

Metod najmanjih kvadrata i fitovanje krivih

Mladenovski Zoran 2017/0632

Rančić Jakov 2017/0363

Sadržaj

1	Uvod	3
2	Regresiona analiza	4
2.1	Korelisane i nekorelisane veličine	4
2.2	Podela regresione analize	4
2.3	Jednostavna (jednostruka) linearna regresija	5
2.4	Regresioni model	5
3	Metoda najmanjih kvadrata	6
3.1	Istorijat metode najmanjih kvadrata	6
3.2	Uvod u metodu	7
3.3	Linija najmanjih kvadrata	8
3.4	Procena parametara modela regresije	9
3.5	Pregled postupaka za rešavanje sistema linearnih jednačina	13
4	Nelinearni problemi i metodi za njihovo rešavanje	14
4.1	Uvod	14
4.2	Određivanje (traženje) krive	15
4.3	Njutnova metoda	16
4.4	Gaus - Njutnova metoda	17
4.5	Levenberg - Marquardt-ov metod	18
5	Primer rešavanja problema u telekomunikacijama metodom najmanjih kvadrata	20
5.1	Polinomska aproksimacija	20
5.2	Linearna predikcija signala	23
5.3	Popravljanje signala oštećenog greškom	25
5.4	Stvaranje nedostataka u signalu usled prolaženja kroz HPF (High pass filter) i LPF (Low pass filter)	28
	Literatura	30

1 Uvod

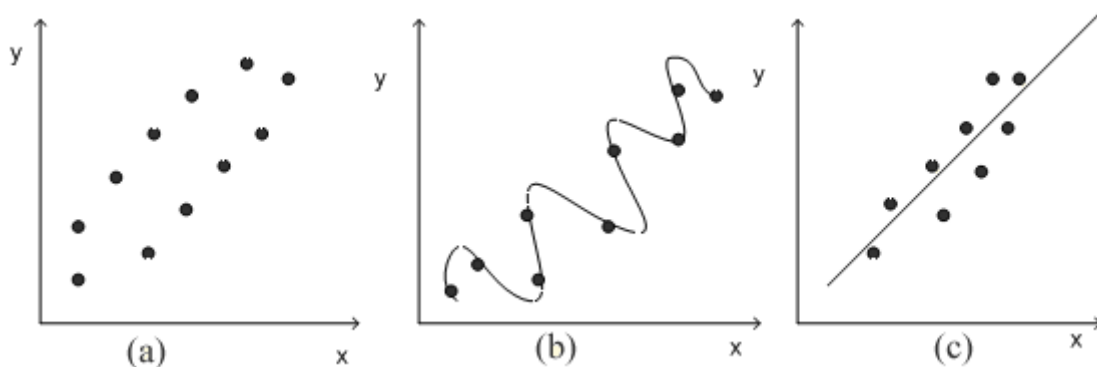
U inženjerskoj praksi, što eksperimentalnoj, što komercijalnoj, jako važnu ulogu imaju ogledi i merenja, koja daju obilne rezultate o onome što je urađeno. Da bi imali ikakve koristi od tih rezultata, mora postojati adekvatan matematički aparat (alat) koji na osnovu tih izmerenih vrednosti raznih fizičkih veličina daje, ljudima "čitljive" i razumljive vrednosti. Tu se postavlja pitanje: „A koji je to matematički aparat adekvatan?“ Koji je kriterijum adekvatnosti matematičkog aparata? Jednim delom odgovora na to pitanja se bavi oblast matematike o kojoj je reč u ovom semestralnom radu. Takođe, poznato je da jedna slika govori više od hiljadu reči i da su dijagrami ljudima mnogo čitljiviji i značajniji za razumevanje samih fizičkih procesa na kojima se radi, od prikaza raznih vrednosti u tabeli, na primer.

Grafici su veoma značajni ljudima jer se sa njih lako i brzo može videti kakvu zavisnost imaju prikazani podaci, dok preciznost (finoća) grafika zavisi od broja ucrtanih tačaka, a to zavisi od samog obima oglada ili merenja. Vratimo se na zavisnost koja se posmatra na grafiku. Ona se može odrediti računanjem nepoznatih parametara (npr. koeficijenata prave), a za to nam je potrebno najmanje dve tačke na prikazanom dijagramu.

S obzirom na to da raspolažemo većim brojem tačaka, suočeni smo sa velikim problemom na koji način odrediti nepoznate parametre i tako utvrditi vezu između prikazanih veličina dobijenih ogledom. Tehnički, ovaj problem se svodi na problem „kako najbolje postaviti pravu kroz tačke koje nisu na istom pravcu, tako da zavisnost koju tačke predstavljaju najbolje bude opisana tom pravom“.

Pretpostavimo da su nam dati podaci $(X_i, Y_i), i = 1, 2, \dots, n$, dobijeni nekim merenjem ili preračunavanjem podataka merenja. Bez preteranog ulaženja u suštinu samog merenja, vidi se da je merenje obavljeno neprecizno, po tome što postoje velika odstupanja u vrednostima između uzastopnih tačaka na grafiku. I inače, poznato je da su podaci dobijeni merenjem ili ogledom često "rastureni" ("raštrkani") nezavisno od toga koliko je merenje ili ogled ozbiljno organizovan, te su odličan primer odstupanja, koja, jednostavno, zbog nesavršenosti merne tehnike i aparature, ne možemo zaobići.

O polinomskoj interpolaciji nema mnogo smisla govoriti u ovakvim situacijama, jer nije prikladna metoda za ovakve primene. Da bi smo bolje razumeli razlog, zbog koga polinomska interpolacija nije prikladna, pogledaćemo grafike na slici 1.



Slika 1: 1.a – "rastureni" podaci dobijeni merenjem; 1.b – Polinomska kriva koja osciluje izvan granice dobijenih podataka; 1.c – Aproksimativna kriva dobijenih podataka

Pogledajmo ponovo, malo detaljnije, sliku 1.a. Sa grafika vidimo trend porasta merene veličine, tj. većim vrednostima X odgovaraju veće vrednosti Y . Ako date podatke fitujemo krivom koju predstavlja polinom (u ovom konkretnom slučaju polinomom osmog stepena), vidimo da kriva tačno prolazi kroz eksperimentalno dobijene tačke, ali ima primetnih oscilacija koje su posledica mernih nesigurnosti samog merenja, slika 1.b. Pritom, tako dobijena kriva ide izvan granica koje određuju dobijene tačke, što obrađene podatke čini još manje preciznim i još manje upotrebljivim.

Dakle, potreban nam je bolji, a samim tim i kompleksniji, način određivanja najprikladnije (optimalne) funkcije koja odgovara eksperimentalno dobijenim vrednostima neke merene veličine. Grafik jedne takve, optimalne, funkcije koja odgovara mernim podacima sa slike 1.a, data je na slici 1.c, na kojoj vidimo da fitovana (optimalna) kriva odgovara obliku i generalnom trendu porasta vrednosti merenih podataka.

Jedna od najpoznatijih i najkorišćenijih metoda za pronalaženje optimalnih krivih, u ovakvim i sličnim situacijama, je metoda najmanjih kvadrata, koja je standardan pristup u regresionoj analizi, pri aproksimaciji rešenja određenog sistema jednačina (sistem jednačina u kome ima više jednačina nego nepoznatih) minimizacijom suma kvadrata ostataka rezultata u svakoj jednačini posebno.

2 Regresiona analiza

2.1 Korelisane i nekorelisane veličine

Pre svega, treba razjasniti da kada imamo dve fizičke veličine, one mogu biti ili nezavisne ili zavisne. Nezavisne su one fizičke veličine kod kojih na osnovu poznavanja jedne ne možemo zaključiti ništa o onoj drugoj. Primer bi bile godine i telesna masa nekog čoveka. Poznavanjem godina nekog čoveka, ne možemo zaključiti ništa o njegovoj telesnoj masi, niti obrnuto, na osnovu telesne mase, ne možemo znati koliko čovek ima godina.

Zavisne su one veličine kod kojih jedna tačno određuje ili ograničava onu drugu. Za to bi važili sledeći primeri: Na osnovu temperature gasa, pri konstantnoj zapremini, tačno možemo odrediti pritisak koji vlada u posudi u kojoj se gas nalazi. Ili, ako posmatramo kružnicu čiji je centar u koordinatnom početku, i ako posmatramo jednu tačku jedne kordinatne ose, njoj odgovaraju tačno dve tačke na kružnici čije su druge kordinate istog modula ali suprotnog znaka. Tako smo poznavanjem jedne veličine, ograničili drugu do reda znaka.

2.2 Podela regresione analize

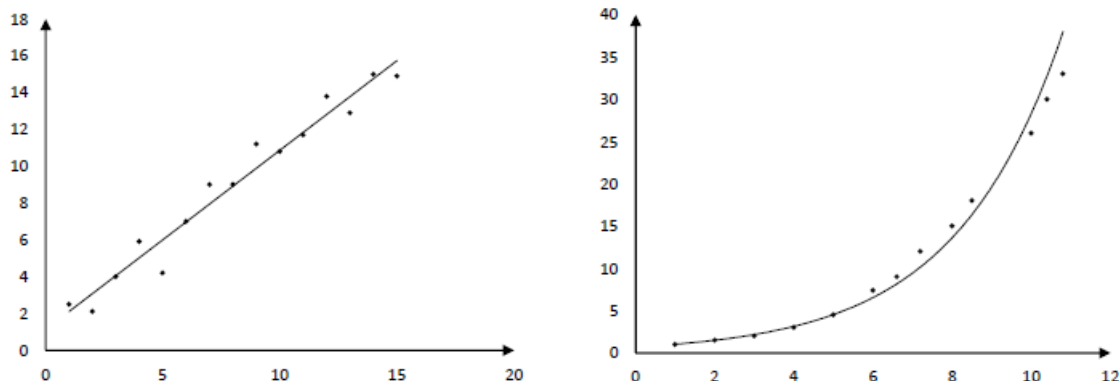
Regresiona analiza nam govori o povezanosti nezavisno promenljive (x), sa zavisno promenljivom (y , odnosno $f(x)$), tako što pokušava da predvidi koju će vrednost poprimiti zavisno promenljiva (y), na osnovu poznate vrednosti nezavisno promenljive (x). Drugim rečima, regresiona analiza preciznije objašnjava kako se vrednosti zavisno promenljive mogu menjati u slučaju kada se vrednosti nezavisno promenljive menjaju.

U slučaju da imamo više nezavisno promenljivih, a samo jedna se menja, dok su ostale konstantne, regresionu analizu nazivamo **jednostavnom** ili **jednostrukom regresionom analizom**. Sa druge strane, u slučaju da imamo više promenljivih, od kojih se više njih menja, regresionu analizu nazivamo **višestrukom regresionom analizom**. Pritom, obe regresione analize (i jednostruka i višestruka), gledano iz aspekta linaernosti, mogu biti **linearne** ili **nelinearne**.

2.3 Jednostavna (jednostruka) linearna regresija

Najosnovniji oblik regresione analize je jednostavna linearna regresija. Ona dovodi u vezu nezavisno i zavisno promenljivu, pri čemu odabira najbolju moguću **linearnu** zavisnost.

Pogledajmo grafike na slici 2. Zapažamo da na oba grafika, kada bi se odstranio određen broj tačaka, može da se ostvari i linearna i nelinearna zavisnost.



Slika 2: Primer linearne i nelinearne zavisnosti promenljivih

Ipak, kada se sve tačke sagledaju, vidimo da na prvoj ipak može da se ostvari samo linearna, a na drugoj samo nelinearna zavisnost. S obzirom na to da se prilikom ozbiljnih merenja ili oglada prikupi veliki broj padataka, kao i na to da je svet oko nas najčešće nelinearan, takvu, jednostavnu linearnu zavisnost je ponekad teško pronaći.

Kako bi smo prevazišli ovaj problem, moramo dopustiti posmatranje problema na višem nivou. To znači da dozvolimo posmatranje više nezavisno promenljivih, od kojih svaka utiče na zavisno promenljivu, ali na različit način. Treba napomenuti i to da taj uticaj može biti linearan ili nelinearan. Sada se postavlja pitanje, kako na osnovu svih tih nezavisno promenljivih formirati traženu matematičku zavisnost (funkciju optimalne krive na grafiku).

2.4 Regresioni model

Regresioni model matematičkim zapisom daje opis odnosa dve ili više promenljivih (nezavisnih i zavisne). Razlikujemo **deterministički** i **statistički** regresioni model.

Kada je reč o **determinističkom** modelu, njegov zapis možemo dati sledećim izrazom:

$$y = a_0 + a_1x \quad (1)$$

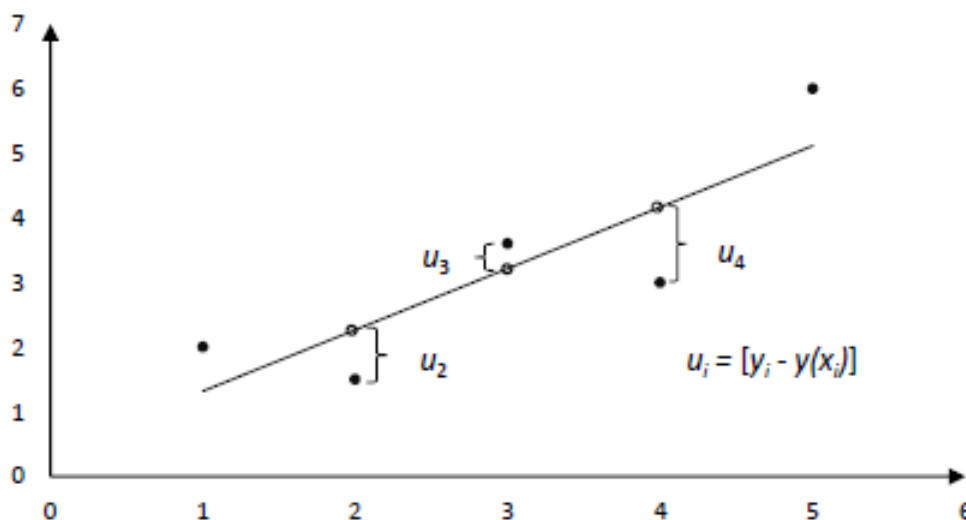
Ovaj model opisuje tačnu zavisnost promenljivih x i y , odnosno dokazuje kako je promenljiva y određena tačno jednom promenljivom x .

Pogledajmo sada **statistički** model, zadat sledećim izrazom:

$$y = a_0 + a_1x + u \quad (2)$$

Primetimo da, pored uobičajenih elemenata, figuriše i promenljiva u , koja predstavlja mernu nesigurnost. Ona obuhvata sve one promenljive koje nisu direktno sadržane u modelu, ali utiču na vrednost zavisno promenljive y i njene slučajne promene. Naravno, sve te promenljive su manje uticajne na promenljivu y , u odnosu na glavnu nezavisno promenljivu x , ali ipak doprinose, u kompoziciji, slučajnom kolebanju njene vrednosti. Sa druge strane, merna nesigurnost u , prikazuje razliku između eksperimentalno izmerenih ili utvrđenih vrednosti promenljive veličine i

teorijskog predviđanja vrednosti te promenljive veličine. Kako bi lakše razumeli i vizualizovali promenljivu u – mernu nesigurnost, pogledajmo sliku 3. Na horizontalnoj osi pratimo vrednosti koje uzima nezavisno promenljiva x , dok na vertikalnoj osi pratimo vrednosti koje uzima zavisno promenljiva y .



Slika 3: Grafički prikaz mernih nesigurnosti

Pokazuje se da primenom **metoda najmanjih kvadrata** dobijamo najmanji uticaj merne nesigurnosti u , na tačnost zavisno promenljive y . Kako bi smo bolje upoznali samu metodu, pogledajmo, pre svega njen istorijat.

3 Metoda najmanjih kvadrata

3.1 Istorijat metode najmanjih kvadrata

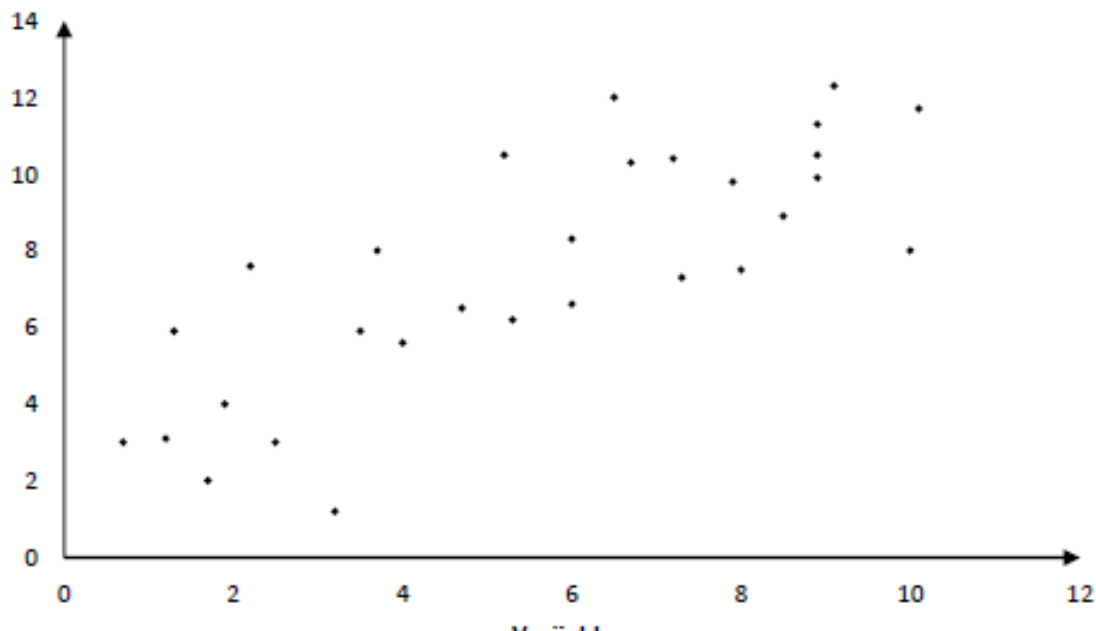
Najraniji oblik regresije dao je veliki francuski matematičar Pjer - Simon Laplas (fran. Pierre - Simon Laplace), 1799. godine kada se bavio principom minimizacije suma apsolutnih grešaka, što se formalno matematički može iskazati kao $\sum_{i=1}^m |r_i| \rightarrow 0$, sa dodatnim uslovom da broj grešaka bude jednak nuli. Pokazalo se, da pod tim uslovima, rešenje x mora mora zadovoljiti tačno n od m jednačina.

Međutim, nemački matematičar Karl Fridrih Gaus (nem. Johann Carl Friedrich Gauß), je tvrdio kako su, po načelima verovatnoće, veće ili manje greške jednako moguće i to u svim jednačinama. Po njegovom mišljenju, bilo je očigledno da se rešenje koje zadovoljava tačno n jednačina, treba smatrati manje preciznim u skladu sa zakonima varovatnoće. Još 1795. godine, Gauss je došao do osnovnih metoda najmanjih kvadrata, iako to nije objavio sve do 1809. godine, kada je metoda izašla u okviru njegovih radova o nebeskoj mehanici, pod naslovom *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium*. Međutim, za vreme od kada je Gaus otkrio metodu i kada ju je objavio, do slične ideje i metode došao je i francuski matematičar Anri - Mari Legendre, koji je 1805. godine izdao knjigu pod naslovom *Nouvelles methods pour la determination des orbites des cometes*, sa dodatkom *Sour la metode des moindres quarres*. Budući da su obe metode bile objavljene u slično vreme, između Gausa i Legendrea su se razvile žustre i neprijatne polemike oko prava i zasluga za otkriće metode. Veliki broj istoričara i matematičara se složio da prvenstvo nad zaslugama za otkriće metode ima Gaus, i to potkrepljuju poznatim primerom njegovog uspešnog predviđanja putanje asteroida

Ceres, 1801. godine. Gaus je, pomoću podataka koje je sakupio italijanski astronom Piaćí (it. Piazzi), četrdesetogodišnjim praćenjem kretanja asteroidea *Ceres*, predvideo kada će on opet doći u vidno polje ljudi na zemlji. Metodom najmanjih kvadrata, Gaus je uradio proračune i na osnovu njih ponovo locirao asteroid. Nakon tog događaja, metoda najmanjih kvadrata postala je jedna od najkorišćenijih i najpoznatijih metoda, tj. postupaka, za analiziranje astronomskih i geodetskih podataka.

3.2 Uvod u metodu

Metoda najmanjih kvadrata jedna je od najvažnijih metoda za obradu eksperimentalno dobijenih rezultata. Regresionim jednačinama, regresionom analizom, nastoji se da se u dijagram rasipanja rezultata ucrtta optimalna kriva, tj. pravac regresije koji će najbolje opisati odnos posmatranih promenljivih. Na slici 4 vidimo jedan takav dijagram pre povlačenja pravca regresije.

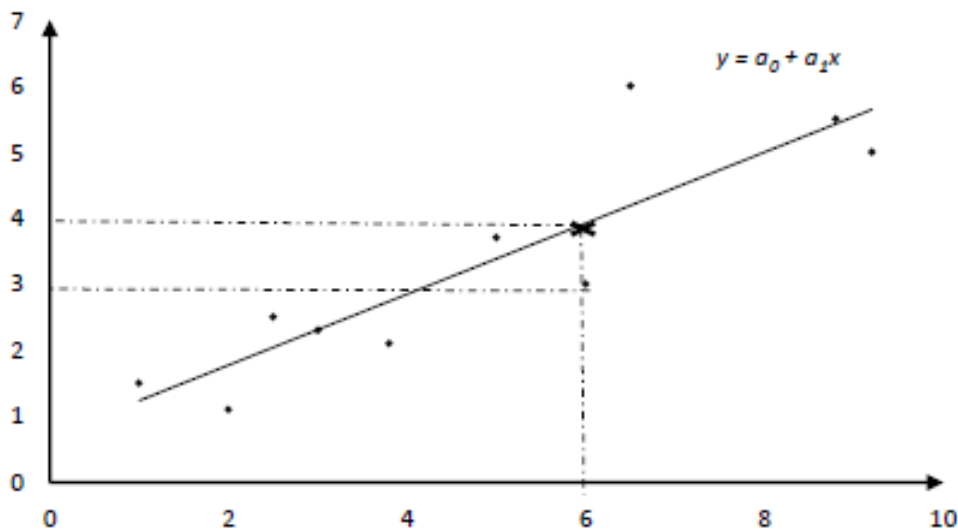


Slika 4: Dijagram rasipanja

Ako u dijagramu prikažemo sve podatke dobijene eksperimentalnim putem, dobili bi smo raspored tačkaka kroz koji je moguće ucrtati beskonačno mnogo pravaca. Svaki od pravaca bio bi određen odgovarajućom jednačinom sa sopstvenim vrednostima parametara a_0 i a_1 , iz jednačine (2).

3.3 Linija najmanjih kvadrata

Ako su koordinate i -te tačke (x_i, y_i) , a oblik aproksimiranog pravca je $y = a_0 + a_1x$, tada je y koordinata najbolje aproksimiranog pravca, za x_i jednaka $y_i = a_0 + a_1x_i$. Grafički prikaz ove postavke vidimo na slici 5.



Slika 5: Dijagram rasipanja sa ucrtanim regresionim pravcem

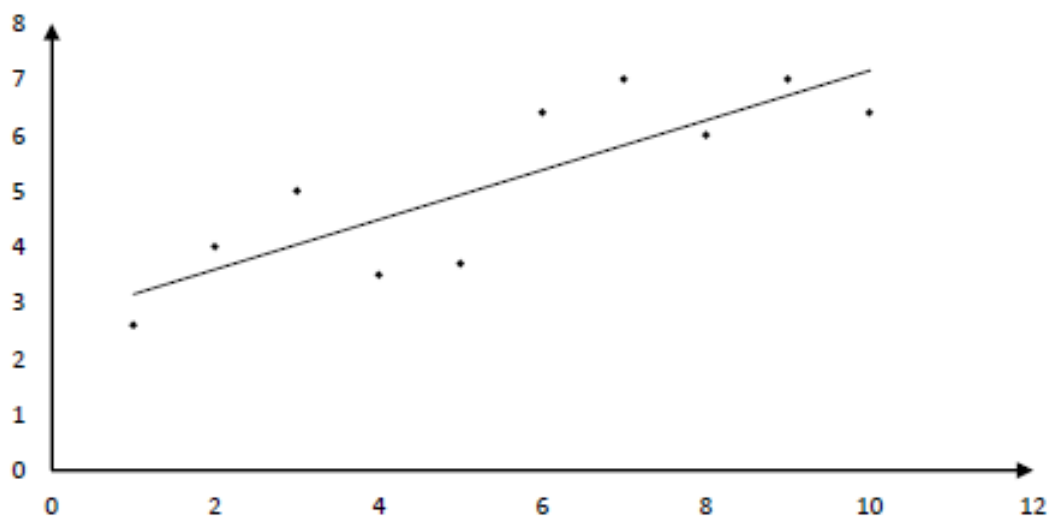
Sa slike se može videti kako za vrednosti x_i nezavisno promenljive x , zavisno promenljiva y_i dobija vrednosti $a_0 + a_1x_i$. Ono što je za analizu bitno je $y_i - (a_0 + a_1x_i)$, jer ova razlika predstavlja ostatak. Postavlja se pitanje (problem), kako odrediti, ili proceniti parametre a_0 i a_1 tako da taj ostatak bude minimalan, tj. da dobijemo optimalnu pravku koja na najbolji mogući način aproksimira realnost (izmerene vrednosti).

Uzmimo slučaj da smo nekako (nebitno kako) došli do odgovarajućih parametara a_0 i a_1 , tako da $y_i - (a_0 + a_1x_i) \rightarrow 0$, ni tada ne možemo odrediti optimalni pravac iako je data razlika minimizovana jer, statistički gledano, ostaci mogu biti i pozitivni i negativni sa istom verovatnoćom, što znači da će postojati slučaj u kome će se u zbiru skratiti. Dakle, moramo baratati sa nenegativnim vrednostima, ali je sada pitanje kako do njih doći, jer postoji više načina. Ako primenimo funkciju apsolutne vrednosti, dobićemo funkciju koja nije diferencijabilna u jednoj tački, što znači da nećemo moći uvek da primenimo tekovine matematičke analize, što je vrlo nepovoljno. Zato se pribegava korišćenju kvadratne funkcije. Pre svega, uvek je diferencijabilna, a pored toga, rezultati su nenegativne vrednosti, baš kako nam je i potrebno.

Pretpostavimo da imamo eksperimentalno dobijene podatke $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$. Parametre a_0 i a_1 , u tom slučaju, procenjujemo metodom najmanjih kvadrata, minimizacijom sume kvadrata ostataka. Pritom mora da važi sledeća jednakost:

$$f(a_0, a_1) = \sum_{i=1}^n [y_i - (a_0 + a_1x_i)]^2 = \min_{(a_0, a_1) \in \mathbb{R}^2} \sum_{i=1}^n [y_i - (a_0 + a_1x_i)]^2 = \min_{(a_0, a_1) \in \mathbb{R}^2} f(a_0, a_1) \quad (3)$$

Najbolja procena (nepoznatog) optimalnog pravca regresije, $y = a_0 + a_1x_0$ data je na slici 6.



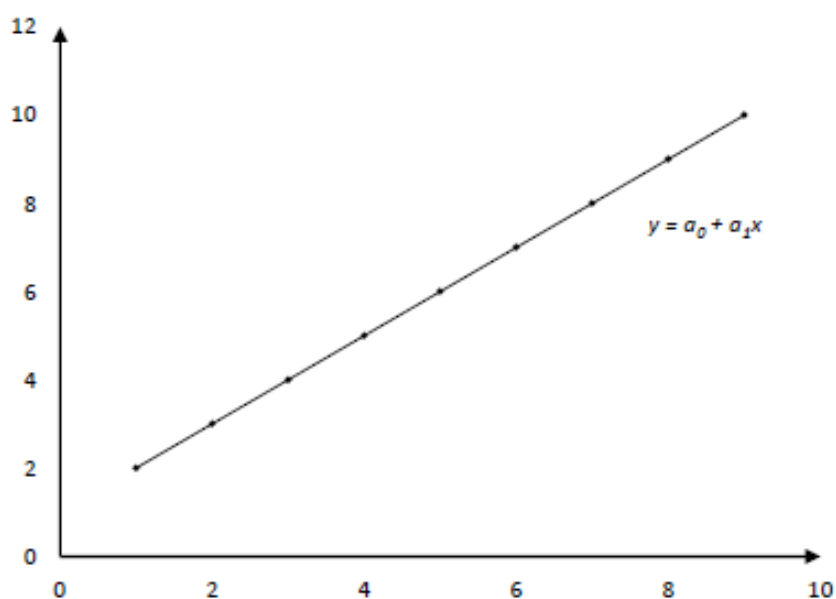
Slika 6: Optimalna prava regresije sa procenjenim parametrima

3.4 Procena parametara modela regresije

Pretpostavimo da smo izmerili x i y , odnosno njihove približne vrednosti, i to više puta. Tako samo dobili seriju mernih rezultata $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$, čijim ucrtavanjem na grafik dobijamo dijagram rasipanja, pa možemo posmatrati kakva je zavisnost između merenih veličina. Ukoliko možemo pretpostaviti linearnu zavisnost, vrednosti možemo povezati izrazom:

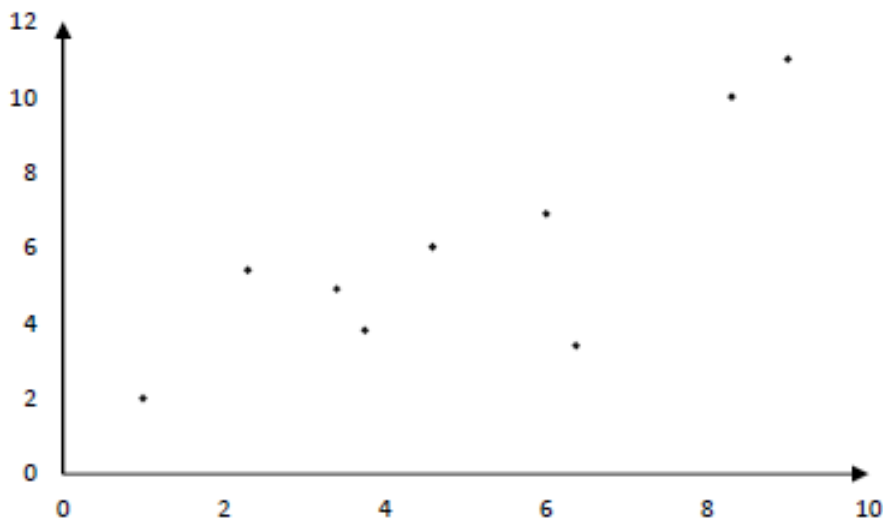
$$f(a_0, a_1) \equiv y = a_0 + a_1x \quad (4)$$

U idealnom slučaju, sve tačke (parovi tačaka) dobijeni merenjem, ležaće na istom pravcu, što možemo videti na slici 7.



Slika 7: Idealan slučaj – tačke leže na jednom pravcu

Međutim, ovaj slučaj je najmanje verovatan u realnosti, a i kada bi se dobio ovakav grafik, to bi značilo da u našem merenju ne postoje greške, što je statistički nemoguće. Nasuprot tome, u realnim i obimnim merenjima, prilikom ucrtavanja podataka na grafik, skoro obavezno se dobije da tačke nisu ravnomerno raspoređene po liniji, što vidimo na slici 8.



Slika 8: Realan slučaj – "oblak" tačaka

Kako bi smo odredili najbolji mogući pravac, greške i njihov uticaj, kao i merne nesigurnosti, moramo svesti na minimum. Zato će koeficijenti a_0 i a_1 , koje ćemo pronaći, biti oni najverovatniji. Odgovarajući pravac $y = a_0 + a_1x_i$ je najverovatniji pravac regresije ako važi da je suma kvadratnog odstupanja minimalna:

$$f(a_0, a_1) = \sum_{i=1}^n [y_i - (a_0 + a_1x_i)]^2 = \min \quad (5)$$

gde je n serijski broj merenja, a $[y_n - (a_0 + a_1x_n)]$ ostatak pri konkretnom merenju u celoj seriji. Dakle tražimo vrednost x koja minimizuje sumu kvadratnog odstupanja.

Ako se podsetimo matematičke analize, znamo da funkcija dostiže svoj maksimum onda kada joj je prvi izvod jednak 0 i kada je drugi izvod pozitivan (funkcija konveksna). Da je funkcija proizvoljna, postojali bi i drugi slučajevi, ali pošto smo uzeli kvadratnu funkciju ostatka, ovo je jedini slučaj i zapravo sada vidimo zašto je bilo toliko birno ne uzeti funkciju apsolutne vrednosti. Saglasno tome, potreban i dovoljan uslov za funkciju više promenljivih je:

$$\frac{\partial f(a_0, a_1)}{\partial a_0} = 0, \frac{\partial f(a_0, a_1)}{\partial a_1} = 0 \quad (6)$$

Rešavanjem parcijalnih izvoda po a_0 i a_1 , dobijamo sledeće jednačine:

$$\frac{\partial f(a_0, a_1)}{\partial a_0} = \sum_{i=1}^n 2[y_i - (a_0 + a_1x_i)] = 0 \quad (7)$$

$$\frac{\partial f(a_0, a_1)}{\partial a_1} = \sum_{i=1}^n 2[y_i - (a_0 + a_1x_i)]x_i = 0 \quad (8)$$

Da bi smo došli do koeficijenta a_0 i a_1 , potrebno je rešiti sledeći sistem jednačina:

$$\sum_{i=1}^n [y_i - (a_0 + a_1 x_i)] = 0 \quad (9)$$

$$\sum_{i=1}^n [y_i - (a_0 + a_1 x_i)] x_i = 0 \quad (10)$$

Iz jednačine (9) izražavamo koeficijent a_0 :

$$a_0 = \frac{\sum_{i=1}^n y_i - a_1 \sum_{i=1}^n x_i}{n} \quad (11)$$

Zatim, a_0 uvrštavamo u jednačinu (10) i dobijamo:

$$\sum_{i=1}^n x_i y_i - a_1 \sum_{i=1}^n x_i^2 - \frac{\sum_{i=1}^n y_i - a_1 \sum_{i=1}^n x_i}{n} \sum_{i=1}^n x_i = 0 \quad (12)$$

što množenjem sa n daje:

$$n \sum_{i=1}^n x_i y_i - a_1 n \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n y_i + a_1 \left(\sum_{i=1}^n x_i \right)^2 = 0 \quad (13)$$

Daljim sređivanjem izražavamo koeficijent a_1 :

$$a_1 n \sum_{i=1}^n x_i^2 - a_1 \left(\sum_{i=1}^n x_i \right)^2 = n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \quad (14)$$

$$a_1 \left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] = n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \quad (15)$$

Pa je konačan izraz za a_1 :

$$a_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (16)$$

Kada se ovaj izraz vrati u izraz za računanje a_0 dobijamo:

$$a_0 = \frac{\sum_{i=1}^n y_i - \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \sum_{i=1}^n x_i}{n} \quad (17)$$

$$a_0 = \frac{\sum_{i=1}^n y_i \left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] - \left(n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \quad (18)$$

$$a_0 = \frac{\sum_{i=1}^n y_i \left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] - \left(n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i \right) \sum_{i=1}^n x_i}{n \left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right]} \quad (19)$$

$$a_0 = \frac{n \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - (\sum_{i=1}^n x_i)^2 \sum_{i=1}^n y_i - n \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i + (\sum_{i=1}^n x_i)^2 \sum_{i=1}^n y_i}{n \left[n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 \right]} \quad (20)$$

$$a_0 = \frac{n (\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i)}{n \left[n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2 \right]} \quad (21)$$

Najzad, dobija se konačni oblik izraza za a_0 :

$$a_0 = \frac{\sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (22)$$

S obzirom da smo iz uslova ekstremuma dobili dve linearne jednačine sa dve nepoznate (a_0, a_1), radi jednostavnijeg zapisa i preglednosti, ovaj sistem možemo rešiti i u matricnom obliku:

$$\mathbf{H} \vec{a} = \vec{b} \Leftrightarrow \begin{bmatrix} H_{0,0} & H_{0,1} \\ H_{1,0} & H_{1,1} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \quad (23)$$

Pri čemu uzimamo u obzir da je:

$$\mathbf{H}_{i,j} = \mathbf{H}_{j,i} = \sum_{n=1}^N \mathbf{x}_n^{i+j} \quad \mathbf{b}_i = \sum_{n=1}^N \mathbf{x}_n^i y_n \quad \mathbf{i}, \mathbf{j} = 0, 1 \dots \quad (24)$$

Ako bolje pogledamo matricu \mathbf{H} , primetićemo da je simetrična (Hessian, Hesse-ova matrica). Navodimo opštu definiciju Hesse-ove matrice: Za skalarnu funkciju f , od n promenljivih, koja poseduje sve parcijalne izvode drugog reda (u tački X iz svog domena), odgovarajuća Hesse-ova matrica (u toj tački) je matrica $\mathbf{H} = \mathbf{H}(\mathbf{f})\mathbf{X}$, koja na preseku i -te vrste i j -te kolone sadrži vrednost $\frac{\partial^2 \mathbf{X}}{\partial x_i \partial y_j}$. Shodno definiciji, funkcija f u stacionarnoj tački X_0 ima lokalni minimum, ako su svi minorni Hesse-ove matrice $\mathbf{H}(\mathbf{f})(\mathbf{X}_0)$ pozitivni.

Zbog toga je rešenje date matricne jednačine dato izrazom:

$$\vec{a} = \mathbf{H}^{-1} \vec{b} \quad (25)$$

pri čemu je \mathbf{H}^{-1} inverzna matrica, matrice \mathbf{H} , data izrazom:

$$\mathbf{H}^{-1} = \frac{1}{\det \mathbf{H}} \mathbf{H}^* \quad (26)$$

\mathbf{H}^* je adjungovana matrica matrice \mathbf{H} , koju dobijamo kada na glavnoj dijagonali zamenimo mesta članova, dok članovima na sporednoj dijagonali zamenimo mesta i promenimo znak:

$$\mathbf{H}^{-1} = \frac{1}{\det H} \begin{bmatrix} H_{1,1} & -H_{1,0} \\ -H_{0,1} & H_{0,0} \end{bmatrix} \quad (27)$$

Prema tome, dobija se:

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \frac{1}{\det H} \begin{bmatrix} H_{1,1} & -H_{1,0} \\ -H_{0,1} & H_{0,0} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \quad (28)$$

Najzad, nepoznati parametri, u ovom slučaju, glase:

$$a_0 = \frac{\mathbf{H}_{1,1}\mathbf{b}_0 - \mathbf{H}_{1,0}\mathbf{b}_1}{\mathbf{H}_{0,0}\mathbf{H}_{1,1} - \mathbf{H}_{0,1}\mathbf{H}_{1,0}} \quad (29)$$

$$a_1 = \frac{\mathbf{H}_{0,0}\mathbf{b}_1 - \mathbf{H}_{0,1}\mathbf{b}_0}{\mathbf{H}_{0,0}\mathbf{H}_{1,1} - \mathbf{H}_{0,1}\mathbf{H}_{1,0}} \quad (30)$$

Ako uvrstimo prave vrednosti za $\mathbf{H}_{i,j}$ i b_i , dobićemo izraze za a_0 i a_1 , iste kao što smo dobili prethodnim izvođenjem iz uslova ekstremuma funkcije više promenljivih, parcijalnim izvodima.

3.5 Pregled postupaka za rešavanje sistema linearnih jednačina

Rešavanje sistema linearnih jednačina i metod najmanjih kvadrata primenjen na linearne funkcije su suštinski dva različita postupka, ali se svode na isto – rešavanje sistema jednačina. Pošto je ovde reč o linearnim funkcijama i njihovom aproksimiranju, i sistemi jednačina koje treba rešiti su linearni, pa dobijamo da se ovaj postupak svodi na rešavanje sistema linearnih jednačina.

Sistem linearnih jednačina zapisaćemo u matričnom obliku, na sledeći način:

$$\mathbf{Ax} = \mathbf{b} \quad (31)$$

gde je \mathbf{A} matrica sistema, dimenzija $m \times n$, dok je \mathbf{b} matrica kolona, dimenzije m . Obe ove matrice su nam poznate, dok je \mathbf{x} nepoznata matrica rešenja, po tipu matrica vrsta, dimenzije m . Kako bi smo lakše shvatili matrični zapis, pišemo ovu jednačinu u proširenom obliku:

$$\mathbf{Ax} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{bmatrix} \mathbf{x} = \begin{bmatrix} a_{1,1} \\ \vdots \\ a_{n,1} \end{bmatrix} \mathbf{x}_1 + \cdots + \begin{bmatrix} a_{1,m} \\ \vdots \\ a_{n,m} \end{bmatrix} \mathbf{x}_m = \sum_{j=1}^m \mathbf{a}_j \mathbf{x}_j = \mathbf{b} \quad (32)$$

Dalji postupak je sličan kao, u gore navedenom, primeru. Tražimo x , sumu minimalnog kvadratnog ostatka (odstupanja). Sledeći korak je traženje minimuma takve funkcije, zatim se dobija sistem od n linearnih jednačina sa n nepoznatih. Rešavanjem tog sistema dobijamo nepoznatu matricu \mathbf{x} , iz koje možemo pročitati nepoznate vrednosti x .

4 Nelinearni problemi i metodi za njihovo rešavanje

4.1 Uvod

Do kraja ove celine, diskutovaćemo o nelinearnim problemima i metodama za njihovo rešavanje. Pre svega, tu je metoda najmanjih kvadrata primenjena na nelinearne probleme.

Pretpostavimo da je data funkcija $f : R^n \mapsto R^m$, gde je $m \geq n$. Kao i kod linearnog slučaja, želimo minimizovati $\|f(x)\|$, ili ekvivalentno, da pronademo $x^* = \operatorname{argmin}_x F(x)$, gde je:

$$F(x) = \frac{1}{2} \sum_{i=1}^m (f_i(x))^2 = \frac{1}{2} \|f(x)\|^2 = \frac{1}{2} f(x)^T f(x) \quad (33)$$

Ovi problemi se mogu rešiti generalnim metodama optimizacije, ali ćemo mi ovde ipak predstaviti specijalne metode koje su mnogo efektivnije u pogledu vremena koje je potrebno da se kod izvrši. Treba napomenuti da se svi ovi koraci u postupku skoro nikad ne rade ručno, već upotrebom računara i nekog, za matematiku pogodnog, programskog jezika.

Za dalje izlaganje, biće nam potrebne sledeće definicije:

Definicija 1: Data je funkcija $f(x)$, pod uslovom da je ona neprekidna i da ima parcijalni izvod drugog reda u svakoj tački svog domena, njen Tejlorov razvoj možemo zapisati kao:

$$f(x+h) = f(x) + J(x)h + O(\|h\|^2) \quad (34)$$

gde je J Jakobijeva matrica i važi $J \in R^{m \times n}$, a to je zapravo matrica koja sadrži parcijalne izvode prvog reda date funkcije:

$$(J(x))_{ij} = \frac{\partial f_i}{\partial x_j}(x) \quad (35)$$

odakle se, na osnovu jednačine (33), može napisati:

$$\frac{\partial F}{\partial x_i}(x) = \sum_{i=1}^m f_i(x) \frac{\partial f_i}{\partial x_i}(x) \quad (36)$$

Napokon, dobijamo gradijent (nagib prave linije):

$$F'(x) = J(x)^T f(x) \quad (37)$$

Svi metodi rešavanja nelinearnih problema su iterativni. Od početne tačke x_0 metoda (koji koristimo), stvara se niz vektora x_1, x_2, \dots , koji, kako se nadamo konvergiraju ka x^* , lokalnom minimumu naše funkcije. Skoro svi metodi imaju način na koji obezbeđuju da je funkcija monotono opadajuća, tj da važi:

$$F(x_{k+1}) < F(x_k) \quad (38)$$

Definicija 2: Kažemo da je h pravac opadanja funkcije $F(x)$, u tački x , ako važi $h^T F'(x) < 0$.

Ako ipak ne postoji takvo h , onda je $F'(x) = 0$, što nam govori da je x stacionarna tačka date funkcije. U suprotnom, moramo izabrati α , tj. moramo odrediti koliko daleko možemo pustiti posmatranu tačku da se odalji od tačke x , a u pravcu tačke h_0 (smer opadanja funkcije), tako da dobijemo opadanje vrednosti funkcije. O ovome će biti reči u narednom odeljku.

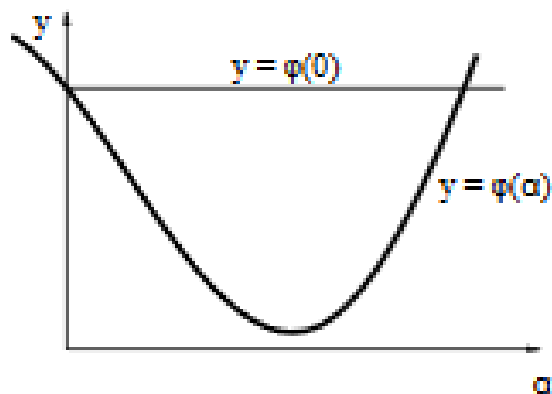
4.2 Određivanje (traženje) krive

Postavlja se pitanje, kako možemo zaključiti na osnovu čega da crtamo krivu i kako da odredimo njen gradijent.

Neka nam je data tačka x i smer opadanja h . Naredna iteracija posmatranu tačku pomera iz tačke x , u smeru opadanja h . Da bi znali koliko "daleko" treba da idemo, proučavamo varijaciju date funkcije duž polovine puta od tačke x do pravca h :

$$\varphi(\alpha) = F(x + \alpha h) \quad x, h = \text{const.} \quad \alpha \geq 0 \quad (39)$$

Primer ponašanja krive $\varphi(\alpha)$ prikazano je na slici 9.



Slika 9: Varijacija neke funkcije duž prave (putanje) traženja krive

Na osnovu grafika i jednačine (39) zaključujemo da se traži minimum u našem primeru i kada se dostigne taj minimum, onda se nekom od metoda, koje ćemo kasnije prikazati, izračunava vrednost i nalazi gradijent za konkretnu tačku minimuma. Pritom treba obratiti pažnju da je h pravac opadanja koji nam obezbeđuje da važi:

$$\varphi'(0) = h^T F'(x) < 0 \quad (40)$$

i pokazuje da dokle god imamo dovoljno malo α , koje nam garantuje da je ispunjen uslov $F(x_{k+1}) > F(x_k)$, što je u stvari ekvivalentno sa $\varphi(\alpha) < \varphi(0)$, važi uslov (40).

Često se uzima inicijalna pretpostavka vrednosti α , (najčešće $\alpha = 1$ kod Njutnove metode).

Postoje tri disjunktna slučaja koja moramo razlikovati:

1. α je jako malo, što nam govori da je vrednost funkcije jako mala i moramo povećati α ,
2. α je preveliko, odnosno $\varphi(\alpha) \geq \varphi(0)$. Moramo smanjiti α kako bi zadovoljili uslov $F(x_{k+1}) < F(x_k)$
3. α je blizu vrednosti lokalnog minimuma ($\varphi(\alpha)$). Prihvatamo tu vrednost α .

Ovo je iterativni proces koji stvara niz $\alpha_1, \alpha_2, \dots$

4.3 Njutnova metoda

Ovaj metod se izvodi iz uslova da je x^* stacionarna tačka. Pošto, po pretpostavci, znamo da je x^* stacionarna tačka, sledi da je $F'(x^*) = 0$. Odavde se dobija nelinearan sistem jednačina i na osnovu Tejlorovog razvoja se dobija:

$$F'(x+h) = F'(x) + F''(x)h + O(\|h\|^2) \quad (41)$$

Ako uzmemo da je $\|h\|$ dovoljno malo, možemo izraz napisati kao:

$$F'(x+h) \simeq F'(x) + F''(x)h \quad (42)$$

Odavde izvodimo Njutnovu metodu. Prvo treba pronaći h_0 , kao rešenje sledeće jednačine:

$$Hh_n = -F'(x) \quad (43)$$

pri čemu je $H = F''(x)$, dok narednu iteraciju računamo kao:

$$x := x + h_0 \quad (44)$$

Uzmimo da je H sigurno pozitivno, odatle sledi da jednačina (43) ima nesingularnih rešenja (implicira da ima jedinstvena rešenja). Pritom, neka je $u^T H u > 0$, za sve nenulte vrednosti u .

Sada, ako pomnožimo obe strane sa h_n^T , dobijamo:

$$0 < h_n^T H h_n = -h_n^T F'(x) \quad (45)$$

odakle zaključujemo da h_n opada, jer zadovoljava definiciju 2.

Njutnova metoda se najbolje pokazuje u poslednjim koracima iteracije, gde je x jako blizu x^* . Ako se zaista pokaže da je Hesse-ova matrica veća od nule, tj. da je njena sopstvena vrednost veća od nule i ako je u okolini x^* , $F''(x)$ pozitivno, tada se dobija kvadratna konvergencija.

Sa druge strane, ako se x nalazi u okolini tačke gde je $F''(x)$ negativno, i ako postoji stacionarna tačka onda će funkcija ipak konvergirati prema stacionarnoj tački koja je u ovom slučaju maksimum funkcije. Kao što znamo kod traženja sume ostatka metodom najmanjih kvadrata, tražimo minimum funkcije.

To je jedna od mana Njutnove metode, koja se prevazilazi uvođenjem Choelsky-eve metode o kojoj ovde neće biti reči. Razlog za ovo je taj što se kod Njutnove metode koristi $F''(x)$, a za komplikovanije probleme je ovaj izvod ili pozitivan ili negativan ili čak može i da ne postoji. Ova mana Njutnove metode je razlog zašto se ona i ne koristi tako često u svetu programiranja optimizacionih metoda. Postoje dodatne metode poput Quasi - Njutn ili Choelsky metode koje donekle prevazilaze ovu manu, ali im je vreme izvršavanja značajno duže u odnosu na Njutnovu metodu.

4.4 Gaus - Njutnova metoda

Ova metoda je temelj nekolicini vrlo dobrih metoda, kojima ćemo se baviti u ovom radu. Zasnovana je na prvom izvodu komponenata vektorske funkcije. U specijalnim slučajevima može dati kvadratnu konvergenciju, kao i Njutnova metoda što daje za generalne metode optimizacije.

Gaus - Njutnova metoda zasnovana je na linearnoj aproksimaciji komponenata funkcije f u okolini tačke x (linearno modelovanje funkcije f). Za dovoljno malo $\|h\|$, možemo primeniti Tejlorov razvoj na sledeći način:

$$f(x+h) \simeq l(h) \equiv f(x) + J(x)h \quad (46)$$

Kada zamenimo po F , iz jednačine (33), vidimo:

$$F(x+h) \simeq L(h) \equiv \frac{1}{2}l(h)^T l(h) = \frac{1}{2}f^T f + h^T J^T f + \frac{1}{2}h^T J^T J h = F(x) + h^T J^T f + \frac{1}{2}h^T J^T J h \quad (47)$$

gde je $f = f(x)$ i $J = J(x)$.

Gaus - Njutnov korak h_{gn} minimizuje $L(h)$, odnosno traži minimum te funkcije na lokalnom domenu, što možemo zapisati kao:

$$h_{gn} = \operatorname{argmin}_h L(h) \quad (48)$$

Ako pažljivije pogledamo, možemo zaključiti da je gradijent i Hesse-ova matrica od L :

$$L'(h) = j^T f + J^T J h, \quad L''(h) = J^T J \quad (49)$$

Poređenjem jednačina (37) i (49), možemo primetiti da $L'(0) = F'(x)$. Dalje, možemo uočiti da je matrica $L''(H)$ nezavisna od vrednosti h . Takođe, simetrična je i ako je J punog ranga, invertibilna, onda je i $L''(h)$ sigurno pozitivno. Ovo nam je vrlo bitno jer implicira da $L(h)$ ima jedinstveni minimum koji se može naći rešavanjem sledeće jednačine:

$$(J^T J)h_{gn} = -J^T f \quad (50)$$

Funkcija je monotono opadajuća jer:

$$h_{gn}^T F'(x) = h_{gn}^T (J^T f) = -h_{gn}^T (J^T J)h_{gn} < 0 \quad (51)$$

Najčešće se koristi formula $x := x + \alpha h_{gn}$, za naredni korak, kod postupka iteracije, gde se α nalazi putem metode traženja krive. Kod Gaus - Njutnove metode se najčešće uzima da je $\alpha = 1$ u svim koracima iteracije. Koristeći se metodom traženja krive, može se pokazati da imamo sigurnu konvergenciju funkcije, pod uslovom da važi:

- $x|F(x) \leq F(0)$ je ograničena funkcija, i to u svim iteracijama
- Jakobijan matrica $J(x)$ je punog ranga, i to u svim iteracijama

4.5 Levenberg - Marquardt-ov metod

Levenberg (1894), i kasnije Marquardt (1963), sugerišu sličan metod zasnovan na Gaus - Njutnovom metodu.

Korak iteracije h_{Im} dobija sledeće izmene u odnosu na jednačinu (50):

$$(J^T J + \mu I)h_{Im} = -g, \quad g = J^T f, \quad \mu \geq 0 \quad (52)$$

gde je $J = J(x)$ i $f = F(X)$, a μ je takozvani "damping" parametar koji nam je od velike važnosti u ispitivanju konvergencije funkcije u svakom koraku iteracije.

Postoje tri bitne karakteristike μ :

1. Za svako $\mu > 0$, koeficijent matrice je pozitivno definisan, i to obezbeđuje da korak iteracije h_{Im} monotono opada i približava se vrednosti lokalnog minimuma.
2. Za veće vrednosti μ dobijamo:

$$h_{Im} \simeq -\frac{1}{\mu}g = -\frac{1}{\mu}F'(x) \quad (53)$$

tj. mali pomak u pravcu lokalnog minimuma, kada je nagib krive veliki. Ovo je dobra osobina ako je iteracija daleko od lokalnog minimuma.

3. Ako je μ vrlo malo, onda $h_{Im} \simeq h_{gn}$, gde je h_{gn} korak iteracije Gaus - Njutnove metode, što je ustvari dobar korak u finalnim iteracijama.

Iz ovoga se može primetiti da parametar μ utiče na pravac i vrednost pomaka koraka iteracije, što nas dovodi do krucijalnog zaključka da nam za ovu metodu nije potrebno traženje krive iz poglavlja (4.3).

Treba pomenuti da je μ proizvoljan faktor, koji mi sami biramo. Teorijski argumenti postoje, koji pokazuju zašto neki izbori garantuju lokalnu konvergenciju algoritma. Bilo kakav izbor μ u novije vreme neće pretežno uticati na vreme izvršavanja programa, zbog naglog poboljšanja i ubrzanja računskih mašinama na kojima se programi puštaju u rad.

Izbor inicijalne vrednosti μ trebao bi da bude na osnovu veličine elemenata iz matrice:

$$\mathbf{A}_0 = \mathbf{J}(\mathbf{x}_0)^T \mathbf{J}(\mathbf{x}_0) \quad (54)$$

i to formulom:

$$\mu_0 = \tau \cdot \max_i \{ a_{ii}^{(0)} \} \quad (55)$$

gde mi biramo τ .

Tokom samih iteracija, veličinu μ možemo promeniti, koristeći se:

$$\varrho = \frac{F(x) - F(x + h_{Im})}{L(0) - L(h_{Im})} \quad (56)$$

gde se imenilac dobija po formuli iz Gaus - Njutnove metode:

$$L(0) - L(h_{Im}) = -h_{Im}^T J^T f - \frac{1}{2} h_{Im}^T J^T J h_{Im} =$$

$$-\frac{1}{2} h_{Im}^T (2g + (J^T J + \mu I - \mu I) h_{Im}) = \frac{1}{2} h_{Im}^T (\mu h_{Im} - g)$$

Obratimo pažnju na to da su $h_{Im}^T h_{Im}$ i $-h_{Im}^T g$ oba pozitivna, dakle sledi da je $L(0) - L(h_{Im})$ garantovano pozitivno, što nam obezbeđuje da je uslov 1. "damping" faktora zadovoljen.

Velika vrednost ρ nam daje do znanja da je $L(h_{Im})$ u dobroj aproksimaciji sa $F(x + h_{Im})$ i možemo smanjiti μ tako da nam u narednoj iteraciji h_{Im} , korak iteracije bude bliži Gaus - Njutnovom koraku.

Ako je ρ eventualno negativno, onda je $L(h_{Im})$ loša aproksimacija, i moramo povećati μ da bude duplo veće nego što je bilo, da bi se približili lokalnom minimumu.

Kriterijum zaustavljanja algoritma je momenat kada vrednost globalnog minimuma $F'(X^*) = g(x^*) = 0$. Kako bi proverili da li je ovaj kriterijum zadovoljen, koristimo sledeću nejednakost:

$$\|g\|_{\infty} \leq \varepsilon_1 \quad (57)$$

gde je ε vrlo mali, pozitivan broj, koji smo sami izabrali (ε okolina).

Još jedan, relevantan, kriterijum jeste da zaustavimo izvršavanje programa, onda kada promena promenljive x postane suviše mala da se ne može razlikovati od numeričkog "šuma" samog računara. Taj kriterijum matematički zapisujemo sledećom nejednakosti:

$$\|x_{novo} - x_{trenutno}\| \leq \varepsilon_2 (\|x\| + \varepsilon_2) \quad (58)$$

Ovaj izraz daje postepenu promenu u odnosu na korak iteracije ε_2 , kada je $\|x\|$ veliko. Pošto je ovo takođe iterativni proces, potrebna nam je zaštita od beskrajne petlje. To postizemo uslovom:

$$k \geq k_{max} \quad (59)$$

U nastavku dajemo Levenberg – Marquardt-ov algoritam:

```

begin
   $k := 0; \quad \nu := 2; \quad \mathbf{x} := \mathbf{x}_0$ 
   $\mathbf{A} := \mathbf{J}(\mathbf{x})^\top \mathbf{J}(\mathbf{x}); \quad \mathbf{g} := \mathbf{J}(\mathbf{x})^\top \mathbf{f}(\mathbf{x})$ 
   $found := (\|\mathbf{g}\|_\infty \leq \varepsilon_1); \quad \mu := \tau * \max\{a_{ii}\}$ 
  while (not  $found$ ) and ( $k < k_{\max}$ )
     $k := k+1; \quad \text{Solve } (\mathbf{A} + \mu \mathbf{I})\mathbf{h}_{lm} = -\mathbf{g}$ 
    if  $\|\mathbf{h}_{lm}\| \leq \varepsilon_2(\|\mathbf{x}\| + \varepsilon_2)$ 
       $found := \mathbf{true}$ 
    else
       $\mathbf{x}_{\text{new}} := \mathbf{x} + \mathbf{h}_{lm}$ 
       $\rho := (F(\mathbf{x}) - F(\mathbf{x}_{\text{new}}))/(L(\mathbf{0}) - L(\mathbf{h}_{lm}))$ 
      if  $\rho > 0$ 
         $\mathbf{x} := \mathbf{x}_{\text{new}}$ 
         $\mathbf{A} := \mathbf{J}(\mathbf{x})^\top \mathbf{J}(\mathbf{x}); \quad \mathbf{g} := \mathbf{J}(\mathbf{x})^\top \mathbf{f}(\mathbf{x})$ 
         $found := (\|\mathbf{g}\|_\infty \leq \varepsilon_1)$ 
         $\mu := \mu * \max\{\frac{1}{3}, 1 - (2\rho - 1)^3\}; \quad \nu := 2$ 
      else
         $\mu := \mu * \nu; \quad \nu := 2 * \nu$ 
    end

```

5 Primer rešavanja problema u telekomunikacijama metodom najmanjih kvadrata

5.1 Polinomska aproksimacija

Jedna od važnijih primena jeste polinomska aproksimacija. Pretpostavimo da su nam dati podaci u formi (t_i, y_i) , pri čemu $1 \leq i \leq N$. Cilj nam je da aproksimiramo podatke i nacrtamo krivu. Na osnovu jednačina (3) i (4), možemo zapisati:

$$f(p(t)) = \sum_{i=1}^N (y_i - p(t_i))^2 \quad (60)$$

kao i:

$$f(a_0, a_1, a_2) = \sum_{i=1}^N [y_n - (a_0 + a_1 t_n + a_2 t_n^2)]^2 = \min \quad (61)$$

gde je $p(t) = a_0 + a_1 t + a_2 t^2$, i gde se umesto t , u našem primeru, može uzeti i x .

Radi lakšeg rešavanja, problem ćemo predstaviti u matičnom obliku:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \approx \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_N & t_N^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \quad (62)$$

Na osnovu jednačine (23), ovaj sistem je ekvivalentan sistemu:

$$\mathbf{H}\vec{a} = \vec{y} \Leftrightarrow \begin{bmatrix} H_{0,0} & H_{0,1} & H_{0,2} \\ H_{1,0} & H_{1,1} & H_{1,2} \\ H_{2,0} & H_{2,1} & H_{2,2} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix} \quad (63)$$

Slično kao i kod rešavanja iz poglavlja 2, za linearne funkcije, tj. polinomsku funkciju u našem slučaju, tako rešavamo i ovu kvadratnu funkciju. Pridržavajući se istih koraka:

$$\det H = H_{0,0}(H_{1,1}H_{2,2} - H_{1,2}H_{2,1}) - H_{1,0}(H_{0,1}H_{2,2} - H_{0,2}H_{2,1}) + H_{2,0}(H_{0,1}H_{1,2} - H_{0,2}H_{1,1}) \quad (64)$$

$$\det H = H_{0,0}H_{1,1}H_{2,2} - H_{0,0}H_{1,2}H_{2,1} - H_{1,0}H_{0,1}H_{2,2} + H_{1,0}H_{0,2}H_{2,1} + H_{2,0}H_{0,1}H_{1,2} - H_{2,0}H_{0,2}H_{1,1}$$

Determinantu trećeg reda dobićemo determinantom drugog reda, pa je konačan izraz za datu determinantu:

$$\det H = H_{0,0}H_{1,1}H_{2,2} + 2H_{2,0}H_{2,1}H_{1,0} - H_{0,0}H_{1,0}^2 - H_{1,0}^2H_{2,2} - H_{2,0} \quad (65)$$

U narednom koraku određujemo adjungovanu matricu, a zatim svaki pojedinačni element u njoj, zamenimo determinantom drugog reda. Klasičan postupak dobijanja adjungovane matrice.

$$H^* = \begin{bmatrix} H_{1,1}H_{2,2} - H_{2,1}H_{1,2} & -(H_{0,1}H_{2,2} - H_{2,1}H_{0,2}) & H_{0,1}H_{1,2} - H_{1,1}H_{0,2} \\ -(H_{1,0}H_{2,2} - H_{2,0}H_{1,2}) & H_{0,0}H_{2,2} - H_{2,0}H_{0,2} & -(H_{0,0}H_{1,2} - H_{1,0}H_{0,2}) \\ H_{1,0}H_{2,1} - H_{2,0}H_{1,1} & -(H_{0,0}H_{2,1} - H_{2,0}H_{0,1}) & H_{0,0}H_{1,1} - H_{1,0}H_{0,1} \end{bmatrix} \quad (66)$$

Prema tome:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \frac{1}{\det H} \begin{bmatrix} H_{1,1}H_{2,2} - H_{2,1}H_{1,2} & -(H_{0,1}H_{2,2} - H_{2,1}H_{0,2}) & H_{0,1}H_{1,2} - H_{1,1}H_{0,2} \\ -(H_{1,0}H_{2,2} - H_{2,0}H_{1,2}) & H_{0,0}H_{2,2} - H_{2,0}H_{0,2} & -(H_{0,0}H_{1,2} - H_{1,0}H_{0,2}) \\ H_{1,0}H_{2,1} - H_{2,0}H_{1,1} & -(H_{0,0}H_{2,1} - H_{2,0}H_{0,1}) & H_{0,0}H_{1,1} - H_{1,0}H_{0,1} \end{bmatrix} \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}$$

dok nepoznate parametre a_0 , a_1 i a_2 dobijamo prema formulama:

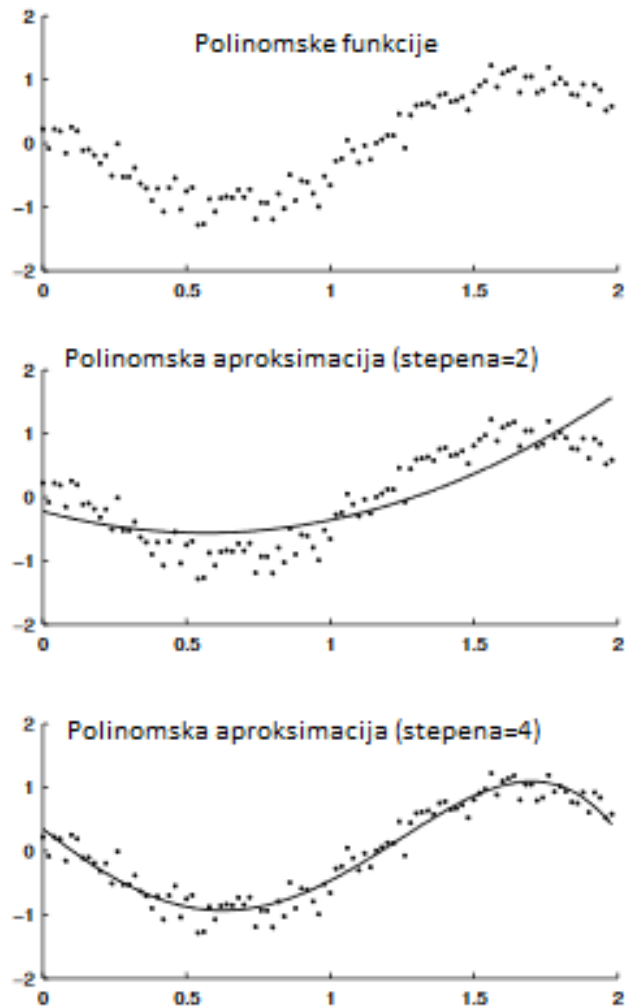
$$a_0 = \frac{(H_{1,1}H_{2,2} - H_{2,1}^2)y_0 + (H_{2,1}H_{0,2} - H_{0,1}H_{2,2})y_1 + (H_{0,1}H_{1,2} - H_{1,1}H_{0,2})y_2}{H_{0,0}H_{1,1}H_{2,2} + 2H_{2,0}H_{2,1}H_{1,0} - H_{0,0}H_{1,2}^2 - H_{1,0}^2H_{2,2} - H_{2,0}^3} \quad (67)$$

$$a_1 = \frac{(H_{2,0}H_{1,2} - H_{1,0}H_{2,2})y_0 + (H_{0,0}H_{2,2} - H_{2,0}^2)y_1 + (H_{1,0}H_{0,2} - H_{0,0}H_{1,2})y_2}{H_{0,0}H_{1,1}H_{2,2} + 2H_{2,0}H_{2,1}H_{1,0} - H_{0,0}H_{1,2}^2 - H_{1,0}^2H_{2,2} - H_{2,0}^3} \quad (68)$$

$$a_2 = \frac{(H_{1,0}H_{2,1} - H_{2,0}H_{1,1})y_0 + (H_{2,0}H_{0,1} - H_{0,0}H_{2,1})y_1 + (H_{0,0}H_{1,1} - H_{1,0}^2)y_2}{H_{0,0}H_{1,1}H_{2,2} + 2H_{2,0}H_{2,1}H_{1,0} - H_{0,0}H_{1,2}^2 - H_{1,0}^2H_{2,2} - H_{2,0}^3} \quad (69)$$

Ovaj postupak važi i za polinome višeg stepena. Sistem bi sveli na matricu odgovarajućeg reda, računajući njenu inverznu matricu pomoću determinante i odgovarajuće adjungovane matrice. Primetimo da se ovom metodom eksponencijalno povećava vreme potrebno za izračunavanje polinoma viših stepena. To u današnjem, modernom, dobu ne predstavlja veliki problem savremenim procesorima i računskim mašinama, za razliku prošlih vremena.

U nastavku dajemo grafike dve polinomske funkcije i njihove aproksimacije:

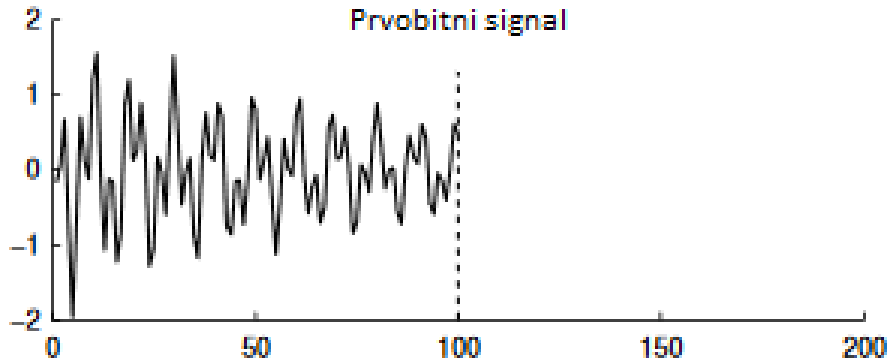


Slika 10: Prikaz aproksimacija polinomskih funkcija različitih stepena

Vidimo da se kod viših stepena polinomske aproksimacije dobija kriva koja bolje odgovara podacima i pokazuje bolju aproksimaciju.

5.2 Linearna predikcija signala

Neka nam je dat signal prikazan na slici 11. Želimo da odredimo buduće vrednosti tog vremenski zavisnog signala, na osnovu linearne predikcije samog tog signala.



Slika 11: Izgled prvobitnog signala

Ako su nam prethodni podaci $y(n)$ o signalu poznati, onda problem pronalaženja a_i se može rešiti metodom najmanjih kvadrata. U slučaju da nisu nije moguće. Uzmimo da su podaci o signalu, $y(n)$, dati sledećom jednačinom:

$$y(n) \approx a_1 y(n-1) + a_2 y(n-2) + a_3 y(n-3) \quad (70)$$

Problem pronalaženja $a = (a_0, a_1, a_2)^T$ može se posmatrati kao rešavanje određenog sistema jednačina. Na primer, ako je $y(n)$ poznato za $0 \leq n \leq N-1$, i mi tražimo linearnu predikciju trećeg stepena, onda je sistem dat kao:

$$\begin{bmatrix} y(3) \\ y(4) \\ \vdots \\ y(N-1) \end{bmatrix} \approx \begin{bmatrix} y(2) & y(1) & y(0) \\ y(3) & y(2) & y(1) \\ \vdots & \vdots & \vdots \\ y(N-2) & y(N-3) & y(N-4) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad (71)$$

Kao i u prethodnim primerima, ovaj sistem možemo zapisati u obliku $\mathbf{y} = \mathbf{H}\mathbf{a}$, gde je H Hesse-ova matrica, dimenzija $(N-3) \times 3$. Na osnovu formule bobijene rešavanjem normalnog sistema, koji glasi:

$$H^T H a = H^T y \quad (72)$$

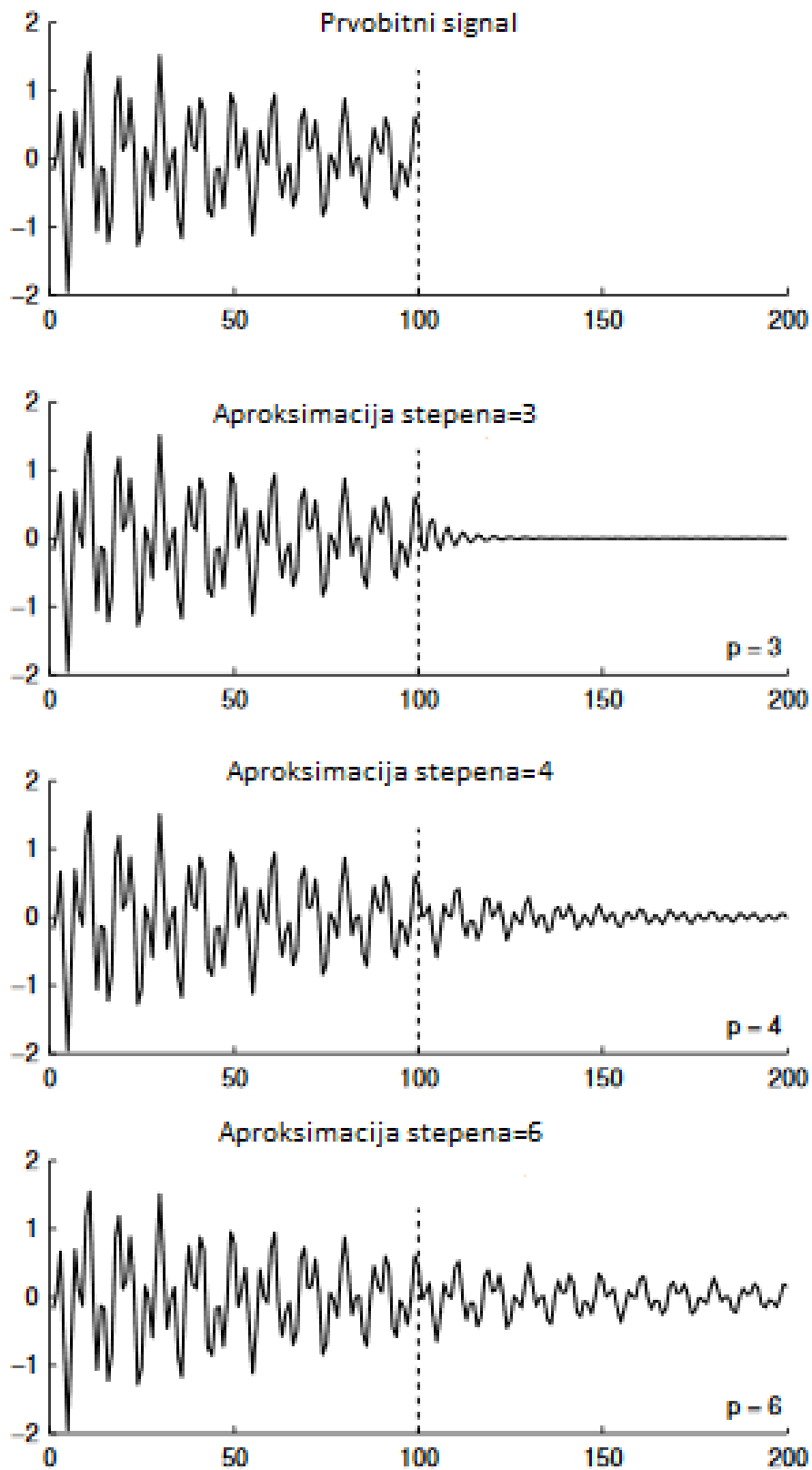
sledi:

$$\mathbf{a} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} \quad (73)$$

Primititi da je $H^T H$ "malo", dimenzija 3×3 samo.

Odavde se dobija \mathbf{a} rešavanjem sistema linearnih jednačina. Kada se pronade koeficijent a_i , onda možemo aproksimirati $y(n) > N$ koristeći rekurzivnu jednačinu datu u primeru (5.1).

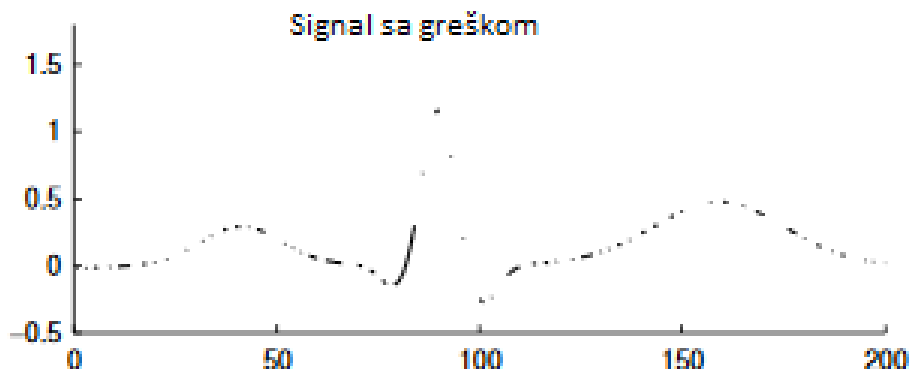
Na slici 12 vidimo originalni signal i nekoliko njegovih prediktovanih aproksimacija.



Slika 12: Predikcija i aproksimacija signala

5.3 Popravljanje signala oštećenog greškom

Usled različitih vrsta grešaka (npr. usled interferencije, šuma ili slabljenja), može se dogoditi da nedostaje deo signala. Grafik primera takvog signala vidimo na slici 13.



Slika 13: Grafik signala oštećenog u prenosu

Kako bi aproksimirali sam signal, moramo posmatrati signal na osnovu onoga što imamo, tj. na osnovu delova koji su nam poznati. Na ovom primeru ćemo ilustrovati kako je moguće aproksimirati sam signal na osnovu metode najmanjih kvadrata. Na slici 13 možemo videti da je signal sačinjen od 200 taktova, vidi se da su neki delovi netaknuti, dok na nekim mestima postoji očigledan prekid u signalu.

Da bi smo formulisali problem pomoću metode najmanjih kvadrata, uvešćemo neke nove promenljive. Neka je x signal dužine N . Pretpostavimo da je poznato K delova tog signala x , gde je $K < N$.

Za K tačaka poznatog signala, y se može napisati kao:

$$\mathbf{y} = \mathbf{S}\mathbf{x} \quad (74)$$

gde je \mathbf{S} selekciona matrica ("matrica delova"), veličine $K \times N$.

Na primer, ako se posmatra samo prvi, drugi ili poslednji element signala x , od pet delova (taktova), tada je matrica \mathbf{S} data kao:

$$\mathbf{s} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (75)$$

Na osnovu formule (74), možemo napisati jednačinu u proširenom obliku:

$$\mathbf{S}\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ x(3) \\ x(4) \end{bmatrix} = \begin{bmatrix} x(0) \\ x(1) \\ x(4) \end{bmatrix} = \mathbf{y} \quad (76)$$

Obratite pažnju da množenjem \mathbf{S} i \mathbf{x} mi otklanjamo dva dela signala od originalnog signala \mathbf{x} , a to su $x(2)$ i $x(3)$. Konkretno u našem primeru, rečeno nam je da su poznati prvi, drugi i poslednji takt signala.

Vektor \mathbf{y} je sastavljen samo od poznatih vrednosti x , odakle sledi da je vektor \mathbf{y} kraći od vektora \mathbf{x} , ($K < N$).

Problem se dalje može formulirati na sledeći način: neka je dat signal \mathbf{y} i matrica \mathbf{S} , naći \mathbf{x} za koje važi $\mathbf{y} = \mathbf{S}\mathbf{x}$. Naravno, odmah se vidi da postoji beskonačno mnogo rešenja ovog sistema, ali mi ćemo ovde prikazati jedinstven način kako da se dobiju samo delovi signala koji nedostaju.

Primetimo da $\mathbf{S}^T\mathbf{y}$ ima efekat da nam postavi vrednosti 0 u matricu za delove signala koji nedostaju. U konkretnom slučaju imamo:

$$\mathbf{S}^T\mathbf{y} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y(0) \\ y(1) \\ y(2) \end{bmatrix} = \begin{bmatrix} y(0) \\ y(1) \\ 0 \\ 0 \\ y(2) \end{bmatrix} \quad (77)$$

U cilju nalaženja delova koji nedostaju potrebna nam je nova matrica, a to je matrica \mathbf{S}_c , odnosno "dopunska" matrica matrice \mathbf{S} . Matrica \mathbf{S}_c je sačinjena isključivo od vrsta koje se ne pojavljuju u originalnoj matrici \mathbf{S} :

$$\mathbf{S}_c = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (78)$$

Sada se aproksimacija signala $\hat{\mathbf{x}}$ može predstaviti kao

$$\hat{\mathbf{x}} = \mathbf{S}^T\mathbf{y} + \mathbf{S}_c^T\mathbf{v} \quad (79)$$

gde \mathbf{y} predstavlja dostupne podatke (taktove) signala, a \mathbf{v} predstavlja delove signala koji nedostaju, tj. koje treba da odredimo i nadoknadimo.

U konkretnom slučaju:

$$\mathbf{S}^T\mathbf{y} + \mathbf{S}_c^T\mathbf{v} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y(0) \\ y(1) \\ y(2) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} v(0) \\ v(1) \end{bmatrix} = \begin{bmatrix} y(0) \\ y(1) \\ v(0) \\ v(1) \\ y(2) \end{bmatrix} \quad (80)$$

Problem je aproksimacija vektora \mathbf{v} , koji je dužine $N - K$.

Vektor \mathbf{v} možemo naći kada tražimo minimum $\|\mathbf{D}\hat{\mathbf{x}}\|_2^2$, gde je \mathbf{D} diferencna matrica drugog reda.

Diferencna matrica drugog reda je diferencna jednačina u kojoj vrednosti vektora promenljivih u nekom momentu vremena su u odnosu sa jednom ili više vrednosti u prethodnom momentu vremena. Red jednačina je maksimalan vremenski razmak između bilo koje dve vrednosti vektora, npr:

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{B}\mathbf{x}_{t-2} \quad (81)$$

Sada možemo naći vrednost \mathbf{v} za naš konkretan primer, a to je:

$$\min_{\mathbf{v}} \| \mathbf{D}(\mathbf{S}^T \mathbf{y} + \mathbf{S}_c^T \mathbf{v}) \|_2^2 \quad (82)$$

odnosno

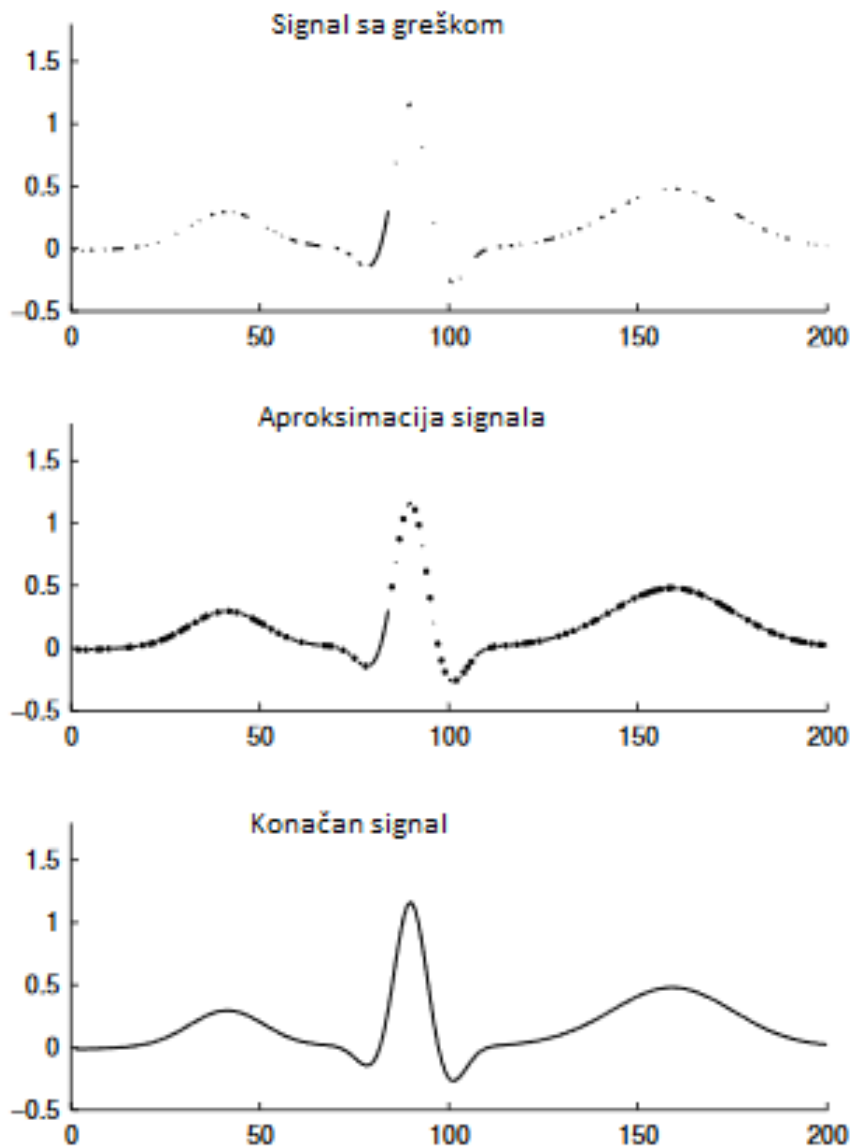
$$\min_{\mathbf{v}} \| \mathbf{D} \mathbf{S}^T \mathbf{y} + \mathbf{D} \mathbf{S}_c^T \mathbf{v} \|_2^2 \quad (83)$$

pa na osnovu jednačine (72), rešenje je dato kao:

$$\mathbf{v} = -(\mathbf{S}_c \mathbf{D}^T \mathbf{D} \mathbf{S}_c^T)^{-1} \mathbf{S}_c \mathbf{D}^T \mathbf{D} \mathbf{S}^T \mathbf{y} \quad (84)$$

Jednom kada se pronađe \mathbf{v} , aproksimacija $\hat{\mathbf{x}}$ se može jednostavno grafički prikazati dodavanjem $\mathbf{v}(\mathbf{i})$, tj. dodavanjem vrednosti za taj takt, u signal \mathbf{y} .

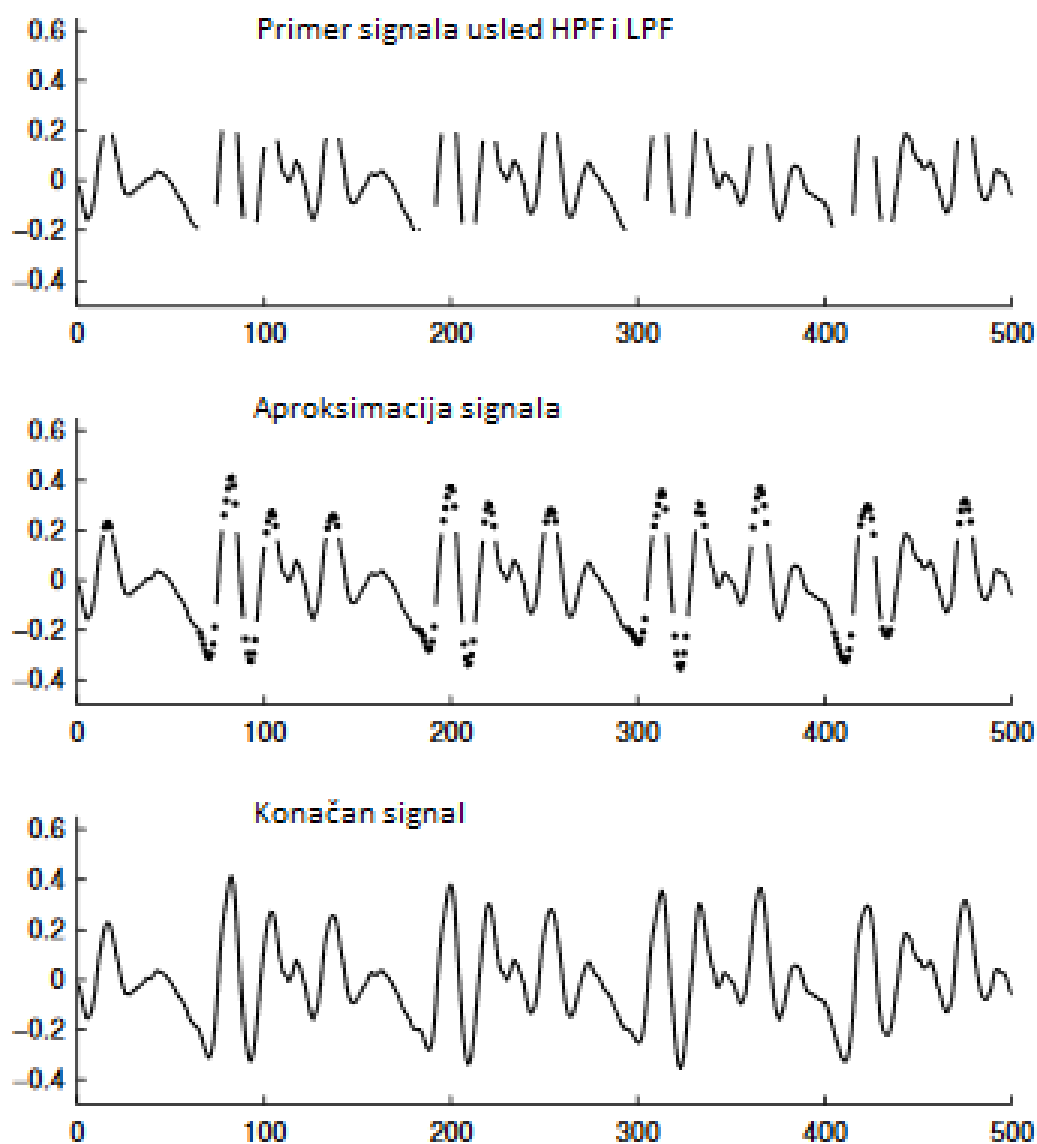
Koristeći se metodom objašnjenom ovde, može se aproksimirati skoro svaki nedostatak signala, bilo koje veličine. U nastavku dajemo sliku 14, na kojoj se vidi popravljen signal.



Slika 14: Grafik signala ispravljenog pomoću metode najmanjih kvadrata

5.4 Stvaranje nedostataka u signalu usled prolaženja kroz HPF (High pass filter) i LPF (Low pass filter)

Česta pojava u telekomunikacijama i obradi signala su HPF i LPF filtri. Ako je amplituda signala prevelika ili ipak previše mala, zapis signala će biti uprošćen. Ovo je česta pojava jer nema opreme koja bi posle određene frekvencije zapisivala signal, pa se on "odseca" da ne bi pravio dodatne smetnje u instrumentima. Na slici 15 vidimo šta se zapravo dogodi sa signalom, kada prođe kroz filter.



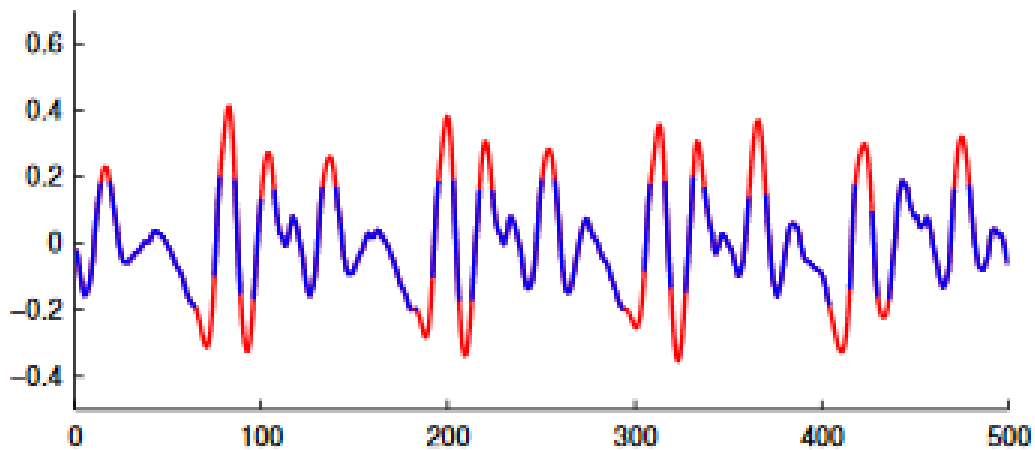
Slika 15: Grafik signala pre i posle prolaženja kroz filter

Da bi dobili delove signala koji nedostaju možemo koristiti metodu iz prethodnog poglavlja, kako bi aproksimirali signal. Kako bi iskoristili jednačinu (84) i kako bi nam ona bila od koristi u ovom primeru, moramo tražiti minimum trećeg izvoda funkcije signala. Na ovaj način podaci koje ćemo dobiti, poprimiće oblik parabole (kada bi se grafički predstavili), jer izvod trećeg reda parabole je 0.

Sve što je potrebno da bi primenili jednačinu (84), jeste da modifikujemo matricu \mathbf{D} , na sledeći način:

$$\mathbf{D} = \begin{bmatrix} 1 & -3 & 3 & -1 & & & & \\ & 1 & -3 & 3 & -1 & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & 1 & -3 & 3 & -1 & \end{bmatrix} \quad (85)$$

Odatle je \mathbf{D}_x aproksimacija izvoda trećeg reda signala \mathbf{x} . I na kraju, dobijamo signal koji vidimo na slici 16.



Slika 16: Potpun signal, nakon neutralizacije efekata HPF-a i LPF-a

Literatura

- [1] <http://graphics.stanford.edu/~jplewis/lscourse/ls.pdf>
- [2] http://eeweb.poly.edu/iselesni/lecture_notes/least_squares/least_squares_P.pdf
- [3] <https://apps.dtic.mil/dtic/tr/fulltext/u2/a201848.pdf>
- [4] <http://math.gmu.edu/~igriva/book/Appendix%20D.pdf>
- [5] <http://homepage.ntu.edu.tw/~ckuan/pdf/et01/ch8.pdf>
- [6] <https://mathworld.wolfram.com/NonlinearLeastSquaresFitting.html>
- [7] <http://www.seas.ucla.edu/~vandenbe/133A/lectures/nlls.pdf>
- [8] https://stanford.edu/class/ee103/lectures/nlls_slides.pdf
- [9] https://www.efunda.com/math/least_squares/least_squares.cfm
- [10] [https://nptel.ac.in/content/storage2/courses/122104019/numerical – analysis/Rathish-kumar/least-square/r1.htm](https://nptel.ac.in/content/storage2/courses/122104019/numerical_analysis/Rathish-kumar/least-square/r1.htm)
- [11] <https://sam.nitk.ac.in/courses/MA608/Curve%20Fitting.pdf>
- [12] https://www.ms.uky.edu/~ma137/Fasttrack19/curve_fitting.pdf
- [13] [https://nptel.ac.in/content/storage2/courses/122104019/numerical – analysis/Rathish-kumar/least-square/r1.htm](https://nptel.ac.in/content/storage2/courses/122104019/numerical_analysis/Rathish-kumar/least-square/r1.htm)
- [14] <https://mathworld.wolfram.com/LeastSquaresFitting.html>