\theta_{k+1}, \xi)$ biti male vrednosti. Njegovim izjednačavanjem sa nulom iz izraza (5.2.26), dobija se sledeći izraz za izračunavanje koeficijenta ξ :

$$(5.2.27) \quad \xi = -\frac{q(\theta_k)}{\left. \frac{dq}{d\xi} \right|_{\xi=0}}$$

Pri izvođenju izraza za izračunavanje koeficijenta ξ nisu uzimani u obzir vrednosti koeficijenata ζ_i , osim što je predpostavljeno da su veći od nule. Levenberg je predložio da bude $\zeta_i = 1$. Ako se uvesti u izraz (5.2.22), dobija se sledeći izraz za Levenbergovu matricu:

$$(5.2.28) \quad \check{H}(\theta_k) = J^T(\theta_k) \cdot J(\theta_k) + \mu I,$$

pri čemu je $\mu = \frac{1}{\xi}$.

Prema tome, Levenbergova aproksimacijska matrica $\check{H}(\theta)$ se dobija jednostavnim dodavanjem koeficijenata μ dijagonalnim članovima matrice $\tilde{H}(\theta)$. Zamenom Hesijan matrice i izrazu (5.2.13), Levenbergovom matricom (5.2.28) dobija se izraz za izračunavanje parametara mreže:

$$(5.2.29) \quad \theta_{k+1} = \theta_k - \check{H}^{-1}(\theta_k) J^T(\theta_k) r^*(\theta_k).$$

Iz izraza (5.2.24) i (5.2.27), uz $\zeta_i = 1$ i $\mu = \frac{1}{\xi}$, dobija se izraz za izračunavanje

koeficijenata:

$$(5.2.30) \quad \mu = \frac{r^{*T}(\theta_k) \cdot J(\theta_k) \cdot J^T(\theta_k) \cdot r^{*T}(\theta_k)}{q(\theta_k)}.$$

Za kriterijumsku funkciju, koja je složenijeg oblika od kvadratnog, predpostavka da je $q(\theta_{k+1}, \mu) = 0$ nije realna, pa su vrednosti koeficijenata μ dobijeni pomoću izraza (5.2.30) preveliki, tako da se metoda previše udaljava od Gauss-Newtonove metode, a približava metodi najbržeg spusta, pri čemu značajno gubi na brzini konvergencije.

Vrlo uspešan i najčešće korišćen, postupak određivanja koeficijenata μ , razradio je Marquardt i zbog toga se ovaj postupak u literaturi i naziva Levenberg-Marquardtova metoda. U svakoj iteraciji se koeficijent μ_k određuje drugim iterativnim postupkom. Početna vrednost koeficijenta μ_k se određuje množenjem njegove optimalne vrednosti iz prethodne iteracije koeficijentom smanjenja μ_d ($\mu_d < 1$) tj. $\mu_k = \mu_d \mu_{k-1}$, osim u prvoj iteraciji kada se koristi početna vrednost μ_0 , koju zadaje korisnik. Uz vrednost koeficijenta μ_k izračunava se i kriterijumska funkcija i upoređuje se sa njenom vrednošću u prethodnoj iteraciji. Ako se njena vrednost smanjila, vrednost koeficijenta μ_k se smatra optimalnom i prelazi se na sledeću iteraciju algoritma učenja mreže. Ako se vrednost kriterijumske funkcije povećala, vrednost koeficijenta μ_k se povećava iterativno, množeći s koeficijentom povećanja μ_i ($\mu_i > 1$) u svakom koraku, dok ne dođe do smanjenja vrednosti kriterijumske funkcije. Ovim postupkom podešavanja vrednosti koeficijenta μ_k , postiže se dobro približavanje metode, Gauss-Newtonovoj metodi uz istovremeno osiguranje njene konvergentnosti i numeričke stabilnosti. Marquardt je predložio sledeće vrednosti koeficijenata $\mu_0 = 0.001$, $\mu_d = 0.1$ i $\mu_i = 10$.

Za male vrednosti μ_k Levenberg-Marquardtova metoda se približava Gauss-Newtonovoj metodi, a za velike vrednosti metodi najbržeg spusta s koeficijentom učenja $\frac{1}{\mu_k}$. Prema tome, povećanjem vrednosti koeficijenta μ_k osigurava se konvergencija metode iz bilo koje početne tačke, koja može biti van oblasti konvergencije Gauss-Newtonove metode, dok se smanjenjem vrednosti u oblasti konvergencije Gauss-Newtonove metode osigurava kvadratna konvergencija. Na taj način se koriste dobra svojstva obe navedene metode, osiguravajući postupan prelaz iz jedne u drugu.

Neke od navedenih metoda i njihovih varijacija su nastale pre više decenija, a pojedine i pre više od jednog veka. Potvrdu njihovog kvaliteta daje i dugačak spisak programskih jezika i softvera u koje su implementirane. Tako na primer C, C++, C#, Delphi, Visual Basic, R programske jezice imaju ugrađene prethodno navedene algoritme za rešavanje nelinearnih problema najmanjih kvadrata. Tu su još i NMath softverski paket za Microsoft .NET Framework, Javanumerics, LMA-package i Apache Commons Math za Java programski jezik kao i procedure ugradene u SAS (Statistical Analysis System), a u koji su implementirani ovi algoritmi.

5.2.5 Primena neuronskih mreža u OCR tehnologiji

U Republičkom zavodu za statistiku je za obradu podataka u Popisu stanovništva 2011., kao i Popisu poljoprivrede 2012. godine, korišćen softverski paket eFlow, kompanije Top Image Systems. Sistem eFlow se sastoji iz više manjih paketa, koji zajedno obrađuju podatke sa upitnika, počevši od skeniranja upitnika do upisa podataka u bazu. Jedan od tih delova je zadužen i za prepoznavanje rukom pisanog teksta. Prepoznavanje teksta (bilo da je mašinski ili rukom pisan) se vrši pomoću tehnologija za optičko i inteligentno prepoznavanje teksta (eng. OCR, ICR i IWR).

OCR tehnologija podrazumeva prevod teksta, koji je u obliku slike, dobijene skeniranjem određenog dokumenta u editabilni tekst ili u kodnu šemu predstavljenu u ASCII ili Unicode kodu. Implementacija OCR-a je patentirana još 1929. godine u Nemačkoj (Gustav Tauschek) i 1933. u USA (Paul W. Handel). Danas većina skenera ima implementiran neki oblik OCR tehnologije, koji omogućava korisniku izmenu skeniranih slika u digitalnom obliku. Prvobitno je OCR tehnologija zahtevala standardne fontove za prepoznavanje znakova. Moderne implementacije OCR tehnologije, za prepoznavanje znakova, ne zahtevaju takav standardizovani font, ali oni ipak utiču na kvalitet i poboljšavaju prepoznavanje. Savremeni OCR endžini (softverska paketi za učenje neuronskih mreža) koriste neuronske mreže za kategorizaciju skupa mogućih znakova koji su identifikovani na slici. Neuronske mreže omogućuju poboljšano prepoznavanje proizvoljnih stilova fontova (nasuprot standardnim fontovima). U cilju učenja tj. treniranja i testiranja mreže potrebno je postaviti početne informacije (prvi skup podataka) na osnovu kojih mreža radi i ispunjava zadatke.

Konkretno u eFlow sistemu, postoji deo koji se zove OCR Visual Designer. U njemu se definišu OCR endžini (eng. OCR engines) koji se koriste za prepoznavanje karaktera, kao i testiranje i fino podešavanje parametara. U ovom slučaju bilo je potrebno zadati ulazne informacije o sprskim slovima (ćiriličnim i latiničnim), a zatim podešavanjem parametara poboljšavati performanse sistema. Ulazne informacije podrazumevaju konkretnе primere popunjениh upitnika (i ćirilicom i latinicom). Preko određene aplikacije i skenera, upitnici se skeniraju, a zatim se dobijene slike (ovde je konkretnо korišćen TIFF format) obrađuju i definišu parametri za prepoznavanje znakova. U toku rada, ovaj samoučeći sistem poboljšava svoje performanse, a rezultat je visok procenat uspešnog prepoznavanja rukom pisanih teksta.

6. Literatura

1. Optimization Theory And Methods – Nonlinear programming; Wenyu Sun, Ya-Xian Yuan (2006)
2. Numerical Methods for Least Squares Problems; Ake Bjork (1996)
3. Numerical Optimization; Jorge Nocedal, Stephen J. Wright (2006)
4. Advanced kalman filtering, least-squares and modeling: A practical handbook; Bruce P. Gibbs (2011)
5. Numerička analiza; Vjeran Hari ,Zlatko Drmač, Mladen Rogina, Miljenko Marušić, Saša Singer, Sanja Singer (2003)
6. Iterative Methods for Optimization; C.T.Kelly (1999)
7. Methods for Non-Linear Least Squares problems; K.Madsen, H.B. Nielsen, O. Tingleff (2004)
8. Inteligentni sustavi upravljanja: Neuronske mreže, evolucijski i genetički algoritmi - skripta; Ivan Petrović, Mato Baotić, Neđeljko Perić (2011)
9. Visual Designer User Guide – eFlow4; TiS Technologies, Inc. (2007)
10. Optical Character Recognition with Neural Networks; Trevor Handley, The University of Kansas (2011)