Uvod

Oduvek je ljudska težnja bila postizanje boljih rezultata nekih procesa. Podešavanjem parametara koji utiču na proces može se postići veća ili manja vrednost funkcije cilja. Postoje brojne matematičke metode za nalaženje vrednosti parametara za koje funkcija cilja postiže optimalnu vrednost (najmanju ili najveću).

Nelinearni matematički modeli se javljaju u gotovo svim naukama, počevši od društvenih nauka i ekonomije, preko inžinjerskih nauka do medicine, u savremnim tehnologijama i problemima iz realnog života. U velikoj većini rešenja tih modela u zatvorenoj formi ne postoje i numerički pristup je jedini moguć.

Najčešći numerički pristup su iterativni postupci. U iterativnim postupcima se kreće od početne aproksimacije, korak po korak, preko vrednosti koje su sve bolje, ka traženoj optimalnoj vrednosti. Iterativne metode mogu imati mali broj iteracija koji brzo dovode do tačnog rešenja a koje su računski skupe jer se u svakoj iteraciji za nalaženje gradijenta računa Hesijan (Njutnova metoda). U cilju smanjenja broja računskih operacija, ili u nedostatku podataka drugog reda, pribegava se modifikacijama Njutnove metode. U traženju globalnog optimuma često je korišćena BFGS metoda koja malim brojem računskih operacija aproksimira Hesijan.

Šum je u određenoj meri uvek prisutan u modelima. Zbog toga se istraživanja usmeravaju na primenu metoda optimizacije u okruženju sa šumom. Najčešće je šum slučajna veličina. Oblast stohastičke optimizacije se bavi rešavanjem problema optimizacije sa slučajnim šumom. Ova oblast privlači pažnju istraživača poslednjih godina zbog savremenih računarskih mogućnosti i dobro razvijene teorije numeričke optimizacije za determinističke probleme. Moderna računarska tehnika je omogućila eksperimentalno generisanje veličina sa šumom pomoću simulacija.

Osnovno pitanje kod metoda stohastičke optimizacije je računska složenost. Broj simulacija (ili veličina uzorka) je veoma velik kod postojećih metoda. Ova činjenica ograničava primenu metoda samo na probleme skromnih dimenzija. Nas zanimaju metode koje će biti značajno jeftinije ali će i dalje davati prihvatljive aproksimacije. Pristup koji ćemo proučavati se zasniva na kombinaciji ideja iz determinističke i stohastičke optimizacije sa ciljem smanjenja broja izračunavanja i očuvanja konvergencije.

Pokazalo se da upotreba Kvazi Njutnovih metoda ima opravdanje u smanjenju složenosti algoritma kao i širini klase problema na koje se može primeniti. Tako da su se Kvazi Njutnovi postupci našli u primeni u velikom broju nauka, u mnogim oblastima u kojima se javlja potreba optimizacije. Stohastička optimizacija, odnosno stohastičko programiranje, svakako, ima širu oblast primenjivosti od determinističke optimizacije. Nalaženje postupaka koji su ekvivalenti Kvazi Njutnovih postupaka u stohastičkoj optimizaciji je jedna od oblasti numeričke matematike na kojima se intenzivno radi u skoroj prošlosti. Kvazi Njutnove metode u determinističkoj matematičkoj optimizaciji igraju veliku ulogu, postoji velika potreba za njihovim analogonima u stohastičkom okruženju. Mnogobrojni sistemi funkcionišu zahvaljujući optimalno podešenom upravljanju. Optimizacija procesa proizvodnje dovodi do velikih ušteda. Računarska simulacija u procesu projektovanja dovodi do značajnog skraćivanja vremena proizvodnje i novčanih ušteda u izradi prototipova.

Tekst ove disertacije je podeljen na tri dela.

I deo sadrži pregled teorijskih rezultata o dve oblasti koje se ovde ukrštaju:

1. Verovatnoća, slučajne promenljive i njihova primena

2. Rešavanje problema (determinističke) nelinearne optimizacije

II deo sadrži u glavi 3 klasične rezultate stohastičke optimizacije, a u glavama 4 i 5 dva originalna rezultata koji su objavljeni u eminentnim časopisima. Takođe, u glavi 5 imamo rezultate o modelima ekvilibriuma iz Ekonomije. Problem ekvilibriuma je rešavan originalnim algoritmom.

III deo sadrži tabele i grafikone numeričkih rezultata primene dva pomenuta originalna algoritma: u glavi 6 DSLS algoritam i u glavi 7 FNM algoritam. U glavi 6 je prvo je dat spisak problema za testiranje stohastičke optimizacije. U glavi 7 je opisana procedura primene FNM algoritma radi dobijanja statistike rešenja u zavisnosti od slučajno generisanih vrednosti parametara.

Standardne oznake

U tekstu će se koristiti standardne matematičke oznake. Podsetićemo ovde samo na neke:

- $2^{\Omega} = \{A : A \subseteq \Omega\}$, Partitivni skup
- $1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$
- $X^+ = \max\{X, 0\}$, $X^- = \max\{-X, 0\}$, pozitivni i negativni deo funkcije
- $\nabla f(x_1, \ldots, x_n) = [\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}]^T$, gradijent
- $\nabla^2 f(x_1, \ldots, x_n) = \dfrac{\partial^2 f}{\partial x \partial x^T} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$, Hesijan
- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ redom skup prirodnih, celih, racionalnih, realnih i kompleksnih brojeva
- $\mathbb{R}^+$ pozitivni realni brojevi
- f' oznaka za prvi izvod ili Jakobijan ako je f vektorska funkcija
- $\|\cdot\|$ vektorska ili matrična norma 2 ako nije drugačije naznačeno
- f^{-1} oznaka za inverzno preslikavanje f

Deo I

Pregled teorijskih rezultata

Glava 1

Verovatnoća, slučajne promenljive, konvergencija niza slučajnih promenljivih

1.1 Prostor verovatnoće

Vrši se neki eksperiment i posmatraju se mogući ishodi. Skup svih ishoda se obeležava Ω. U teoriji verovatnoće Ω je proizvoljan neprazan skup, njegove podskupove nazivamo **događaji**.

Primenom skupovnih operacija na događaje A i B dobijaju se novi događaji, na primer: unija $A \cup B$, presek $A \cap B$, koji se kraće obeležava AB, skupovna razlika $A \backslash B$, komplement $\overline{A} := \Omega \backslash A$, simetrična razlika $A \Delta B = (A \backslash B) \cup (B \backslash A)$.

Ako posmatramo prebrojivu familiju događaja, pišemo $A_1, A_2, \ldots$.

DEFINICIJA 1 *Neprazna familija $\mathscr{F}$ podskupova nepraznog skupa Ω je **Borelovo polje** nad Ω ako*

(i) $\forall A \in \mathscr{F}, \overline{A} \in \Omega$,

(ii) *za prebrojivu familiju* $A_1, A_2, \ldots \in \mathscr{F}, \bigcup_{j=1}^{\infty} A_j \in \mathscr{F}$.

Borelovo polje može sadržati sve podskupove od Ω: $\mathscr{F} = 2^{\Omega}$, a može biti i $\mathscr{F} = \{\emptyset, \Omega\}$. Za proizvoljnu familiju $\mathscr{A}$ podskupova Ω presek svih Borelovih polja koja sadrže $\mathscr{A}$ je minimalno Borelovo polje koje sadrži $\mathscr{A}$, kažemo da ga familija $\mathscr{A}$ **generiše**.

Koristeći (i) i (ii) mogu se dokazati druge osobine Borelovog polja, na primer: skupovna razlika i simetrična razlika elemenata Borelovog polja ostaju u polju. Takođe, prebrojiv presek elemenata Borelovog polja ostaje unutar tog Borelovog polja.

Ako za prebrojivu familiju $A_1, A_2, \ldots$ važi

$$\forall k, j \in \{1, 2, \ldots\}, k \neq j \Rightarrow A_k A_j = \emptyset,$$

kažemo da su događaji te familije **disjunktni po parovima**.

PRIMER 1 *Skup intervala oblika* $(a,b]$, $(-\infty,a]$, (b,∞), *gde su* $-\infty < a < b < \infty$, *generiše Borelovo polje nad* $\mathbb{R}$ *koje ćemo obeležavati* $\mathscr{B}_1$, *zvaćemo ga* ***Euklidsko Borelovo polje****.*

Ako dozvolimo da granice intervala budu i vrednosti $-\infty$ *i* ∞, *onda se nad* $\mathbb{R}^* = [-\infty,\infty]$ *generiše* ***prošireno Euklidsko Borelovo polje****, polje koje ćemo obeležavati* $\mathscr{B}^*$.

PRIMER 2 *Skup svih disjunktnih intervala oblika* $(a,b]$, *gde su* $0 < a < b \leq 1$ *generiše Borelovo polje* $\mathscr{F}$ *nad* $\Omega = (0,1]$. *Ako želimo da isti poluotvoreni intervali generišu Borelovo polje nad* $\mathscr{U} = [0,1]$, *dodaćemo i jednoelementni skup* $\{0\}$.

Nad događajima iz $\mathscr{F}$ se posmatra verovatnosna mera, kraće verovatnoća.

DEFINICIJA 2 *Za Borelovo polje* $\mathscr{F}$ *nad nepraznim skupom* Ω, ***verovatnoća****, odnosno* ***verovatnosna mera****, je funkcija* $P : \mathscr{F} \to \mathbb{R}$ *koja zadovoljava*

1. $\forall A \in \mathscr{F}, P(A) \geq 0.$
2. *Za prebrojivu familiju događaja disjunktnih po parovima* $A_1, A_2, \ldots \in \mathscr{F}$,
   $P\left(\bigcup_k A_k\right) = \sum_k P(A_k).$
3. $P(\Omega) = 1.$

Koristeći osobine iz definicije 2: 1. nenegativnost, 2. sigma aditivnost i 3. normiranost, mogu se dokazati i druge **osobine verovatnoće**:

Za proizvoljne događaje A i B:

4. $P(A) \leq 1$
5. $P(\emptyset) = 0$
6. $P(\overline{A}) = 1 - P(A)$
7. $P(A \cup B) = P(A) + P(B) - P(AB)$
8. $A \subseteq B \Rightarrow P(A) \leq P(B)$

Za prebrojivu familiju događaja $A_1, A_2, \ldots$ i događaj A

9. $A_k \uparrow A$ ili $A_k \downarrow A \Rightarrow P(A_k) \to P(A)$

10. $P\left(\bigcup\limits_k A_k\right) \le \sum\limits_k P(A_k)$

Ako je $(\Omega, \mathscr{F})$ Borelovo polje nad kojim je definisana verovatnoća $P(\cdot)$, uređenu trojku $(\Omega, \mathscr{F}, P)$ nazivamo **prostor verovatnoće**.

PRIMER 3 *Za Borelovo polje $(\Omega, \mathscr{F})$ iz primera 2, Borel-Lebegova mera m je verovatnosna mera, $(\Omega, \mathscr{F}, m)$ je prostor verovatnoće.*

PRIMER 4 *Neka je skup ishoda konačan: $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, i neka je Borelovo polje skup svih podskupova $\mathscr{F} = 2^\Omega$.*

Neka je $p_k = P(\{\omega_k\}) \ge 0$, $k = 1, 2, \ldots, n$ i $\sum_{k=1}^n p_k = 1$.

Verovatnoća je definisana

$$P(A) = \sum_{k:\, \omega_k \in A} p_k.$$

*Ovaj prostor zovemo **diskretni prostor verovatnoće**.*

*Ako važi i $p_1 = p_2 = \cdots = p_n = 1/n$, dobijamo $P(A) = \#A/\#\Omega$, to je **klasična definicija** verovatnoće. U klasičnoj definiciji verovatnoće se kaže da je verovatnoća broj povoljnih podeljen sa brojem mogućih ishoda. Klasična definicija verovatnoće se koristi kada imamo konačno mnogo jednako verovatnih ishoda, odnosno, kada se vrši slučajan izbor.*

Ako u prostoru verovatnoće neka osobina važi nad skupom čija verovatnoća je jednaka 1, kažemo da važi **skoro sigurno**. Pišemo **a. s.** od engleskog *almost surely*.

Skup čija verovatnoća je 0 zovemo **nemoguć događaj**. Ako podskupovi nemogućeg događaja pripadaju $\mathscr{F}$, kažemo da je prostor **kompletan**. Ako nije kompletan, prostor se može kompletirati proširivanjem.

1.1.1 Funkcije raspodele

Posmatraćemo **neopadajuće** funkcije $F : \mathbb{R} \to \mathbb{R}$:

$$\forall x_1, x_2 \in \mathbb{R}, x_1 < x_2 \Rightarrow F(x_1) \le F(x_2).$$

Za neopadajuće funkcije u svakoj tački postoje konačna leva i desna granična vrednost: redom

$$F(x-) = \lim_{t \uparrow x} F(t) \text{ i } F(x+) = \lim_{t \downarrow x} F(t).$$

Takođe, postoje i

$$F(-\infty) = \lim_{x \downarrow -\infty} F(x) \text{ i } F(\infty) = \lim_{x \uparrow \infty} F(x),$$

koji mogu biti redom $-\infty$ i ∞. To je tako zato što

$$F(x-) = \sup_{t \in (-\infty, x)} F(t) \text{ i } F(x+) = \inf_{t \in (x, \infty)} F(t).$$

Neopadajuća funkcija F je neprekidna ako i samo ako

$$\forall x \in \mathbb{R}, F(x-) = F(x+).$$

U protivnom, ako postoji $x \in \mathbb{R}, F(x-) \neq F(x+)$, onda to mora biti prekid prve vrste, takozvani **skok** čija je **veličina** $F(x+) - F(x-) > 0$ i $F(x) \in [F(x-), F(x+)]$.

Neopadajuća funkcija može imati najviše prebrojivo skokova.

Neka su F_1 i F_2 neopadajuće funkcije i neka je D skup svuda gust u $\mathbb{R}$. Ako

$$\forall x \in D, F_1(x) = F_2(x),$$

onda F_1 i F_2 imaju iste tačke skokova koji su iste veličine. Vrednost F_1 i F_2 može se razlikovati jedino u samoj tački skoka.

Ako je $\forall x \in \mathbb{R}, F(x) = F(x+)$, kažemo da je F **neprekidna sa desna**.

Ako je $\forall x \in \mathbb{R}, F(x) = F(x-)$, kažemo da je F **neprekidna sa leva**.

DEFINICIJA 3 *Kažemo da je $F : \mathbb{R} \to \mathbb{R}$ **funkcija raspodele** ako je*

(a) *neopadajuća,*

(b) $F(-\infty) = 0$ *i* $F(\infty) = 1$,

(c) *neprekidna sa desna.*

Verovatnosnoj meri μ nad Euklidskim Borelovim poljem $(\mathbb{R}, \mathscr{B}_1)$ pridružujemo **funkciju raspodele** $F : \mathbb{R} \to \mathbb{R}$ formulom

$$F(x) = \mu((-\infty, x]), x \in \mathbb{R}. \tag{1.1}$$

TEOREMA 1 *Za funkciju raspodele za sve $a < b \in \mathbb{R}$ važi:*

$$\begin{aligned} \mu((a,b]) &= F(b) - F(a) \\ \mu((a,b)) &= F(b-) - F(a) \\ \mu([a,b)) &= F(b-) - F(a-) \\ \mu([a,b]) &= F(b) - F(a-). \end{aligned} \tag{1.2}$$

NAPOMENA 1. Ako jednakost (1.1) važi za još neku funkciju raspodele, ali za x iz svuda gustog podskupa $\mathbb{R}$, onda je to ista funkcija raspodele, jer, ako su neopadajuće funkcije, obe neprekidne sa desna, jednake nad svuda gustim skupom, onda su jednake svuda.

NAPOMENA 2. Bilo koja od jednakosti (1.2) za sve a i b nad svuda gustim podskupom $\mathbb{R}$ definiše istu funkciju raspodele.

Osim što verovatnosna mera definiše funkciju raspodele, važi i obrnuto.

TEOREMA 2 *Svaka funkcija raspodele određuje verovatnoću definisanu na $\mathscr{B}_1$ preko bilo koje od jednakosti* (1.2) *ili* (1.1).

Dokaz se daje preko teorije Lebeg-Stiltjesove mere.

Lako se vidi da za prebrojivu familiju disjunktnih intervala $(a_k, b_k], k = 1, 2, \ldots$

$$\mu\left(\bigcup_k (a_k, b_k]\right) = \sum_k (F(b_k) - F(a_k)).$$

Takođe, za elementarni događaj $\{a\}$ verovatnoća je

$$F(a) - F(a-) = F(a+) - F(a-),$$

a to je baš veličina skoka u tački a.

Funkcija raspodele je neopadajuća, stoga ima najviše prebrojivo prekida prve vrste, odnosno, skokova. Neka su za funkciju raspodele F vrednosti $x_1, x_2, \ldots$

tačke u kojima ima skokove i neka su veličine tih skokova redom $p_1, p_2, \ldots$. Označićemo

$$F_d(x) = \sum_k 1_{[x_k,\infty)}(x) p_k. \tag{1.3}$$

Lako se vidi da je F_d neopadajuća, neprekidna sa desna i da je

$$F_d(-\infty) = 0, F_d(\infty) = \sum_k b_k \leq 1.$$

F_d je **diskretni deo** funkcije raspodele. Označimo i **neprekidni deo**

$$F_c(x) = F(x) - F_d(x). \tag{1.4}$$

Može se dokazati da je F_c nenegativna, neopadajuća i neprekidna funkcija.

Svaka funkcija raspodele se na jedinstven način može predstaviti kao zbir neprekidne funkcije i funkcije oblika (1.3), vidi Chung [14].

Ako se funkcija raspodele F može predstaviti u obliku

$$F = \sum_k 1_{[x_k,\infty)} p_k,$$

za prebrojivu familiju vrednosti x_k, gde su p_k nenegativni brojevi i $\sum_k p_k = 1$, kažemo da je F **diskretna funkcija raspodele**.

Ako je funkcija raspodele F neprekidna svuda nad R, kažemo da je F **neprekidna funkcija raspodele**.

Svaka funkcija raspodele se može na jedinstven način predstaviti kao konveksna kombinacija jedne neprekidne i jedne diskretne funkcije raspodele.

Kažemo da je funkcija $F : \mathbb{R} \to \mathbb{R}$ **apsolutno neprekidna** ako postoji Lebeg integrabilna funkcija $f \in L^1(\mathbb{R})$ tako da za svako $x_1, x_2 \in \mathbb{R}$

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)dt. \tag{1.5}$$

Integral u formuli (1.5) je Lebegov integral, a mera koja se koristi je Lebegova mera, obeležavamo je m. Funkcije f koje imaju konačan (Lebegov) integral $\int_a^b f(t)dt$ su, kažemo, **(Lebeg) merljive**, odnosno, pripadaju klasi $L^1(a,b)$.

Ako neka osobina važi nad celim $\mathbb{R}$ osim na skupu mere 0, kažemo da važi **skoro svuda**, odnosno **a. e.** od engleskog *almost everywhere*.

Svaka apsolutno neprekidna funkcija je uniformno neprekidna pa stoga i neprekidna. Svaka Lišic neprekidna funkcija je apsolutno neprekidna.

Ako je funkcija F apsolutno neprekidna nad $\mathbb{R}$, odnosno, ako važi (1.5), onda F ima izvod skoro svuda i on je jednak sa f, odnosno $F' = f$ a. e.

Ako je funkcija raspodele F apsolutno neprekidna, onda funkciju f iz (1.5) nazivamo **gustina** za funkciju raspodele F, i važi

$$f \geq 0 \text{ a. e. i } \int_{-\infty}^{\infty} f(t)dt = 1. \tag{1.6}$$

Obrnuto, ako neka funkcija $f : \mathbb{R} \to \mathbb{R}$ zadovoljava (1.6), onda je

$$F(x) = \int_{-\infty}^{x} f(t)dt \tag{1.7}$$

apsolutno neprekidna funkcija raspodele.

Kažemo da je funkcija F **singularna** ako nije identički jednaka nuli, a njen izvod postoji i jednak je nuli skoro svuda:

$$F \not\equiv 0 \text{ i } F' = 0 \text{ a. e.}$$

Svaka funkcija raspodele može se na jedinstven način predstaviti kao konveksna kombinacija jedne diskretne, jedne singularne neprekidne i jedne apsolutno neprekidne funkcije raspodele.

PRIMER 5 *Ako u Borelovom polju $\mathscr{B}$ generisanom poluotvorenim intervalima i skupom $\{0\}$ nad $\mathscr{U} = [0,1]$ iz primera 2 posmatramo verovatnosnu meru μ, dobijamo prostor verovatnoće $(\mathscr{U}, \mathscr{B}, \mu)$.*

Drugi način da dobijemo isto je da definišemo funkciju raspodele F koja će biti 0 za $x < 0$ i koja će biti 1 za $x \geq 1$. Verovatnoća koja odgovara takvoj funkciji raspodele ima nosač na intervalu $\mathscr{U}$ jer je $\mu((-\infty, 0)) = \mu((1, \infty)) = 0$ (posledica (1.1)).

Za dobijeni prostor verovatnoće možemo reći da je $(\mathscr{U}, \mathscr{B}, \mu)$ restrikcija Euklidskog Borelovog polja $(\mathbb{R}, \mathscr{B}_1, \mu)$ na $\mathscr{U}$ u smislu da je Borelovo polje restrikcija Euklidskog Borelovog polja $\mathscr{B}_1$ na $\mathscr{B}$ i da je verovatnoća restrikcija verovatnoće na $\mathscr{U}$.

Ako uzmemo funkciju raspodele uniformne raspodele na $\mathscr{U}$ *:*

$$F(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

dobijamo verovatnoću koja je jednaka Borelovoj meri na $\mathscr{U}$*. Ako se dobijena Borelova mera kompletira tako da podskupovi nemogućeg događaja budu merljivi i u skladu sa tim dodefiniše verovatnoća, dobijamo uobičajenu Lebegovu meru.*

Analogno se može postupiti sa proizvoljnim konačnim intervalom $[a,b]$ *umesto* $\mathscr{U}$*.*

1.1.2 Uslovna verovatnoća, nezavisnost

Za događaj $B \in \mathscr{F}$, za koji je $P(B) > 0$, definišemo **uslovnu verovatnoću**

$$P(A|B) = \frac{P(AB)}{P(B)}. \tag{1.8}$$

$P(A|B)$ čitamo: *"Verovatnoća događaja A pod uslovom da se desio događaj B".*

Nad familijom događaja $B \cap \mathscr{F} = \{AB | A \in \mathscr{F}\}$ uslovna verovatnoća čini prostor verovatnoće $(B, B \cap \mathscr{F}, P(\cdot|B))$.

Potpun sistem događaja u Borelovom polju $(\Omega, \mathscr{F})$ je konačna ili prebrojiva familija $H_1, H_2, \ldots \in \mathscr{F}$ događaja disjunktnih po parovima koja pokriva skup ishoda Ω: $\bigcup_{k=1}^{\infty} H_k = \Omega$,

Formula totalne verovatnoće
U prostoru verovatnoće $(\Omega, \mathscr{F}, P)$, za potpun sistem događaja $H_1, H_2, \ldots$ i događaj $A \in \mathscr{F}$ važi

$$P(A) = \sum_{k=1}^{\infty} P(H_k) P(A|H_k). \tag{1.9}$$

Ako je $P(A) > 0$ važi i **Bayesova formula**

$$P(H_k|A) = \frac{P(H_k)P(A|H_k)}{P(A)}, \; k = 1, 2, \ldots \tag{1.10}$$

Ako je $P(A|B) = P(A)$, realizacija događaja A ne zavisi od realizacije događaja B. Uvodimo definicije:

Kažemo da su događaji A i B **nezavisni** ako $P(AB) = P(A)\,P(B)$.

Kažemo da je $A_1, A_2, \ldots$ **nezavisna familija događaja** ako za proizvoljni niz indeksa $\{i_1, i_2, \ldots, i_n\}$ važi

$$P(A_{i_1}A_{i_2}\cdots A_{i_n}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_n}).$$

Lema Borel-Cantelli

Za familiju događaja $A_1, A_2, \ldots \in \mathscr{F}$ u prostoru verovatnoće $(\Omega, \mathscr{F}, P)$ važi:

1. Ako $\sum_{k=1}^{\infty} P(A_k) < \infty$, onda $P\left(\bigcap_{k=1}^{\infty}\bigcup_{j=i}^{\infty} A_j\right) = 0$.
2. Ako su $A_1, A_2, \ldots$ nezavisni i $\sum_{k=1}^{\infty} P(A_k) = \infty$, onda $P\left(\bigcap_{k=1}^{\infty}\bigcup_{j=i}^{\infty} A_j\right) = 1$.

1.1.3 Slučajne promenljive

Posmatraćemo funkcije koje ishodima iz Ω pridružuju realne brojeve.

Definicija 4 ***Slučajna promenljiva*** *nad prostorom verovatnoće* $(\Omega, \mathscr{F}, P)$ *je funkcija* $X : \Omega \to \mathbb{R}$ *za koju važi*

$$\forall B \in \mathscr{B}_1, X^{-1}(B) = \{\omega : X(\omega) \in B\} \in \mathscr{F}.$$

Da podsetimo: $\mathscr{B}_1$ je Euklidsko Borelovo polje, generisano poluotvorenim intervalima nad $\mathbb{R}$ (Primer 1).

Malo opštija definicija je **proširena slučajna promenljiva**: umesto $\mathscr{B}_1$ i $\mathbb{R}$ koristi prošireno Euklidsko Borelovo polje $\mathscr{B}^*$ nad $\mathbb{R}^* = [-\infty, \infty]$, uz zahtev da je funkcija X skoro sigurno konačna.

Za funkciju $X : A_1 \to A_2$, gde su $(A_1, \mathscr{F}_1)$ i $(A_2, \mathscr{F}_2)$ Borelova polja, kaže se da je **merljiva** ako važi

$$X^{-1}(\mathscr{F}_2) \subseteq \mathscr{F}_1.$$

Slučajna promenljiva je merljiva funkcija iz skupa ishoda u Euklidsko Borelovo polje.

Osobine inverznog preslikavanja

Za funkciju $X : \Omega \to \mathbb{R}$, važi:

$$X^{-1}(A^C) = (X^{-1}(A))^C$$

$$X^{-1}\left(\bigcup_{j\in\mathscr{J}} A_j\right) = \bigcup_{j\in\mathscr{J}} X^{-1}(A_j)$$
$$X^{-1}\left(\bigcap_{j\in\mathscr{J}} A_j\right) = \bigcap_{j\in\mathscr{J}} X^{-1}(A_j)$$

za proizvoljni indeksni skup $\mathscr{J}$ i elemente $A_j \in \mathscr{B}_1$ i $A \in \mathscr{B}_1$.

Koristeći osobine inverznog preslikavanja možemo dokazati:

TEOREMA 3 *X je slučajna promenljiva nad $(\Omega, \mathscr{F})$ ako i samo ako za sve x iz nekog svuda gustog podskupa $\mathbb{R}$ važi*

$$X^{-1}((-\infty, x]) \in \mathscr{F}. \tag{1.11}$$

Slučajna promenljiva definiše verovatnosnu meru, t.j. verovatnoću nad $\mathscr{B}_1$:

TEOREMA 4 *Ako je X slučajna promenljiva nad prostorom verovatnoće $(\Omega, \mathscr{F}, P)$, onda je $(\mathbb{R}, \mathscr{B}_1, \mu)$ prostor verovatnoće, gde je za $B \in \mathscr{B}_1$*

$$\mu(B) = P(X^{-1}(B)). \tag{1.12}$$

Ako je X slučajna promenljiva, onda je familija $\{X^{-1}(B) : B \in \mathscr{B}_1\}$ Borelovo polje, kažemo da je **generisano** slučajnom promenljivom X. Isto to polje generiše familija skupova $\{\{\omega : X(\omega) \leq x\} : x \in \mathbb{R}\}$.

Verovatnoća iz (1.12) se zove **mera raspodele verovatnoće** slučajne promenljive X.

Slučajnoj promenljivoj X pridružujemo funkciju raspodele:

$$F(x) = P(X^{-1}((-\infty, x])) = \mu((-\infty, x]). \tag{1.13}$$

Ovo pridruživanje je, očigledno, jednoznačno. Obrnuto nije tačno: jednoj funkciji raspodele odgovara više slučajnih promenljivih, za koje kažemo da su **jednako raspoređene**.

Kažemo da je slučajna promenljiva **diskretna** ako postoji najviše prebrojiv skup B takav da je $P(\{\omega : X(\omega) \in B\}) = 1$. To je ekvivalentno sa tim da je njena funkcija raspodele diskretna, odnosno oblika (1.3).

Isto tako, ako slučajnoj promenljivi odgovara apsolutno neprekidna funkcija raspodele, kažemo da je slučajna promenljiva **apsolutno neprekidna** ili kraće **neprekidna**.

PRIMER 6 *Za proizvoljan događaj A iz prostora verovatnoće $(\Omega, \mathscr{F}, P)$ slučajnu promenljivu definisanu 1_A zovemo* ***indikator događaja*** *A.*

PRIMER 7 *Za konačan skup $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ i diskretni prostor verovatnoće $(\Omega, \mathscr{F}, P)$ iz primera 4 proizvoljna funkcija $X : \omega_k \mapsto x_k \in \mathbb{R}$, $k = 1, 2, \dots, n$, je slučajna promenljiva. Za slučajnu promenljivu X funkcija raspodele je*

$$F(x) = \sum_k 1_{[x_k, \infty)}(x) p_k, \text{ gde je } p_k = P(\{\omega : X(\omega) = x_k\}), k = 1, 2, \dots, n.$$

Ako je $f : \mathbb{R} \to \mathbb{R}$ merljiva funkcija u odnosu na Euklidsko Borelovo polje $\mathscr{B}_1$ i ako je X slučajna promenljiva u prostoru verovatnoće $(\Omega, \mathscr{F}, P)$, onda je $f(X) = f \circ X : \Omega \to \mathbb{R}$ slučajna promenljiva u prostoru verovatnoće $(\Omega, \mathscr{F}, P)$.

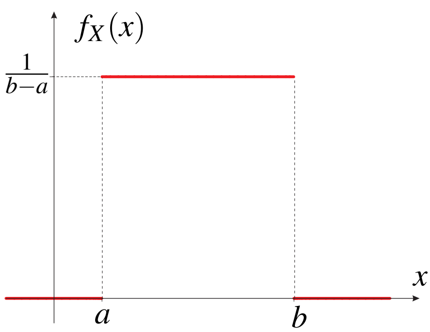
PRIMER 8 *Daćemo primere nekoliko poznatih raspodela.*

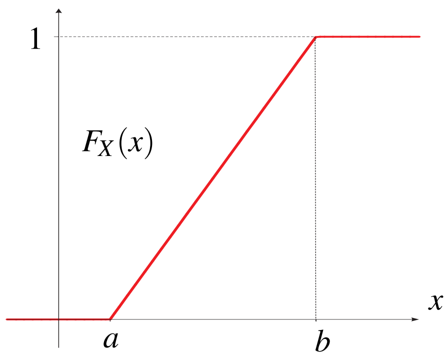
Binomna raspodela $\mathscr{B}(n, p)$, $n \in \mathbb{N}$ i $0 < p < 1$ uzima vrednosti iz skupa $\mathscr{R}_X = \{0, 1, \dots, n\}$, sa verovatnoćama $P(X = k) = \binom{n}{k} p^k q^{n-k}$, $k \in \mathscr{R}_X$, $q = 1 - p$.

Poasonova raspodela $\mathscr{P}(\lambda)$, $\lambda > 0$, $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, \dots$.

Uniformna raspodela $\mathscr{U}(a, b)$, $a < b \in \mathbb{R}$ ima gustinu i funkciju raspodele

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b), \\ 0, & x \notin (a, b); \end{cases} \qquad F_X(x) = \begin{cases} 0, & x \le a, \\ \frac{x-a}{b-a}, & a < x \le b, \\ 1, & x > b. \end{cases}$$

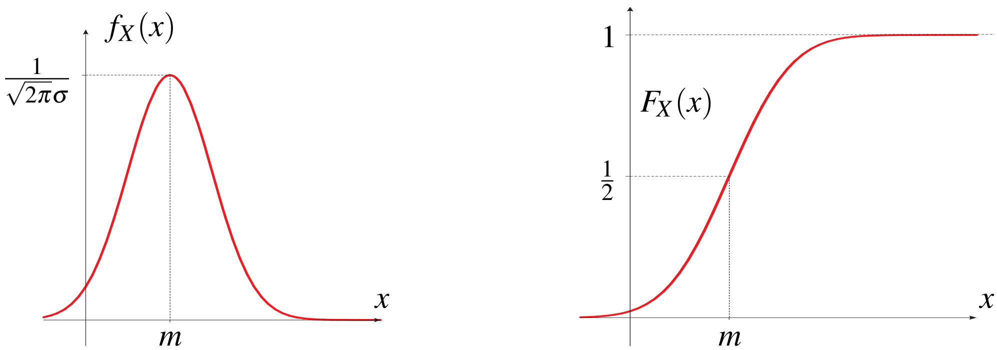




Slika 1.1: Uniformna raspodela, gustina i funkcija raspodele

Normalna (Gausova) raspodela $\mathcal{N}(m,\sigma)$, $m,\sigma\in\mathbb{R}$, $\sigma>0$

$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-m)^2}{2\sigma^2}}$, $x\in\mathbb{R}$; $\qquad F_X(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(t-m)^2}{2\sigma^2}}dt$, $x\in\mathbb{R}$.



Slika 1.2: Normalna raspodela, gustina i funkcija raspodele

Slučajni vektori

Slučajni vektor je vektor čije su sve komponente slučajne promenljive.

Posmatramo dvodimenzionalni slučajni vektor (X,Y) nad $(\Omega,\mathscr{F},P)$ i nad $\mathbb{R}^2$ posmatramo dvodimenzionalno Euklidsko Borelovo polje $\mathscr{B}_2$ generisano skupovima oblika $B_1\times B_2$, gde su $B_1,B_2\in\mathscr{B}_1$.

Ako su X i Y slučajne promenljive nad prostorom verovatnoće $(\Omega,\mathscr{F},P)$ onda inverzno preslikavanje

$$(X,Y)^{-1}(A)=\{\omega:(X(\omega),Y(\omega))\in A\},\ \text{za } A\in\mathscr{B}_2$$

definiše verovatnoću nad $(\Omega,\mathscr{B}_2)$, isto kao u jednodimenzionalnom slučaju.

Dalje, merljiva funkcija $f(x,y):\mathbb{R}^2\to\mathbb{R}$ u smislu da inverzne slike iz $\mathscr{B}_1$ pripadaju $\mathscr{B}_2$ primenjena na slučajne promenljive X i Y definiše novu slučajnu promenljivu koju označavamo $Z=f(X,Y)$.

Tako dobijamo da ako su X i Y slučajne promenljive, onda su to i

$$f(X,Y),\ \text{ako je } f \text{ neprekidna},\ e^{aX},\ e^{itX},\ \text{za } a,t\in\mathbb{R},$$

$$\max(X,Y),\ \min(X,Y),\ X+Y,\ XY,\ \text{kao i } X/Y \text{ za nenula } Y.$$

Prelazak iz dvodimenzionalnog slučaja u višedimenzionalan je direktan.

Prelazak na prebrojiv slučaj daje da ako su $X_1, X_2, \ldots$ prebrojiva familija slučajnih promenljivih, onda su

$$\inf_k X_k,\ \sup_k X_k,\ \liminf_k X_k,\ \limsup_k X_k,$$

slučajne promenljive (koje ne moraju biti skoro svuda konačne). Takođe

$$\lim_{k\to\infty} X_k$$

je slučajna promenljiva na skupu gde je moguća samo konvergencija ili divergencija ka $\pm\infty$.

Višedimenzionalnoj slučajnoj promenljivi (X,Y) se pridružuje višedimenzionalna funkcija raspodele:

$$F(x,y) = P(\{\omega : X(\omega) < x, Y(\omega) < y\}),$$

koja može biti definisana preko višedimenzionalne funkcije gustine raspodele $f(x,y) \geq 0$:

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv.$$

Prelazak sa $n = 2$ dimenzije na $n > 2$ dimenzija je direktan.

PRIMER 9 ***Višedimenzionalna normalna raspodela*** *X je definisana funkcijom gustine raspodele za n-dimenzionalni vektor $x = (x_1, \ldots x_n)$*

$$f_X(x) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right),$$

gde je μ n-dimenzionalni vektor, a Σ je $n \times n$, simetrična, pozitivno definitna matrica.

Važi tvrđenje da višedimanezionalna slučajna promenljiva X ima normalnu raspodelu ako i samo ako proizvoljna linearna kombinacija $c^T X$, za konstantni vektor c ima normalnu raspodelu. Ovde su dozvoljene i deformisane normalne raspodele za $c = 0$.

Nezavisnost slučajnih promenljivih

Slučajne promenljive $X_1, X_2, \dots, X_n$ na prostoru verovatnoće $(\Omega, \mathscr{F}, P)$ su **nezavisne** ako za sve Borelove skupove $B_1, B_2, \dots, B_n \in \mathscr{B}_1$ važi

$$P(\{X_1 \in B_1\} \cap \{X_2 \in B_2\} \cap \dots \cap \{X_n \in B_n\}) =$$

$$= P(\{X_1 \in B_1\}) P(\{X_2 \in B_2\}) \cdots P(\{X_n \in B_n\}) .$$

Familija slučajnih promenljivih $\{X_i\}_{i \in I}$ je **nezavisna** ako su za proizvoljni skup indeksa $\{i_1, i_2, \dots, i_n\} \subseteq I$ slučajne promenljive $X_{i_1}, X_{i_2}, \dots, X_{i_n}$ nezavisne. Familija slučajnih promenljivih $\{X_i\}_{i \in I}$ je **nezavisna po parovima** ako su za svaka dva indeksa i, j, $i \neq j$, slučajne promenljive X_i i X_j nezavisne.

Često se posmatraju slučajne promenljive $X_1, X_2, \dots, X_n$ koje su nazavisne i imaju istu raspodelu, kažemo da su **i. i. d.** od engleskog **independent and identically distributed**.

Ako su $X_1, X_2, \dots, X_n$ nezavisne, apsolutno neprekidne slučajne promenljive sa redom gustinama raspodele $f_1, f_2, \dots, f_n$, onda je $(X_1, X_2, \dots, X_n)$ apsolutno neprekidni slučajni vektor sa gustinom raspodele

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n) \text{ za } (x_1, x_2, \dots, x_n) \in \mathbb{R}^n .$$

PRIMER 10 *U primeru 9, svaka komponenta n-dimenzionalne normalne raspodele ima normalnu raspodelu. Obrnuto ne mora da važi, moguće je naći dve ili više (zavisnih) normalnih raspodela čija zajednička raspodela nema višedimenzionalnu normalnu raspodelu.*

1.1.4 Matematičko očekivanje slučajne promenljive

Matematičko očekivanje (kraće "očekivanje") je uopštenje pojma integrala. Pretpostavimo da je slučajna promenljiva koju posmatramo konačna skoro svuda.

U prostoru verovatnoće $(\Omega, \mathscr{F}, P)$, P je mera u Borelovom polju $\mathscr{F}$. Slučajna promenljivu $X : \Omega \to \mathbb{R}$ je merljiva funkcija. **Matematičko očekivanje** slučajne promenljive X definišemo kao Lebegov integral nad Ω u odnosu na meru P u oznaci

$$E(X) = \int_\Omega X(\omega) P(d\omega), \tag{1.14}$$

kad taj integral funkcije X postoji.

Koristeći očekivanje (1.14) može se definisati i integral funkcije X nad proizvoljnim događajem $A \in \mathscr{F}$:

$$\int_A X(\omega)P(d\omega) = E(1_A X), \tag{1.15}$$

kada je taj integral konačan. Kraće obeležavamo $\int_A XdP$, čitamo: "integral od X u odnosu na meru P na skupu A".

Definicija očekivanja uključuje kao moguće i vrednosti $+\infty$ i $-\infty$.

Po definiciji Lebegovog integrala, vrednost (1.15) postoji ako i samo ako postoji $\int_A |X|\,dP$.

Ako za neku slučajnu promenljivu postoji očekivanje (1.14), onda postoji i integral (1.15).

Kada se posmatra nad prostorom verovatnoće $(\mathbb{R}, \mathscr{B}_1, \mu)$, integral (1.15) je Lebeg-Stiltjesov integral, obično obeležavamo $f = X$, $x = \omega$:

$$\int_A XdP = \int_A f(x)\,\mu(dx).$$

Kad je oblast integracije $A = (a,b]$ i kada je F funkcija raspodele verovatnosne mere μ, onda se poslednji Lebeg- Stiltjesov integral zapisuje i kao

$$\int_A XdP = \int_{(a,b]} f(x)\,dF(x), \text{ odnosno } \int_A XdP = \int_{a+}^{b+} f(x)\,dF(x).$$

Poslednji zapis se modifikuje ako se traži integral nad: $[a,b]$, (a,b), $[a,b)$ u, redom: $\int_{a-}^{b+}, \int_{a+}^{b-}, \int_{a-}^{b-}$.

Ako posmatramo prostor $(\mathscr{U}, \mathscr{B}, m)$, sa Lebegovom merom m, onda je to običan Lebegov integral nad $\mathbb{R}$

$$\int_A XdP = \int_a^b f(x)m(dx) = \int_a^b f(x)dx,$$

što je moguće bez opisivanja da li su krajnje tačke uključene, jer Lebegova mera "ne oseti" tačku, odnosno, skupove mere 0.

Praktično se očekivanje najčešće računa koristeći rezultat sledeće teoreme:

TEOREMA 5 *Neka je $(\Omega, \mathscr{F}, P)$ prostor verovatnoće i neka je $f : \mathbb{R} \to \mathbb{R}$ merljiva funkcija.*

1. *Ako je X diskretna slučajna promenljiva nad $(\Omega, \mathscr{F}, P)$ sa prebrojivim skupom vrednosti $x_1, x_2, \ldots$ i verovatnoćama $p_k = P(\{\omega | X(\omega) = x_k\})$, za $k = 1, 2, \ldots$, za koje važi $\sum_k p_k = 1$ i ako je $\sum_k |f(x_k)|\, p_k < \infty$, onda je*

$$E(f(X)) = \sum_k f(x_k)\, p_k, \ \text{ specijalno, za } f(x) = x \ \text{ imamo } E(X) = \sum_k x_k\, p_k.$$

2. *Ako je X apsolutno neprekidna slučajna promenljiva sa gustinom raspodele $\varphi(x)$ i važi $\int_{\mathbb{R}} |f(x)|\, \varphi(x)\, dx < \infty$, onda je*

$$E(f(X)) = \int_{\mathbb{R}} f(x)\, \varphi(x)\, dx, \ \text{ za } f(x) = x \ \text{ imamo } E(X) = \int_{\mathbb{R}} x\, \varphi(x)\, dx.$$

Očekivanje je vrednost oko koje se grupišu moguće vrednosti slučajne promenljive.

Postoje i druge bitne karakteristike slučajne promenljive koje se mogu opisati običnim i centralnim momentima k-tog reda:

DEFINICIJA 5 ***Momenat** k**-tog reda** slučajne promenljive X je $E(X^k)$.*
***Centralni momenat** k**-tog reda** slučajne promenljive X je $E((X - E(X))^k)$.*

Specijalno, centralni momenat drugog reda se zove **varijansa** i označava $\mathrm{Var}(X) = E((X - E(X))^2)$, a koren iz varijanse se zove **standardna devijacija** slučajne promenljive, $\sigma(X) = \sqrt{\mathrm{Var}(X)}$.

Postoje dva zanimljiva numerička obeležja slučajnih promenljivih koja se vezuju za oblik krive gustine raspodele. To su

$$\textbf{zakrivljenost (kurtosis)}\text{: } \mathrm{Kurt}\,(X) = \frac{E((X - E(X))^4)}{(E((X - E(X))^2))^2}$$

$$\text{i } \textbf{nagnutost (skewness)}\text{: } \mathrm{Skew}\,(X) = \frac{E((X - E(X))^3)}{(E((X - E(X))^2))^{(3/2)}}\,.$$

Zakrivljenost krive normalne raspodele je 3 i ona se koristi kao etalon za poređenje raspodela.

Nagnutost za simetrične gustine raspodele je 0, na primer normalna raspodela. Odstupanje od 0 meri koliko je neka kriva gustine raspodele pozitivno ili negativno nagnuta.

U statistici se pomoću ova dva parametra može prepoznati da li uzorak neke raspodele odgovara normalnoj raspodeli ili kojoj raspodeli ili grupi raspodela posmatrani uzorak odgovara.

TEOREMA 6 ***Nejednakost Čebiševa.*** *Za nenegativnu (a. s.) slučajnu promenljivu X nad prostorom verovatnoće $(\Omega, \mathscr{F}, P)$ i proizvoljno $\varepsilon > 0$ važi*

$$P(X \geq \varepsilon) \leq \frac{E(X)}{\varepsilon}.$$

Ako je $E(X) < \infty$ i postoji $\mathrm{Var}(X)$, onda za proizvoljno $\varepsilon > 0$ važi

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

PRIMER 11 *Za slučajne promenljive iz primera 8 postoje očekivanje i varijansa i dati su u tabeli što sledi.*

	$E(X)$	$\mathrm{Var}(X)$
$\mathscr{B}(n,p)$	np	$np(1-p)$
$\mathscr{P}(\lambda)$	λ	λ
$\mathscr{U}(a,b)$	$(a+b)/2$	$(b-a)^2/12$
$\mathscr{N}(m,\sigma)$	m	σ^2

DEFINICIJA 6 *Za slučajne promenljive X i Y sa pozitivnim, konačnim, varijansama,* ***kovarijansa*** *je*

$$\mathrm{cov}(X,Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y),$$

a ***koeficijent korelacije*** *je*

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$

Može se dokazati: $|\rho(X,Y)| \leq 1$. Takođe, ako su X i Y nezavisne, onda je $E(XY) = E(X)E(Y)$, onda je $\mathrm{cov}(X,Y) = \rho(X,Y) = 0$. Obrnuto ne mora biti tačno.

DEFINICIJA 7 *Ako je $X = (X_1, X_2, \ldots, X_n)$ slučajni vektor i važi $E(X_k^2) < \infty$, $k = 1,2,\ldots,n$,* ***kovarijansna matrica*** *slučajnog vektora X je*

$$\mathrm{cov}(X) = [\mathrm{cov}(X_i, X_j)]_{i,j=1,2,\ldots,n}.$$

Može se dokazati da je $\mathrm{cov}(X)$ simetrična, pozitivno semidefinitna matrica.

Uslovno očekivanje

Neka je $(\Omega, \mathscr{F}, P)$ prostor verovatnoće. Ako je $\mathscr{F}_1 \subseteq \mathscr{F}$ i $\mathscr{F}_1$ je σ-algebra, kažemo da je $\mathscr{F}_1$ σ-podalgebra σ-algebre $\mathscr{F}$.

Teorema 7 ***Radon-Nikodym teorema.*** *Neka je $(\Omega, \mathscr{F}, P)$ prostor verovatnoće i neka je $\mathscr{F}_1$ σ-podalgebra σ-algebre $\mathscr{F}$. Za slučajnu promenljivu X, za koju je barem jedno od $E(X^+)$ i $E(X^-)$ konačno, postoji jedinstvena $\mathscr{F}_1$ merljiva slučajna promenljiva koju označavamo $E(X|\mathscr{F}_1)$ takva da za sve $A \in \mathscr{F}_1$ važi*

$$\int_A X\,dP = \int_A E(X|\mathscr{F}_1)\,dP \text{ a. s.}$$

Slučajnu promenljivu $E(X|\mathscr{F}_1)$ iz prethodne teoreme zovemo **uslovno očekivanje** slučajne promenljive X nad $\mathscr{F}_1$.

Ako je Y slučajna promenljiva onda uslovno očekivanje definišemo

$$E(X|Y) = E(X|\sigma(Y)) \text{ a. s.}$$

gde je $\sigma(Y)$ σ-algebra generisana slučajnom promenljivom Y, što znači da je generisana familijom skupova $\{\omega : Y(\omega) \leq x\}_{x\in\mathbb{R}}$.

Za proizvoljni događaj $A \in \mathscr{F}_1 \subseteq \mathscr{F}$ definišemo uslovnu verovatnoću u odnosu na $\mathscr{F}_1$

$$P(A|\mathscr{F}_1) = E(1_A|\mathscr{F}_1).$$

Osobine uslovnog očekivanja

Neka je $(\Omega, \mathscr{F}, P)$ prostor verovatnoće i neka je $\mathscr{F}_1$ σ-podalgebra σ-algebre $\mathscr{F}$. Neka su X i Y slučajne promenljive i neka postoje $E(X)$ i $E(Y)$. Onda važi

1. Ako je $X = c = const$ a. s. onda je $E(X|\mathscr{F}_1) = c$ a. s.
2. Ako je $X \leq Y$ a. s. onda je $E(X|\mathscr{F}_1) \leq E(Y|\mathscr{F}_1)$ a. s.
3. $|E(X|\mathscr{F}_1)| \leq E(|X|\,|\mathscr{F}_1)$ a. s.
4. Za proizvoljno $a, b \in \mathbb{R}$, $E(aX + bY|\mathscr{F}_1) = aE(X|\mathscr{F}_1) + bE(Y|\mathscr{F}_1)$ a. s.
5. $E(X|\{\emptyset, \Omega\}) = E(X)$ a. s.

6. $E(X|\mathscr{F}) = X$ a. s.

7. $E(E(X|\mathscr{F}_1)) = E(X)$

8. Ako su $\mathscr{F}_1$ i $\mathscr{F}_2$ σ-podalgebre $\mathscr{F}$ i važi $\mathscr{F}_2 \subseteq \mathscr{F}_1$, onda $E(E(X|\mathscr{F}_1)|\mathscr{F}_2) = E(X|\mathscr{F}_2)$ a. s.

9. Ako je Y $\mathscr{F}_1$-merljiva, $E(|X|) < \infty$ i $E(|Y|) < \infty$, onda $E(XY|\mathscr{F}_1) = X\,E(Y|\mathscr{F}_1)$ a. s.

1.1.5 Konvergencija niza slučajnih promenljivih

Konvergencija niza slučajnih promenljivih $X_1, X_2, \ldots$ (niza $\{X_n\}$), nad prostorom verovatnoće $(\Omega, \mathscr{F}, P)$ se može definisati na više načina. Već smo pomenuli definiciju "skoro svuda", ubuduće ćemo reći "skoro sigurno".

Definicija 8 *Kažemo da niz slučajnih promenljivih $\{X_n\}$ konvergira **skoro svuda** ili **skoro sigurno** ka slučajnoj promenljivi X (pišemo $X_n \to X$ a. s.) ako postoji skup mere nula N tako da*

$$\forall \omega \in \Omega \backslash N, \ \lim_{n\to\infty} X_n(\omega) = X(\omega) \neq \pm\infty\,.$$

Mi posmatramo slučajne promenljive koje uzimaju vrednosti $\pm\infty$ na skupu mere nula. Onda naš zahtev da granica bude konačna nije problem, jer sve promenljive u nizu koji posmatramo mogu uzeti vrednost $\pm\infty$ na prebrojivoj uniji skupova mere 0 jer prebojiva unija ostaje mere 0, može se čak dodati na skup N iz definicije.

Daćemo bez dokaza teoremu iz Chung [14] koja karakteriše skoro sigurnu konvergenciju.

Teorema 8 *Niz slučajnih promenljivih $\{X_n\}$ konvergira skoro sigurno ka slučajnoj promenljivi X ako i samo ako $\forall \varepsilon > 0$*

$$\lim_{m\to\infty} P(\{\omega : \forall n \geq m,\, |X_n(\omega) - X(\omega)| \leq \varepsilon\}) = 1$$

ili ekvivalentno

$$\lim_{m\to\infty} P(\{\omega : \exists n \geq m,\, |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0\,.$$

DEFINICIJA 9 *Kažemo da niz slučajnih promenljivih $\{X_n\}$ **konvergira u verovatnoći** ka slučajnoj promenljivi X (pišemo $X_n \to X$ pr.) ako $\forall \varepsilon > 0$*

$$\lim_{n\to\infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}) = 0\,.$$

DEFINICIJA 10 *Kažemo da niz slučajnih promenljivih $\{X_n\}$ sa funkcijama raspodele $F_{X_n}(x)$ **konvergira u raspodeli** ka slučajnoj promenljivi X sa funkcijom raspodele $F_X(x)$ (pišemo $X_n \to X$ dist.) ako $\forall x \in C_F$*

$$\lim_{n\to\infty} F_n(x) = F(x),$$

gde je C_F skup tačaka neprekidnosti F_X.

TEOREMA 9 *Ako $X_n \to X$ a. s. onda $X_n \to X$ pr.*

TEOREMA 10 *Ako $X_n \to X$ pr. onda $X_n \to X$ dist.*

DEFINICIJA 11 *Kažemo da niz slučajnih promenljivih $\{X_n\}$ **konvergira u L^p**, $0 < p < \infty$, ka slučajnoj promenljivi X, pišemo $X_n \to X$ u L^p, ako $X_n \in L^p$, $n = 1, 2, \ldots$, $X \in L^p$ i važi*

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0\,.$$

TEOREMA 11 *Ako $X_n \to X$ u L^p, $0 < p < \infty$, onda $X_n \to X$ pr.*

TEOREMA 12 *Ako $X_n \to 0$ pr. i postoji slučajna promenljiva $Y \in L^p$ tako da je $|X_n| \le Y$ a. s. onda $X_n \to 0$ u L^p.*

Postoje primeri nizova slučajnih promenljivih koji pokazuju da konvergencija u verovatnoći ne implicira skoro sigurnu konvergenciju.

Takođe, skoro sigurna konvergencija i konvergencija u L^p ne impliciraju jedna drugu.

1.1.6 Granične teoreme

Zakoni velikih brojeva daju konvergenciju aritmetičke sredine niza slučajnih promenljivih.

Zajedno sa centralnim graničnim teoremama, koje daju uslove pod kojima niz slučajnih promenljivih teži ka normalnoj raspodeli, često se koriste u statistici i drugim primenama teorije verovatnoće.

Za niz slučajnih promenljivih $\{X_n\}$ kažemo da važi **slabi zakon velikih brojeva** ako niz slučajnih promenljivih

$$\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(x_k) \tag{1.16}$$

konvergira u verovatnoći ka 0. Ako niz (1.16) konvergira ka 0 skoro sigurno (a. s.), kažemo da za niz $\{X_n\}$ važi **jaki zakon velikih brojeva**.

TEOREMA 13 ***Slabi zakon velikih brojeva Čebiševa.*** *Ako je $\{X_n\}$ niz po parovima nezavisnih slučajnih promenljivih sa konačnim varijansama Var (X_n) i važi Var $(X_n) \leq c$, $n = 1, 2, \ldots$ za neko c, onda za niz $\{X_n\}$ važi slabi zakon velikih brojeva.*

TEOREMA 14 ***Slabi zakon velikih brojeva Hinčina.*** *Ako je $\{X_n\}$ niz i. i. d. slučajnih promenljivih sa konačnim matematičkim očekivanjem, onda za niz $\{X_n\}$ važi slabi zakon velikih brojeva.*

TEOREMA 15 ***Jaki zakon velikih brojeva Kolmogorova.*** *Ako je $\{X_n\}$ niz nezavisnih slučajnih promenljivih takvih da je red $\sum_{n=1}^{\infty} \frac{\mathrm{Var}(X_n)}{n^2}$ konvergentan, onda za niz $\{X_n\}$ važi jaki zakon velikih brojeva.*

Posledica ove teoreme je Borelov jaki zakon velikih brojeva koji kaže da niz $\frac{S_n}{n} \to p$ a. s. gde $S_n : \mathscr{B}(n, p)$.

TEOREMA 16 ***Centralna granična teorema.*** *Ako je $\{X_n\}$ niz i. i. d. slučajnih promenljivih takvih da za sve $n = 1, 2, \ldots$ važi $E(X_n) = m$ i Var $(X_n) = \sigma^2$ onda*

$$\frac{\sum_{k=1}^{n} X_k - nm}{\sigma\sqrt{n}} \to Z \ \ dist.$$

gde je Z slučajna promenljiva sa normalnom $\mathscr{N}(0, 1)$ raspodelom.

1.1.7 Martingali

Filtracija je familija σ-algebri $\{\mathscr{F}_t\}_{t\in T}$ za koje važi $\forall t \leq s \in T$, $\mathscr{F}_t \subseteq \mathscr{F}_s$, gde je T beskonačan skup uređen relacijom $\leq$.

Familiju slučajnih promenljivih $\{X_t\}_{t\in T}$ zovemo **slučajni proces**. Mi ćemo posmatrati slučajne procese sa diskretnim "vremenom" $T = \{0,1,2,\ldots\}$, a često se koristi i $T = \mathbb{R}^+ \cup \{0\}$.

Ako je za slučajni proces $\{X_t\}$ svaka slučajna promenljiva X_t $\mathscr{F}_t$-merljiva, kažemo da je $(X_t, \mathscr{F}_t)$ **adaptirani** proces.

Ako je $\forall t \in T$, $E(|X_t|) < \infty$ onda adaptirani proces $(X_t, \mathscr{F}_t)$ zovemo
martingal ako važi $\forall t > s$, $E(X_t|\mathscr{F}_s) = X_s$ a. s.
supermartingal ako važi $\forall t > s$, $E(X_t|\mathscr{F}_s) \leq X_s$ a. s.
submartingal ako važi $\forall t > s$, $E(X_t|\mathscr{F}_s) \geq X_s$ a. s.

Adaptirani proces $(X_t, \mathscr{F}_t)$ zovemo **martingal niz razlika** ako važi

$$E(X_t|\mathscr{F}_{t-1}) = 0 \text{ a. s. za } t \geq 1.$$

Na primer, niz nezavisnih slučajnih promenljivih koje imaju očekivanje 0 je martingal niz razlika.

Teoreme o konvergenciji martingala

TEOREMA 17 ***Dubova (Doob) teorema konvergencije martingala***. *Neka je $(X_n, \mathscr{F}_n)_{n\in\mathbb{N}}$ submartingal i neka važi $\sup_n E(X_n^+) < \infty$ a. s. Onda postoji slučajna promenljiva X za koju važi $E(|X|) < \infty$ i*

$$\lim_{n\to\infty} X_n = X \text{ a. s.}$$

Posledica 1: Ako je $(X_n, \mathscr{F}_n)$ nenegativan supermartingal ili nepozitivan submartingal, onda postoji slučajna promenljiva X za koju

$$\lim_{n\to\infty} X_n = X \text{ a. s. i važi } E(X) < \infty.$$

Posledica 2: Ako je $(X_n, \mathscr{F}_n)$ martingal za koji važi $\sup_n E(X_n) < \infty$ a. s. onda postoji slučajna promenljiva X za koju

$$\lim_{n\to\infty} X_n = X \text{ a. s. i važi } E(|X|) < \infty.$$

Posledica 2 se često zove **teorema o konvergenciji martingala**.

Ova posledica sledi iz teoreme 17 jer:

1. martingal je submartingal

2. $E(X_n) = E(X_1) \Rightarrow$
$E(|X_n|) = E(X_n^+) + E(X_n^-) = 2E(X_n^+) - E(X_n) = 2E(X_n^+) - E(X_1) \Rightarrow$
$\sup_n E(X_n^+) = \frac{1}{2}\sup_n E(E(|X_n|) + E(X_1)) = \frac{1}{2}\sup_n E(|X_n|) + \frac{1}{2}E(X_1) < \infty$

Sledeća posledica sledi iz posledice 1 i 2 i koristi se u teoriji konvergencije stohastičke aproksimacije, Chen [13].

Posledica 3: Neka su $(X_n, \mathscr{F}_n)$ i $(Y_n, \mathscr{F}_n)$ dva nenegativna adaptirana procesa.

1. Ako je $E(X_{n+1}|\mathscr{F}_n) \le X_n + Y_n$ i $E(\sum_n Y_n) < \infty$, onda X_n konvergira skoro sigurno i ima konačnu granicu.

2. Ako je $E(X_{n+1}|\mathscr{F}_n) \le X_n - Y_n$ onda $E(\sum_n Y_n) < \infty$ a. s.

Sledi i teorema o konvergenciji martingala koja se najčešće koristi u teoriji metoda stohastičke aproksimacije. To je teorema o skoro sigurnoj konvergenciji nenegativnih skoro supermartingala, Robbins i Siegmund [50].

TEOREMA 18 *Neka je $(\Omega, \mathscr{F}, P)$ prostor verovatnoće i neka je $\mathscr{F}_1 \subseteq \mathscr{F}_2 \subseteq \cdots \subseteq \mathscr{F}$ niz σ-podalgebri $\mathscr{F}$. Neka su U_n, β_n, ξ_n i ζ_n, $n = 1, 2, \ldots$ nenegativne $\mathscr{F}_n$-merljive slučajne promenljive za koje važi*

$$E(U_{n+1}|\mathscr{F}_n) \le (1 + \beta_n)U_n + \xi_n - \zeta_n, \; n = 1, 2, \ldots$$

Onda na skupu $\{\sum_n \beta_n < \infty, \sum_n \xi_n < \infty\}$ niz U_n konvergira skoro sigurno ka slučajnoj promenljivi i $\sum_n \zeta_n < \infty$ a. s.

Glava 2

Problemi nelinearne optimizacije bez ograničenja

2.1 Problem matematičkog programiranja

DEFINICIJA 12 *Problem **matematičkog programiranja** je nalaženje $x^* \in X$ za koje je*

$$f(x^*) = \min_{x \in X} f(x), \tag{2.1}$$

*gde je $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ **funkcija cilja**, $X \subseteq \mathbb{R}^n$ je **skup dopustivih vrednosti**.*

Drugim rečima, matematičko programiranje je minimizacija realne funkcije nad skupom dopustivih vrednosti. Problem je ekvivalentan maksimizaciji funkcije $-f$.

Ako je skup dopustivih vrednosti $X = \mathbb{R}^n$, govorimo o **optimizaciji bez ograničenja**.

Za probleme sa ograničenjima osnovu postupaka rešavanja čine postupci za probleme bez ograničenja [20, 47]. Ovde nećemo razmatrati probleme sa ograničenjima.

Problem (2.1) može imati više rešenja.

DEFINICIJA 13 *Ako postoji $x^* \in X$ takvo da za sve $x \in X$ važi $f(x^*) \leq f(x)$, kažemo da je x^* **globalni minimum** funkcije f nad X.*

DEFINICIJA 14 *Ako postoji $x^* \in X$ i okolina $\mathscr{N}(x^*)$ takva da za sve $x \in \mathscr{N}(x^*) \cap X$ važi $f(x^*) \leq f(x)$, kažemo da je x^* **lokalni minimum** za f.*

Algoritmi koje ćemo posmatrati su kreirani tako da nalaze stacionarnu tačku, koja je pod dodatnim uslovima lokalni minimum.

Nekad se lokalni minimum iz definicije 14 naziva **slabi** lokalni minimum nasuprot **strogog** lokalnog minimuma.

DEFINICIJA 15 *Kažemo da je x^* **strogi lokalni minimum** funkcije f ako postoji okolina $\mathscr{N}(x^*)$ takva da za sve $x \in \mathscr{N}(x^*) \backslash \{x^*\} \cap X$ važi $f(x^*) < f(x)$.*

DEFINICIJA 16 *Tačka $x^* \in X$ je* ***izolovani lokalni minimum*** *funkcije f ako postoji okolina $\mathcal{N}(x^*)$ takva da je u njoj x jedini lokalni minimum.*

Izolovani lokalni minimum mora biti strogi lokalni minimum.

Glavna alatka koju ćemo koristiti u dokazima i ispitivanju ponašanja funkcije je **Tejlorova teorema**, koja daje razvoj funkcije f u okolini tačke x:

TEOREMA 19 (Taylor) *Neka je funkcija $f : \mathbb{R}^n \to \mathbb{R}$ neprekidno diferencijabilna i neka $p \in \mathbb{R}^n$. Tada, za neko $t \in (0,1)$*

$$f(x+p) = f(x) + \nabla f(x+tp)^T p\,. \tag{2.2}$$

Ako je, pritom, f dva puta neprekidno diferencijabilna, onda

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp)p\,dt\,, \tag{2.3}$$

odnosno, za neko $t \in (0,1)$

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2}p^T \nabla^2 f(x+tp)p\,. \tag{2.4}$$

Koristeći Tejlorovu teoremu mogu se dokazati teoreme o potrebnim uslovima i teorema o dovoljnom uslovu za lokalni ekstrem funkcije.

TEOREMA 20 (Potreban uslov prvog reda) *Neka je x^* lokalni minimum funkcije f i neka je f neprekidno diferencijabilna u otvorenoj okolini od x^*. Onda je $\nabla f(x^*) = 0$.*

DEFINICIJA 17 *Za neprekidno diferencijabilnu funkciju $f : \mathbb{R}^n \to \mathbb{R}$ tačka $x^* \in X$ je* ***stacionarna tačka*** *ako $\nabla f(x^*) = 0$.*

Posledica teoreme 20 je da su za neprekidno diferencijabilnu funkciju svi lokalni minimumi stacionarne tačke.

TEOREMA 21 (Potreban uslov drugog reda) *Neka je x^* lokalni minimum funkcije f i neka je $\nabla^2 f$ neprekidno u otvorenoj okolini od x^*. Onda je $\nabla f(x^*) = 0$ i $\nabla^2 f(x^*)$ je pozitivno semidefinitno.*

Prvi deo ove teoreme je posledica teoreme 20, drugi deo sledi iz Tejlorove teoreme.

TEOREMA 22 (Dovoljan uslov drugog reda) *Neka je $\nabla^2 f$ neprekidno u otvorenoj okolini x^*, neka je $\nabla f(x^*) = 0$ i neka je $\nabla^2 f(x^*)$ pozitivno definitno. Onda je x^* strogi lokalni minimum od f.*

TEOREMA 23 *Za konveksnu funkciju f lokalni minimum je i globalni minimum. Ako je f diferencijabilna i konveksna, stacionarna tačka je i globalni minimum.*

Ova teorema omogućava da globalni optimum za diferencijabilnu konveksnu funkciju dobijemo rešavajući sistem jednačina $\nabla f(x) = 0$.

Sa druge strane, rešavanje **sistema jednačina** $r(x) = 0$, gde je r vektorska funkcija n promenljivih $r : \mathbb{R}^n \to \mathbb{R}^n$ se može postaviti kao problem minimizacije funkcije

$$\min_x \|r(x)\|^2 .$$

Traži se x^*, minimum funkcije $\|r(x^*)\|^2$, za koji važi $\|r(x^*)\|^2 = 0$.

U praksi se vrlo često pojavljuju problemi rešavanja sistema (nelinearnih) jednačina sa predefinisanim sistemom (kad imamo više jednačina od nepoznatih). U tom slučaju obično ne postoji rešenje. Ipak, tada se mogu naći vrednosti parametara koje "najbolje" odgovaraju datom modelu. Pritom najbolje mislimo u smislu sume kvadrata odstupanja.

Ako su date jednačine $r_j(x) = 0$, $j = 1, 2, \ldots, m$, gde su r_j glatke funkcije i $m \geq n$, **problem najmanjih kvadrata (least squares problem)** je minimizacija funkcije

$$f(x) = \frac{1}{2} \min_x \sum_{j=1}^{m} r_j^2(x) .$$

Problem najmanjih kvadrata se pojavljuje u puno oblasti, kad želimo da zadati model koji zavisi od n parametara najbolje prilagodimo datim vrednostima u m tačaka.

Sistem jednačina je specijalni slučaj problema najmanjih kvadrata, a oba su specijalni slučaj problema matematičkog programiranja.

2.2 Postupci za rešavanje problema optimizacije bez ograničenja

Posmatraćemo problem minimizacije glatke funkcije $f : \mathbb{R}^n \to \mathbb{R}$ bez ograničenja na vrednost argumenta:

$$\min_{x \in \mathbb{R}^n} f(x) . \tag{2.5}$$

Većina postupaka traži lokalni minimum, odnosno stacionarnu tačku koja je rešenje sistema jednačina

$$\nabla f(x) = 0 . \tag{2.6}$$

Za konveksnu funkciju problem traženja lokalnog i globalnog minimuma su ekvivalentni.

Postupci optimizacije bez ograničenja koje ćemo posmatrati nalaze lokalni minimum, obeležavaćemo ga x^*.

Postupke možemo podeliti u zavisnosti od toga koliko izvoda funkcije cilja koriste. Postupke koji koriste $f(x)$, $\nabla f(x)$, $\nabla^2 f(x)$ nazivamo **postupci sa drugim izvodom**; ako koriste $f(x)$, $\nabla f(x)$, to su **gradijentni postupci**; odnosno, ako koriste samo proceduru za računanje $f(x)$, to su **postupci bez izvoda**.

U zavisnosti od toga koliko podataka koriste, postupci zahtevaju manji ili veći stepen glatkosti funkcije cilja.

Postupci koje posmatramo su **iterativni postupci**. To znači da se zadaje početna tačka x_0, a algoritam generiše iterativni niz $\{x_k\}$ koji konvergira ka x^*.

Vrednosti funkcije, gradijenta i Hesijana u tekućoj tački iterativnog postupka ćemo obeležavati redom

$$f_k = f(x_k), \nabla f_k = \nabla f(x_k), \nabla^2 f_k = \nabla^2 f(x_k), k = 1, 2, \ldots \tag{2.7}$$

Obično je niz vrednosti funkcije $\{f_k\}$ koji odgovara generisanom iterativnom nizu **monotono opadajući**. Postoje postupci koji ne zahtevaju da se u svakom iterativnom koraku generiše tačka u kojoj je vrednost funkcije manja, mada se i kod njih posle nekog broja iteracija vrednost funkcije mora smanjiti. To su **nemonotoni postupci**.

Pri generisanju niza $\{x_k\}$ u iterativnom koraku algoritam daje pravilo kako se, koristeći dotad generisane tačke $x_1, x_2, \ldots, x_k$, generiše tačka x_{k+1}. U praksi se pri tome koriste dva glavna pristupa: **linijsko pretraživanje** i **oblast poverenja**.

2.2.1 Postupci oblasti poverenja

U k-tom koraku iterativnog postupka koristeći informacije o funkciji cilja kreira se m_k, model funkcije f u okolini tačke x_k. Zatim se traži pravac p u okolini x_k koji minimizuje model funkciju

$$\min_p m_k(x_k + p), \text{ za } \|p\| < \Delta_k \,. \tag{2.8}$$

Dobijeni pravac daje kandidata za sledeću tačku. Obično se za oblast poverenja u koraku k uzima lopta poluprečnika Δ: $\|p\|_2 \leq \Delta_k$, gde je Δ_k **poluprečnik oblasti poverenja**. Za model funkciju se uzima kvadratna funkcija oblika

$$m_k(x_k + p) = f_k + p^T \nabla f_k + \frac{1}{2} p^T B_k p \,, \tag{2.9}$$

gde su f_k i ∇f_k vrednost funkcije i gradijenta kao u (2.7) i gde je $B_k = \nabla^2 f_k$, Hesijan funkcije f u tački x_k ili njegova aproksimacija.

Aproksimacija funkcije f funkcijom m_k iz (2.9) je dobra u blizini tačke x_k. Zato pokušavamo naći pravac p koje minimizuje model funkciju u regionu poverenja (trust region).

Ako dobijeni pravac ne daje dovoljno smanjenje funkcije cilja, smanjujemo oblast poverenja smanjujući poluprečnik Δ i ponovo tražimo kandidata, inače se za novu iteraciju uzima $x_{k+1} = x_k + p_k$, gde je p_k rešenje problema (2.8).

2.2.2 Postupci linijskog pretraživanja

Glavna razlika između postupaka Oblast poverenja i Linijsko pretraživanje je u redosledu izbora: 1 - pravca od tekuće do sledeće iteracije i 2 - dužine koraka do sledeće iteracije. Neka su iteracije oblika:

$$x_{k+1} = x_k + \alpha_k p_k \,. \tag{2.10}$$

U postupku Oblast poverenja prvo se postavlja maksimalna dužina koraka (poluprečnik oblasti poverenja), a potom se određuje pravac u kome se ostvaruje najveće smanjenje funkcije cilja.

Postupak Linijsko pretraživanje radi obrnuto. Pretpostavljamo da je dat pravac p koji sa gradijentom zaklapa tup ugao, odnosno, za koji važi $\nabla f_k^T p_k < 0$. Kažemo da je to **opadajući pravac**.

Za zadati opadajući pravac p_k traži se dužina koraka α_k koja će dovesti do najvećeg smanjenja funkcije cilja po tom pravcu:

$$\min_{\alpha>0} f(x_k+\alpha p_k)\,. \tag{2.11}$$

Na osnovu Tejlorove teoreme sigurno postoji $\alpha > 0$ za koje je $f_{k+1} < f_k$.

Pri tome nije neophodno naći tačno rešenje problema (2.11), već je moguće da je računski efikasnije naći njegovo približno rešenje. Obično se u jednoj iteraciji generiše nekoliko dužina koraka dok se ne nađe dovoljno dobra aproksimacija minimuma iz (2.11).

Metode linijskog pretraživanja daju konvergenciju za proizvoljan izbor opadajućih pravaca.

Ipak, nekoliko izbora pravaca je više zastupljeno.

2.2.3 Pravac najbržeg opadanja

Najjednostavniji opadajući pravac se može dobiti aproksimacijom Tejlorovog razvoja prvog reda (2.2)

$$f(x_k+\alpha p) = f_k + \alpha p^T \nabla f(x_k+tp)\,,$$

sa $t=0$: $f_{k+1} \approx f_k + \alpha p^T \nabla f_k$. Ako tražimo jedinični pravac, $\|p\|=1$, najveće smanjenje funkcije cilja se dobija kada se reši $\min\limits_p p^T \nabla f_k$.

Kako je $p^T\nabla f_k = \|p\|\,\|\nabla f_k\| \cos\phi$, gde je ϕ ugao između pravca p i gradijenta, vidimo da se najveće smanjenje postiže za

$$\cos\phi = -1\,,\ \text{što se dobija za } p_k = -\frac{1}{\|\nabla f_k\|}\nabla f_k\,.$$

U praksi se nakon određivanja pravca ovom metodom određuje dužina koraka koja skalira pravac, pa se zbog jednostavnosti za opadajući pravac uzima **pravac negativnog gradijenta**

$$p_k := -\nabla f_k\,. \tag{2.12}$$

Postupci Linijskog pretraživanja u kombinaciji sa ovim pravcem se nazivaju **postupci najbržeg opadanja**. Njihova prednost je što u implementaciji nisu potrebni proračuni za izvode drugog reda.

Ipak, ovi postupci su za neke probleme prespori.

2.2.4 Njutnov pravac

Ako u Tejlorovoj teoremi (2.4) stavimo $t = 0$, dobijamo aproksimaciju

$$f(x_k + p) \approx f_k + p^T f_k + \frac{1}{2} p^T \nabla^2 f_k p =: m_k(p) \,. \tag{2.13}$$

Ako pretpostavimo da je $\nabla^2 f_k$ pozitivno definitno, Njutnov pravac se dobija nalaženjem minimuma za $m_k(p)$. Izjednačavanjem izvoda od $m_k(p)$ po p sa nulom, dobija se sistem čije rešenje je

$$p_k^N = -\left(\nabla^2 f_k\right)^{-1} \nabla f_k \,. \tag{2.14}$$

Za dovoljno glatku funkciju aproksimacija Hesijana $\nabla^2 f(x_k + tp) \approx \nabla^2 f_k$ je dobra za male $\|p\|$, što daje brzu konvergenciju u blizini rešenja. Pravac (2.14) nazivamo **Njutnov pravac**.

Za pozitivno definitno $\nabla^2 f_k$ Njutnov pravac je opadajući pravac jer:

$$\nabla f_k^T p_k^N = -p_k^{N^T} \nabla^2 f_k p_k^N \leq -\sigma_k \|p_k^N\|^2 \,, \tag{2.15}$$

za neko $\sigma_k > 0$.

Za Njutnov pravac dužina koraka iteracije (2.10) koji daje najveće smanjenje funkcije je $\alpha_k = 1$. Iterativni postupak sa Njutnovim pravcem i korakom $\alpha_k = 1$ nazivamo **Njutnov iterativni postupak**.

Ako $\nabla^2 f_k$ nije pozitivno definitno, moguće je da Njutnov pravac ne postoji, ili, ako postoji, da nije opadajući pravac.

U Njutnovom postupku se u svakoj iteraciji izračunava i faktoriše Hesijan. To je računski zahtevno, pored toga što zahteva poznavanje formule Hesijana.

Ako se pribegne računanju Hesijana pomoću konačnih razlika, nije potrebno poznavanje formule Hesijana, ali je postupak i dalje računski skup.

2.2.5 Kvazi Njutnovi postupci

Kvazi Njutnovi postupci namesto Hesijana $\nabla^2 f_k$ u formuli (2.14) koriste neku njegovu aproksimaciju B_k.

U kvazi Njutnovim metodama opadajući pravac se dobija po formuli

$$p_k = -B_k^{-1} \nabla f_k \,, \tag{2.16}$$

gde je B_k simetrična, nesingularna matrica koja aproksimira Hesijan. Obično se za matricu B_k u prvoj iteraciji uzima jedinična matrica, a potom se u svakoj iteraciji ažurira koristeći nove vrednosti funkcije i gradijenta.

Osnovna ideja kvazi Njutnovih postupaka je da se uopšti ideja sečice iz postupka za rešavanja jednodimenzionalnih jednačina. Aproksimacija Hesijana se bira tako da zadovoljava **jednačinu sečice**:

$$B_{k+1}s_k = y_k, \text{ gde je } s_k = x_{k+1} - x_k\,,\; y_k = \nabla f_{k+1} - \nabla f_k\,. \tag{2.17}$$

Matrica B_k treba da je simetrična (isto kao Hesijan).

Za aproksimaciju Hesijana često se koristi **formula SR1, odnosno Symmetric-Rank-one formula**:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}\,. \tag{2.18}$$

Ova formula daje simetrične matrice koje zadovoljavaju jednačinu sečice. Razlika između dve susedne iteracije je matrica ranga 1.

Još popularnija je BFGS formula koju su izumeli Broyden, Fletcher, Goldfarb i Shanno:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k^T}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}\,. \tag{2.19}$$

Ova formula takođe daje simetrične matrice koje zadovoljavaju jednačinu sečice. Razlika između dve susedne iteracije je matrica ranga 2.

Praktična implementacija BFGS postupka umesto aproksimacije B_k koristi aproksimaciju inverznog Hesijana $H_k = B_k^{-1}$.

$$H_{k+1} = (I - \rho_k s_k y_k^T) H_k (I - \rho_k y_k s_k^T) + \rho_k s_k s_k^T\,,\; \text{gde je } \rho_k = \frac{1}{y_k^T s_k}\,. \tag{2.20}$$

Koristeći ovu aproksimaciju inverznog Hesijana pravci iterativnog koraka su

$$p_k = -H_k \nabla f_k\,.$$

2.2.6 Dužina koraka linijskog pretraživanja

Posle određivanja pravca, u svakoj iteraciji (2.10) se linijskim pretraživanjem određuje dužina koraka.

Pravac najbržeg opadanja i Njutnov pravac su specijalni slučajevi formule (2.16), sa $B_k = I$ i $B_k = \nabla^2 f_k$ redom.

Pri određivanju dužine koraka se polazi od koraka $\alpha_k = 1$ koji bi odgovarao Njutnovom pravcu. Ako taj korak ne zadovoljava uslove koji garantuju konvergenciju, korak se skraćuje.

Za dati opadajući pravac, idealno je da se dužina koraka α_k odredi kao globalni minimum funkcije cilja duž tog pravca

$$\phi(\alpha) = f(x_k + \alpha p_k)\,. \tag{2.21}$$

Određivanje tog minimuma ćemo zvati **tačno linijsko pretraživanje**. Tačno linijsko pretraživanje je računski skupo.

Zato se primenjuje **netačno linijsko pretraživanje** koje određuje dužinu koraka sa zadovoljavajućim smanjenjem funkcije (2.21) u odnosu na broj računanja vrednosti funkcije i gradijenta.

Da bi se obezbedila konvergencija postupka potrebno je postaviti uslove na dužinu koraka. Armijo uslov zahteva da dužina koraka α_k pruži **dovoljno umanjenje** funkcije cilja:

$$f(x_k + \alpha p_k) \le f(x_k) + c_1 \alpha \nabla f_k^T p_k\,, \text{ za neko } c_1 \in (0,1)\,. \tag{2.22a}$$

Odnosno: dopuštene dužine koraka α su one za koje je $\phi(\alpha) \le l(\alpha)$, gde je $l(\alpha)$, desna strana nejednakosti (2.22a). To su na slici 2.1 vrednosti za koje je grafik funkcije $\phi(\alpha)$ ispod prave $l(\alpha)$. U praksi se c_1 uzima malo, recimo $c_1 = 10^{-4}$.

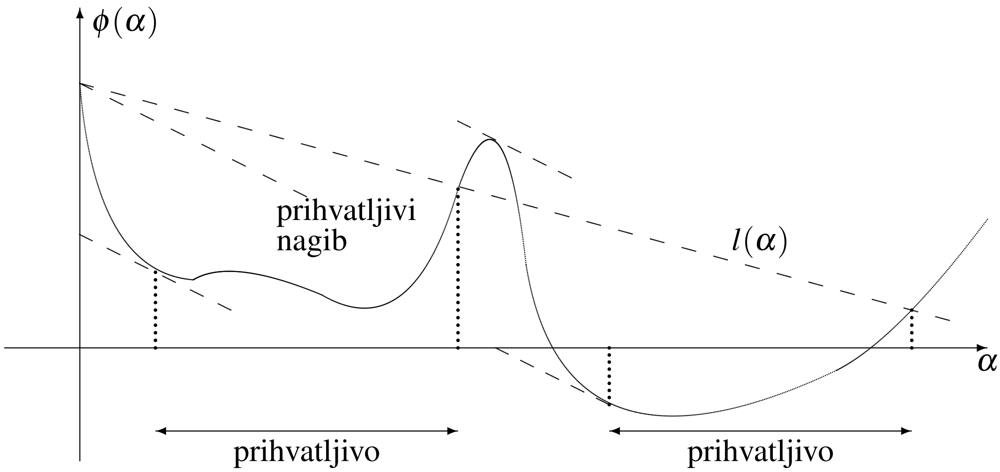
Da se ne bi za dužinu koraka uzimale jako male vrednosti koje ne dovode do napretka u iteracijama, postavlja se **uslov nagiba**

$$\nabla f(x_k + \alpha p_k)^T p_k \ge c_2 \alpha \nabla f_k^T p_k\,, \text{ za neko } c_2 \in (c_1, 1)\,. \tag{2.22b}$$

Leva strana nejednakosti (2.22b) je izvod $\phi'(\alpha)$, treba da je veća od umnoška $\phi'(0)$. To su na slici 2.1 apscise tačaka krive u kojima nagib nije strmiji od prihvatljivog nagiba. U ovom uslovu se traži da je $c_2 > c_1$, zato da bi uslov (2.22b) mogao biti ispunjen istovremeno sa uslovom (2.22a).

Zajedno se uslovi (2.22a), (2.22b) nazivaju **Volfovi (Wolfe) uslovi**.

U praktičnoj primeni, obično se uzima $c_2 = 0.1$ za gradijentni korak ili korak konjugovanog gradijenta; odnosno $c_2 = 0.9$ za Njutnov ili kvazi Njutnove korake.



Slika 2.1: Slika oblasti za α koje zadovoljavaju Wolfe uslove

LEMA 1 *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ neprekidno diferencijabilna. Neka je p_k opadajući pravac u x_k i neka je f ograničena odole na skupu $\{x_k + \alpha p_k | \alpha > 0\}$. Ako je $0 < c_1 < c_2 < 1$, onda postoji interval vrednosti α koje zadovoljavaju Volfove uslove* (2.22a), (2.22b).

Za dokaz konvergencije metode linijskog pretraživanja koristi se teorema Zoutendijka [69]:

TEOREMA 24 (Zoutendijk) *Neka je iterativni niz $\{x_k\}$ definisan iteracijom* (2.10) *sa opadajućim pravcima p_k i neka α_k zadovoljava Volfove uslove* (2.22a), (2.22b).

Neka je f ograničena odole na $\mathbb{R}^n$ i neprekidno diferencijabilna na otvorenom skupu $\mathscr{N}$ koji sadrži skup $\mathscr{L} = \{x : f(x) \le f(x_0)\}$, gde je x_0 data početna iteracija.

Neka je, dalje, gradijent ∇f Lipšic neprekidan nad $\mathscr{N}$, odnosno:

$$\forall x, \tilde{x} \in \mathscr{N}, \ \|\nabla f(x) - \nabla f(\tilde{x})\| \le L\|x - \tilde{x}\|, \ \text{za neko } L > 0. \tag{2.23}$$

Onda

$$\sum_{k \ge 0} \cos^2 \theta_k \, \|\nabla f(x_k)\|^2 < \infty, \tag{2.24}$$

gde je θ_k ugao koji zaklapaju p_k i $-\nabla f_k$. Njegov kosinus se dobija po formuli

$$\cos\theta_k = \frac{-\nabla f_k^T p_k}{\|\nabla f_k\| \, \|p_k\|} \,. \tag{2.25}$$

Pored uslova (2.22a), (2.22b), postoji verzija ove teoreme sa drugim skupom uslova ([47]):

Na primer, **jaki Volfovi uslovi**:

$$f(x_k + \alpha_k p_k) \le f_k + c_1 \alpha_k \nabla f_k^T p_k \tag{2.26a}$$
$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \le c_2 |\nabla f_k^T p_k| \,, \tag{2.26b}$$

za neke $0 < c_1 < c_2 < 1$.

Ili, na primer, **Goldštajnovi (Goldstein) uslovi**:

$$f_k + (1-c)\alpha_k \nabla f_k^T p_k \le f(x_k + \alpha_k p_k) \le f_k + c\alpha_k \nabla f_k^T p_k \,, \tag{2.27}$$

za neko $c \in (0, \frac{1}{2})$.

Ista posledica: (2.24) koju nazivamo **Zoutendijkov uslov** sledi iz para uslova (2.26a), (2.26b), odnosno iz uslova (2.27).

Prosta posledica Zoutendijkovog uslova je

$$\lim_{k\to\infty} \cos^2\theta_k \, \|\nabla f_k\|^2 = 0 \,. \tag{2.28}$$

Da bi se obezbedila konvergencija niza $\{x_k\}$ ka stacionarnoj tački, potrebno je postaviti još uslov na pravce p_k. Niz uglova između pravaca p_k i gradijenata ∇f_k ne sme se nagomilavati kod pravog ugla:

$$\forall k \in \{0,1,2,\ldots\},\ \cos\theta_k \ge \delta \,,\ \text{ za neko } \ \delta > 0. \tag{2.29}$$

Direktna posledica Zoutendijkovog uslova i uslova (2.29) je:

$$\lim_{k\to\infty} \|\nabla f(x_k)\| = 0 \,. \tag{2.30}$$

Da podsetimo: globalna konvergencija, odnosno uslov (2.30), znači da postupak konvergira ka x^*, stacionarnoj tački funkcije f. Tek ako se na f postave dodatni uslovi sledi da je x^* lokalni, odnosno globalni minimum.

Za postupke najbržeg opadanja (pravci negativnog gradijenta), važi $\cos\theta_k = 1$. Za te postupke je dovoljno linijskim pretraživanjem uzimati dužine koraka koje zadovoljavaju Volfove uslove (ili Jake Volfove uslove ili Goldštajnove uslove) i važiće globalna konvergencija.

Za kvazi Njutnove postupke, iz pozitivne definitnosti aproksimacije Hesijana B_k sledi da su im pravci opadajući. Za Njutnove pravce (koji su u neku ruku specijalni slučaj kvazi Njutnovih pravaca) to imamo izvedeno u (2.15).

Dodajmo pretpostavku da matrice B_k imaju **uniformno ograničen uslovni broj**. To znači da:

$$\forall k \in \{0,1,2,\ldots\},\ \|B_k\|\,\|B_k^{k-1}\| \le M\,,\ \text{ za neko } M > 0. \tag{2.31}$$

Lako se može pokazati iz (2.25) da iz (2.31) sledi $\cos\theta_k \ge \dfrac{1}{M}$.

Dakle: U iterativnom nizu (2.10), Njutnov ili kvazi Njutnov postupak sa pravcima (2.16), uz uslov (2.31) čije dužine koraka daju Zoutendijkov uslov jesu globalno konvergentni.

Teorema 24 daje i više: Ako neki algoritam generiše iterativni niz (2.10) koji u svakoj iteraciji daje smanjenje funkcije cilja i svakih m koraka uzima pravac najbržeg opadanja $p_k = -\nabla f_k$ sa dužinom koraka koja zadovoljava Volfove ili Goldštajnove uslove, može se dokazati

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0\,. \tag{2.32}$$

2.2.7 Red konvergencije

Da bi se za pravce zadovoljio uslov (2.29), u tekućoj iteraciji se može "ispraviti" pravac za koji je $\cos\theta_k < \delta$, za neko unapred određeno $\delta > 0$. Takvom modifikacijom algoritma se može dobiti globalna konvergencija.

Ipak, to utiče na red konvergencije. Za probleme sa loše uslovljenim Hesijanom, baš pravci blizu ortogonalnih na pravac negativnog gradijenta daju brzu konvergenciju, a oni bi mogli biti odbačeni nekim izborom δ. Za takve probleme postupak najbržeg opadanja sporo konvergira.

Sa druge strane, za neke probleme, Njutnovi pravci koji u blizini rešenja daju brzu konvergenciju, u tačkama koje su dalje od rešenja možda neće biti ni opadajući.

Bilo bi dobro definisati algoritam koji će zadovoljiti oba aspekta: globalnu konvergenciju i brzu konvergenciju blizu rešenja.

DEFINICIJA 18 *Neka je* $\{x_k\}$ *niz u* $\mathbb{R}^n$ *koji konvergira ka* x^*.

Kažemo da je konvergencija ***Q-linearna*** *ako postoji* $r \in (0,1)$ *tako da*

$$\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|} \le r\,,$$

za sve dovoljno velike k.

Kažemo da je konvergencija ***Q-superlinearna*** *ako*

$$\lim_{k\to\infty} \frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|} = 0\,.$$

Kažemo da je konvergencija ***Q-kvadratna*** *ako postoji* $M > 0$ *tako da*

$$\frac{\|x_{k+1}-x^*\|}{\|x_k-x^*\|^2} \le M\,,$$

za sve dovoljno velike k.

Brzina konvergencije zavisi od r i M. Ipak, za sve r i M Q-kvadratna konvergencija je bolja od Q-linearne.

Ako niz konvergira Q-kvadratno, onda konvergira Q-superlinearno. Ako niz konvergira Q-superlinearno onda konvergira i Q-linearno.

Obično Kvazi Njutnovi postupci konvergiraju Q-superlinearno, Njutnovi konvergiraju Q-kvadratno. Postupci najbržeg opadanja konvergiraju Q-linearno, a ako je postupak loše uslovljen, konstanta r je blizu 1.

U daljem tekstu ćemo izostaviti prefiks Q i samo reći linearno, superlinearno, kvadratno.

Red konvergencije postupka najbržeg opadanja

Imamo sledeću teoremu.

TEOREMA 25 *Neka je* $f : \mathbb{R}^n \to \mathbb{R}$ *dva puta neprekidno diferencijabilna i neka je* $\{x_k\}$ *iterativni niz metode opadajućih pravaca sa dužinama koraka određenim tačnim linijskim pretraživanjem.*

Neka $\{x_k\}$ konvergira ka x^* i neka je Hesijan $\nabla^2 f(x^*)$ pozitivno definitan.

Neka $r \in \left(\dfrac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}, 1\right)$, gde su $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ karakteristične vrednosti Hesijana $\nabla^2 f(x^*)$.

Onda za sve dovoljno velike k važi $f(x_{k+1}) - f(x^*) \leq r^2[f(x_k) - f(x^*)]$.

Vidimo da je konvergencija linearna. Ipak, u nekim slučajevima postupak najbržeg opadanja može biti jako spor. To su slučajevi kada je uslovni broj matrice Hesijana u rešenju velik, posledično, tada je i vrednost izraza $\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1}$ blizu jedinice.

Red konvergencije Njutnovog postupka

Posmatramo Njutnov iterativni postupak: (2.10) sa izborom pravaca (2.14).

Poznato je da ako je Hesijan u rešenju $\nabla^2 f(x^*)$ pozitivno definitan, onda postoji okolina rešenja nad kojom je Hesijan pozitivno definitan. Unutar te okoline ćemo posmatrati konvergenciju postupka.

TEOREMA 26 *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ dva puta diferencijabilna i neka je Hesijan $\nabla^2 f(x)$ Lipšic neprekidan u nekoj okolini rešenja u kojoj su zadovoljeni uslovi teoreme 22.*

Neka je $\{x_k\}$ iterativni niz (2.10) *sa Njutnovim pravcima* (2.14) *i dužinom koraka $\alpha_k = 1$. Tada važi:*

(i) Ako je početna tačka x_0 dovoljno blizu x^* onda $\{x_k\}$ konvergira ka x^*.

(ii) Brzina konvergencije je kvadratna.

(iii) Niz normi gradijenata $\{\|\nabla f_k\|\}$ kvadratno konvergira ka nuli.

U okolini rešenja Volfovi uslovi ili Goldštajnovi uslovi će počevši od nekog k dopuštati izbor dužine koraka $\alpha_k = 1$. U praktičnoj implementaciji se koristi ta činjenica. Njutnov postupak se implementira sa linijskim pretraživanjem koje prvo proverava da li korak dužine $\alpha_k = 1$ zadovoljava uslove, i počevši od nekog k postavlja $\alpha_k = 1$.

Red konvergencije kvazi Njutnovih postupaka

Daćemo prvo opštiju teoremu, koja važi za opadajuće pravce sa dodatnim uslovom, a potom ćemo dati teoremu o potrebnom i dovoljnom uslovu za superlinearnu konvergenciju kvazi Njutnovih pravaca.

TEOREMA 27 *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ dva puta neprekidno diferencijabilna.*

Neka je $\{x_k\}$ iterativni niz (2.10), *gde su p_k opadajući pravci, i α_k zadovoljavaju* (2.22a), (2.22b), *sa konstantom $c_1 \geq 1/2$.*

Neka $\{x_k\}$ konvergira ka tački x^* za koju je $\nabla f(x^*) = 0$ i $\nabla^2 f(x^*)$ pozitivno definitno.

Neka pravci p_k zadovoljavaju uslov

$$\lim_{k\to\infty} \frac{\|\nabla f_k + \nabla^2 f_k p_k\|}{\|p_k\|} = 0 . \tag{2.33}$$

Onda

(i) Počev od nekog k_0 dužina koraka $\alpha_k = 1$ je dopustiva.

(ii) Ako je za $k > k_0$, $\alpha_k = 1$ onda $\{x_k\}$ konvergira ka x^* superlinearno.

Ako su p_k kvazi Njutnovi pravci definisani sa (2.16), onda je uslov (2.33) ekvivalentan uslovu

$$\lim_{k\to\infty} \frac{\|\left(B_k - \nabla^2 f(x^*)\right) p_k\|}{\|p_k\|} = 0 . \tag{2.34}$$

Iz toga vidimo da je dovoljan uslov za superlinearnu konvergenciju da je B_k dovoljno dobra aproksimacija Hesijana $\nabla^2 f(x^*)$ samo duž pravca p_k.

Štaviše, uslov (2.34) je i potreban i dovoljan za superlinearnu konvergenciju:

TEOREMA 28 *Neka je $f : \mathbb{R}^n \to \mathbb{R}$ dva puta neprekidno diferencijabilna.*

Neka je iterativni niz definisan sa (2.10), *i neka su p_k kvazi Njutnovi pravci, definisani sa* (2.16).

Neka niz $\{x_k\}$ konvergira ka x^* i neka $\nabla f(x^*) = 0$ i neka je $\nabla^2 f(x^*)$ pozitivno definitno.

Onda $\{x_k\}$ konvergira superlinearno ako i samo ako važi (2.34).

2.2.8 Bektreking linijsko pretraživanje

Kada su pravci linijskog pretraživanja Njutnovi ili kvazi Njutnovi, u želji da postupak ima kvadratnu brzinu konvergencije Njutnovog postupka, treba i korake α_k birati bliže Njutnovom koraku $\alpha_k = 1$. Zato u Linijskom pretraživanju krećemo sa korakom $\bar{\alpha} = 1$.

Jedan praktičan algoritam za određivanje dužine koraka koji može zadovoljiti prvi Volfov uslov (2.22a), a da koraci istovremeno ne budu previše kratki, zove se **bektreking linijsko pretraživanje**, algoritam Backtracking Line Search.

Algoritam 1 Backtracking Line Search

Input: $\bar{\alpha} > 0, 0 < \rho_{lo} < \rho_{hi} < 1, c \in (0, \frac{1}{2})$
Output: α koje zadovoljava (2.22a)
  $\alpha := \bar{\alpha}$;
  while $f(x_k + \alpha p_k) > f(x_k) + c\alpha \nabla f_k^T p_k$ **do**
    choose $\rho \in [\rho_{lo}, \rho_{hi}]$;
    $\alpha := \rho\alpha$;
  end while
  return

Ova jednostavna procedura za nalaženje dužine koraka α_k duž opadajućeg pravca p_k se poziva sa početnom vrednošću koraka $\bar{\alpha} = 1$ za Njutnovu i kvazi Njutnove metode. Može se odrediti i manja početna vrednost za drugi izbor opadajućih pravaca. Kad se procedura bektreking linijskog pretraživanja zaustavi sa $\alpha_k = \alpha$, dužina koraka će biti takva da korak α_k/ρ ne zadovoljava uslov (2.22a), a korak α_k zadovoljava.

Vrednost koeficijenta c se bira tako da i malo smanjenje funkcije cilja može da zaustavi petlju traženja koraka α, recimo, $c = 10^{-4}$.

Faktor kontrakcije ρ može da se menja tokom iteracija unutar linijskog pretraživanja. Jedino treba da se sačuva $\rho \in [\rho_{lo}, \rho_{hi}]$, za unapred odabrane konstante $0 < \rho_{lo} < \rho_{hi} < 1$.

2.2.9 Algoritmi linijskog pretraživanja

Linijsko pretraživanje je postupak nalaženja minimuma jednodimenzionalne funkcije (2.21) uz zadovoljenje uslova (2.22a), (2.22b) ili nekih drugih uslova koji garantuju konvergenciju.

Za konveksnu kvadratnu funkciju $f(x) = \frac{1}{2}x^T Qx - b^T x$ tačan minimum duž pravca $\{x_k + \alpha p_k, \alpha > 0\}$ imamo dat formulom

$$\alpha_k = \frac{\nabla f_t^T p_k}{p_k^T Q p_k} . \tag{2.35}$$

Uglavnom se minimum ne može egzaktno odrediti, a iterativno nalaženje minimuma je računski skupo. Ipak, dovoljna je i aproksimacija minimuma koja zadovoljava uslove date u glavi 2.2.6 da bi postupak konvergirao.

Da bismo dobili dužine koraka koje daju brzu konvergenciju izmenićemo algoritam Backtracking Line Search. Izbor koeficijenta smanjenja koraka ρ u iteracijama (linija choose $\rho \in [\rho_{lo}, \rho_{hi}]$) se vrši minimizacijom modela funkcije (2.21) kvadratnim ili kubnim polinomom. Algoritam Line Search u nastavku to radi.

Ova procedura vraća zadovoljenost jednog od tri izlazna kriterijuma kao vrednost $retcode$:

$retcode = 0$, ako α koje vraća zadovoljava uslov (2.22a),

$retcode = 1$, ako postupak dobija za λ vrednosti manje od $\lambda_{\min}$,

$retcode = 2$, ako je potrošena kvota računanje funkcije f,

$retcode = 3$, negativna diskriminanta za minimum kubne interpolacije.

U slučajevima $retcode = 1$, $retcode = 2$, $retcode = 3$ vrednost iteracije x_1 i vrednost funkcije f_1 koje Line Search vraća su nepouzdane, a za $retcode = 0$ se vraćaju da se ne bi ponovo računale u iterativnom postupku.

Ova procedura kreće od vrednosti $\alpha = 1$ koju smanjuje korakom između 0.1 i 0.5, sve dok se ne zadovolji jedan od izlaznih kriterijuma. Korak smanjenja se nalazi kao minimum modela funkcije (2.21) kubnim, a u prvom koraku kvadratnim polinomom.

Nekad iterativni proces blizu rešenja bude usporen zbog malih dužina koraka koje daje ovaj algoritam. Štaviše, u blizini rešenja greška zaokruživanja može da dovede do zaglavljivanja iterativnog procesa (*stall*). Zato se kao ulazni parametar ove procedure daje $\alpha_{\min}$. U slučaju kad algoritam 2 smanji korak α ispod $\alpha_{\min}$ iterativni proces je verovatno došao do rešenja, treba da stane, ali treba i proveriti da li je rešenje traženi minimum. Tada se vraća poruka $retcode = 1$.

Korenovanje negativnog broja može se pojaviti u proceduri Line search kada na vrednost funkcije f utiče šum ili kad se pojavi greška zaokruživanja u determinističkom slučaju. Mi ćemo, kada se pojavi korenovanje negativnog broja, vraćati $retcode = 3$. Obrada ovog povratnog koda će biti ista kao kada iterativni proces zaglavljuje zbog grešaka zaokruživanja blizu rešenja.

Brojanje računanja funkcije se vodi u lokalnoj promenljivi $fcalcls$. Ovu mogućnost nude mnoge implementacije linijskog pretraživanja, na primer u Matlabu. Vrednost $fcalcls$ je jedan od izlaznih parametara ovog algoritma, da bi glavni iterativni program mogao da vodi računa o ukupnoj kvoti računanja funkcije cilja.

Algoritam 2 Line Search

Input: $x_0 \in \mathbb{R}^n,\ f_0 \in \mathbb{R}(= \hat{f}(x_0)),\ f(\cdot) : \mathbb{R}^n \to \mathbb{R},\ g_0 \in \mathbb{R}^n(= \hat{g}(x_0)),\ p \in \mathbb{R}^n,$ $\lambda_{\min} \in \mathbb{R}^+$, $maxfcalcls \in \mathbb{Z}^+$

Output: x_1, f_1, $retcode \in \mathbb{Z}^+$, $\lambda \in \mathbb{R}^+$, $fcalcls \in \mathbb{Z}^+$

  $c := 10^{-4}$;
  $initslope := g_0^T\, p$;
  $retcode := 1$;
  $\alpha := 1$;
  $fcalc := 0$;
  while $retcode > 0$ **do**
    $x_1 := x_0 + \alpha\, p$;
    $f_1 := f(x_1)$; $flalc := fcalc + 1$;
    if $f_1 \le f_0 + c\,\alpha\, initslope$ **then** ▷ Nađeno α
      $retcode := 0$;
    else if $\alpha < \alpha_{\min}$ **then** ▷ Premalo α zaglavljuje
      $x_1 := x_0$; $f_1 := f_0$; $retcode := 1$;
    else if $fcalc \ge maxfcalc$ **then** ▷ maksimalni broj računanja f
      $x_1 := x_0$; $f_1 := f_0$; $retcode := 2$;
    else ▷ Smanjujemo α
      if $\alpha = 1$ **then** ▷ Prvi put kvadratni model
        $\alpha_{temp} := -\frac{initslope}{2(f_1 - f_0 - initslope)}$;
      else ▷ Inače kubni model

$$\begin{bmatrix} A \\ B \end{bmatrix} := \frac{1}{\alpha - \alpha_{prev}} \begin{bmatrix} \frac{1}{\alpha^2} & \frac{-1}{\alpha_{prev}^2} \\ \frac{-\alpha_{prev}}{\alpha^2} & \frac{\alpha}{\alpha_{prev}^2} \end{bmatrix} \begin{bmatrix} f_1 - f_0 - \alpha\, initslope \\ f_{prev} - f_0 - \alpha_{prev}\, initslope \end{bmatrix};$$

Algoritam 2 Line Search (nastavak)

```
            disc := B^2 − 3 A initslope;
            if disc < 0 then                              ▷ Šum ili zaokruživanje
                x_1 := x_0; f_1 := f_0; retcode := 3;
            end if
            if A = 0 then                                 ▷ Treba pravi kubni model
                α_temp := − initslope / (2B);
            else
                α_temp := (−b + √disc) / (3A);
            end if
            if α_temp > 0.5α then                         ▷ Ne želimo premalo smanjenje
                α_temp := 0.5α;
            end if
        end if
        α_prev := α;
        f_prev := f_1;
        if α_temp ≤ 0.1α then                             ▷ Ni preveliko smanjenje
            α := 0.1α;
        else
            α := α_temp;
        end if
    end if
end while
return
```

2.3 Njutnov postupak za rešavanje sistema nelinearnih jednačina

Posmatramo problem

$$g(x) = 0, \tag{2.36}$$

gde je $g : \mathbb{R}^n \to \mathbb{R}^n$ vektorska funkcija,

$$g(x) = \left[g_1(x), g_2(x), \dots, g_n(x)\right]^T,$$

i gde su g_k, $k = 1, 2, \ldots, n$ glatke funkcije.

Vektor x^* koji zadovoljava (2.36) zove se **rešenje** ili koren.

U glavi 2.1 je data veza sistema linearnih jednačina sa problemom matematičkog programiranja.

Daćemo Njutnov algoritam za sistem linearnih jednačina.

Neka je $g'(x)$ Jakobijan vektorske funkcije $g(x)$.

Algoritam 3 Njutnov algoritam za sistem nelinearnih jednačina

---

Korak 1 Odaberi x^0.

Korak 2 For $k = 0, 1, 2, \ldots$

Korak 3 Reši s^k iz sistema linearnih jednačina $g'(x^k)\, s^k = -g(x^k)$.

Korak 4 Postavi $x^{k+1} = x^k + s^k$.

Korak 5 End For

Da smo primenili algoritam 3 na gradijent $g = \nabla f$ funkcije f iz problema (2.1), dobili bismo Njutnov metod iz glave 2.2.4. Pri tome je $g' = \nabla^2 f$.

Daćemo bez dokaza teoremu o konvergenciji Njutnovog postupka.

TEOREMA 29 *Neka je g neprekidno diferencijabilna u konveksnom otvorenom skupu $\mathscr{D} \subseteq \mathbb{R}^n$. Neka je $x^* \in \mathscr{D}$ nedegenerisano rešenje* (2.36) *i neka je $\{x^k\}$ niz generisan algoritmom 3.*

Ako je $x^k \in \mathscr{D}$ dovoljno blizu x^*, onda $\{x^k\}$ superlinearno konvergira ka x^*.

Ako je g Lipšic neprekidno diferencijabilna u okolini x^*, onda je konvergencija kvadratna.

Deo II

Kvazi Njutnove metode stohastičke optimizacije

Glava 3

Pregled rezultata stohastičke optimizacije

Mi ćemo pretpostaviti da na funkciju koju želimo minimizovati $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ utiče slučajni šum $\xi(x)$ koji predstavlja slučajnu promenljivu nad prostorom verovatnoće $(\Omega, \mathscr{F}, P)$. Tako da raspolažemo samo vrednostima

$$F(x) = f(x) + \xi(x)\,. \tag{3.1}$$

Za n-dimenzionalnu funkciju $g(x) : \mathbb{R}^n \to \mathbb{R}^n$ čiju nulu tražimo takođe ćemo pretpostaviti da raspolažemo samo vrednostima na koje utiče šum $\varepsilon(x)$. Šum je slučajni vektor nad prostorom verovatnoće $(\Omega, \mathscr{F}, P)$,

$$G(x) = g(x) + \varepsilon(x)\,. \tag{3.2}$$

Slučajna promenljiva $\xi(x)$ i slučajni vektor $\varepsilon(x)$ zavise od vrednosti $x \in \mathbb{R}^n$.

Slučajni šum mora zadovoljavati neke osobine, na pr. mora biti centriran oko nule. Obeležavamo

$$\xi_k(x) = \xi(x, \omega_k),\ \varepsilon_k(x) = \varepsilon(x, \omega_k). \tag{3.3}$$

Osnovni problem koji ćemo posmatrati je nalaženje vrednosti x^* za koju se ostvaruje

$$\min_{x \in \mathbb{R}^n} f(x)\,, \tag{3.4}$$

za funkciju $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$.

Posmatraćemo probleme bez ograničenja, tako da možemo pretpostaviti da je f diferencijabilna i da je $D = \mathbb{R}^n$.

3.1 Stohastička aproksimacija (SA)

Hronološki, prvi algoritam stohastičke optimizacije je Robbins-Monro algoritam stohastičke aproksimacije (SA) iz 1951. godine, [49]. Ovaj algoritam služi za nalaženje rešenja x^* sistema n jednačina sa n nepoznatih

$$g(x) = 0, \tag{3.5}$$

za $g : \mathbb{R}^n \to \mathbb{R}^n$ kad se raspolaže samo vrednostima g na koje utiče šum.

Problem (3.5) se pojavljuje pri minimizaciji funkcije $f : \mathbb{R}^n \to \mathbb{R}$ kada je $g(x) = \nabla f(x)$.

Stohastička aproksimacija je stohastičko uopštenje determinističkog problema (3.5). Pretpostavlja se da je funkcija g očekivanje slučajne veličine čija raspodela nam nije poznata. Mi smo u mogućnosti samo da za konkretnu vrednost x dobijemo realizovanu vrednost posmatrane slučajne veličine.

U pionirskom radu u ovoj oblasti [49], Robbins i Monro su dokazali L^2 konvergenciju iterativnog niza za nalaženje korena monotone funkcije u jednoj dimenziji. Iz L^2 konvergencije sledi konvergencija u verovatnoći. Wolfowitz u radu [63] iz 1952. daje oslabljene uslove za konvergenciju iterativnog niza u srednjem.

Skoro sigurnu konvergenciju iterativnog niza je dokazao Blum 1954. u [6], koristeći rezultate Loève iz rada [43] iz 1951. Takođe, 1954. Blum u [7] dokazuje skoro sigurnu konvergenciju iterativnog niza slučajnih vektora ka rešenju sistema n jednačina sa n nepoznatih (3.5).

Kiefer i Wolfowitz u [35], 1952. rešavaju problem nalaženja maksimuma unimodalne funkcije u jednoj dimenziji postupkom koji potiče iz SA.

Algoritam SA polazi od unapred zadate početne vrednosti x_0 i niza dužina koraka $\{a_k\}, k = 0, 1, \dots$ koji zadovoljava

$$a_k > 0, k \in \mathbb{N}, \ \sum_{k=0}^{\infty} a_k = \infty, \ \sum_{k=0}^{\infty} a_k^2 < \infty. \tag{3.6}$$

Za nizove koji zadovoljavaju (3.6) kažemo da su **tipa** $\frac{1}{k}$, a najjednostavniji primer je $a_k = \frac{a}{k+1}$, za $a > 0$ i $k = 0, 1, \dots$.

U SA postupku za proizvoljno izabrano x_0 kreira se stohastički iterativni niz rekurzivnom formulom

$$x_{k+1} = x_k - a_k G_k, \tag{3.7}$$

gde je G_k slučajna promenljiva sa raspodelom koja zavisi od vrednosti x_k, pišemo $G_k = G(x_k)$, odnosno, u skladu sa oznakama (3.2), (3.3) imamo

$$G_k = g(x_k) + \varepsilon_k(x_k), \ k = 0, 1, 2, \dots$$

sa realizovanom vrednošću šuma $\varepsilon_k(x_k)$.

Umesto dobijanja bolje procene za $g(x_k)$ traženjem aritmetičke sredine realizovanih vrednosti $G(x_k)$, Robbins-Monro algoritam vrši *usrednjavanje kroz iteracije*. Iako se kroz iteracije menja vrednost x_k, tako da ovo nije pravo usrednjavanje, ovaj postupak se pokazao efikasnim, jer za nalaženje rešenja ne troši puno računanja g.

Primena SA algoritma je vrlo široka. Problemi nalaženja nepoznatih parametara modeliranog procesa u kojem se raspolaže samo merenjima na koja je uticao šum se lako pretvaraju u probleme nalaženja rešenja sistema jednačina.

Sedamdesetih godina prošlog veka, sa ciljem oslabljenja dovoljnih uslova za konvergenciju SA postupka prišlo se problemu konvergencije na drugi način, preko običnih diferencijalnih jednačina. Na taj način je oslabljen uslov brzine rasta A4, ali je uvedena pretpostavka o ograničenosti niza $\{x_k\}$.

Postoje i drugi pristupi dokazivanju konvergencije SA postupka: Chen [13], Spall [55].

Mogućnost primene, te pouzdanost i stabilnost nekog algoritma na što široj klasi problema zovemo **robustnost**. Neki algoritmi se posebno konstruišu da bi na određenom tipu problema bili efikasniji. Česta je trgovina između robustnosti i efikasnosti u izborima algoritama za rešavanje nekih klasa problema.

Naime, algoritmi koji su efikasni za jednu klasu problema manje su efikasni za neku drugu klasu. To je posledica takozvane "No Free Lunch" teoreme ("Nema besplatnog ručka"), Wolpert & Macready, [64].

Ipak, za određenu klasu problema neki algoritmi su efikasniji od drugih.

3.1.1 Konvergencija SA postupka

Razni autori (vidi [13], [55]) su dokazali konvergenciju iterativnog niza (3.7) za razne kombinacije dovoljnih uslova.

Mi ćemo ovde dati teoremu iz [13].

Prvo ćemo dati standardan skup uslova pod kojima postupak (3.7) konvergira.

Neka je za zadato x_0 niz $\{x_k\}$ nastao postupkom (3.7). Označavaćemo sa $\mathscr{F}_k$ σ-algebru generisanu realizovanim vrednostima $x_0, x_1, \ldots, x_k$.

A1 Niz dužina koraka $\{a_k\}$ zadovoljava (3.6).

A2 Postoji neprekidno diferencijabilna funkcija Ljapunova $v(\cdot):\mathbb{R}^n \to \mathbb{R}$ koja zadovoljava uslove:

(i) drugi izvod v je ograničen,

(ii) $v(x) > 0$ za $x \neq x_0$, $v(x_0) = 0$ i $v(x) \to \infty$ kada $\|x\| \to \infty$.

(iii) Za svako $\varepsilon > 0$ postoji $\beta_\varepsilon > 0$ takvo da

$$\inf_{\|x-x_0\|>\varepsilon} \nabla v^T g(x) = \beta_\varepsilon > 0\,, \tag{3.8}$$

gde je ∇v oznaka za gradijent $v(\cdot)$.

A3 Niz šumova sa generisanim sigma algebrama $(\varepsilon_k, \mathscr{F}_k)$ je niz martingal razlika za neopadajući niz sigma algebri $\mathscr{F}_k$ takav da važi

$$E(\varepsilon_k|\mathscr{F}_{k-1}) = 0,\ E(\|\varepsilon_k\|^2) < \infty.$$

A4 Funkcija g i uslovni drugi momenat realizovanog šuma su ograničeni odgore:

$$\|g(x)\|^2 + E(\|\varepsilon_k\|^2|\mathscr{F}_{k-1}) < c(1+v(x)), k = 0,1,\ldots,$$

za neko $c > 0$.

Uslov A1 se postavlja da bi dužina koraka a_k prigušivala šum, ali ne previše. Trgovina između prigušenja šuma smanjivanjem koraka i poboljšanja rešenja povećavanjem koraka je ključna za brzinu konvergencije SA algoritama. Poznavanje raspodele u pojedinom koraku bi omogućilo ocenu greške i nalaženje optimalne dužine koraka a_k.

Uslov A2 je standardan uslov na oblik funkcije g koji se postavlja kad se dokazuje konvergencija.

Uslov A3 koji zahteva da je srednja vrednost šuma nula je strog. Ako se radi sa šumom koji osim slučajne komponente sadrži i strukturnu grešku, potrebno je oslabljenje ovog uslova.

Za funkciju $v(\cdot)$ se može uzeti funkcija

$$v(x) = \|x - x_0\|^2. \tag{3.9}$$

U tom slučaju iz uslova A4 sledi da rast $g(x)$ kad $\|x\| \to \infty$ ne sme biti brži od linearnog što bi moglo biti ograničenje. Ipak, ako uvedemo pretpostavku da je niz $\{x_k\}$ ograničen, onda to nije problem.

TEOREMA 30 *Neka su zadovoljeni uslovi A1 - A4. Onda za proizvoljnu početnu vrednost x_0 niz $\{x_k\}$ generisan SA algoritmom* (3.7) *konvergira skoro sigurno ka x^*, rešenju problema* (3.5) *kad $k \to \infty$.*

Ova teorema je dokazana u [13].

SA algoritam može da se primeni kada se traži minimum funkcije f tako što se traži rešenje sistema $g(x) = \nabla f(x) = 0$, a poznate su nam vrednosti sa šumom $G(x)$ iz (3.2).

U tom slučaju i ako je funkcija $v(\cdot)$ data sa (3.9), uslov (3.8) postaje:

$\overline{\text{A2}}$ $\overline{\text{(iii)}}$ Za svako $\varepsilon > 0$ postoji $\beta_\varepsilon > 0$ takvo da

$$\inf_{\|x-x_0\|>\varepsilon} (x-x_0)^T g(x) = \beta_\varepsilon > 0\,. \tag{3.10}$$

Brzina konvergencije SA postupka

Izbor dužina koraka $\{a_k\}$ u okviru tipa $\frac{1}{k}$ je bitan za brzinu konvergencije algoritma. Primetimo da izbori

$$a_k = \frac{a}{\sqrt{k+1}} \quad \text{ili} \quad a_k = \frac{a}{(k+1)^2} \quad \text{za} \quad a > 0$$

ne zadovoljavaju uslov A1 odnosno (3.6).

Spal (Spall) u [55] predlaže da se podešavanjem parametara $a > 0$, $\alpha \in (0.5, 1]$ i $A \geq 0$ u okviru standardnog oblika

$$a_k = \frac{a}{(k+1+A)^\alpha} \tag{3.11}$$

kontroliše robustnost i brzina konvergencije SA algoritma.

U radu [26] Fabian je pokazao da pod pogodnim uslovima $k^{\alpha/2}(x_k - x^*)$ konvergira u raspodeli za $k \to \infty$ ka slučajnom vektoru sa normalnom raspodelom $\mathcal{N}(0,\Sigma)$ za neku kovarijansnu matricu Σ.

Može se razumeti da je brzina kojom x_k teži ka x^* asimptotski u stohastičkom smislu $k^{-\alpha/2}$. Da bi uslov A1 ostao ispunjen, maksimalnu brzinu možemo dobiti za $\alpha = 1$ i ona iznosi $O(1/\sqrt{k})$.

Za optimizaciju nad konačnim uzorkom (finite-sample) maksimalna brzina se dobija za $\alpha < 1$. Postoji postupak za fino podešavanje vrednosti parametara a, α i A, Spall, [57].

Veličina A u (3.11), takozvana konstanta stabilnosti, se najčešće bira da bude $A = 0$. Ipak, taj izbor omogućava velike korake u ranim iteracijama, čak i sa relativno malom vrednosti a što zna dovesti do nestabilnosti.

Stoga se preporučuje (Spall, [55]) uzeti nenula vrednost A, recimo 5 do 10% ukupnog očekivanog broja iteracija. Ionako u kasnijim iteracijama A postaje zanemarljivo u odnosu na $k+1$, pa se dovoljna veličina dužine koraka može postići povećanjem vrednosti a.

SA algoritam koristi unapred definisane dužine koraka, odnosno: bez obzira koliko je trenutna iteracija x_k blizu rešenja x^*, dužina koraka se ne menja. Postoje razne metode za određivanje dužina koraka a_k koji se prilagođavaju trenutnoj iteraciji.

U radu [34] iz 1958. Kesten je razradio postupak u jednodimenzionalnom SA za prilagođavanje dužina koraka a_k na osnovu znaka razlike $x_k - x_{k-1}$. Kad je x_k blizu rešenja, znak se češće menja pa se može smanjivati dužina koraka. Kesten je dokazao skoro sigurnu konvergenciju svog postupka.

Višedimenzionalnu verziju Kestenove ideje su obradili Delyon i Juditsky u radu [19] iz 1993.

Ako se dužine koraka uzmu konstantne, postupak neće konvergirati, ali se može pokazati da konstantni koraci vode niz u okolinu rešenja. Ova ideja je dovela do upotrebe kaskadnih dužina koraka u SA šemi, Yousefian et al. [68]. Koriste se konstantne dužine koraka dok se ne dođe do neke okoline rešenja.

3.1.2 SA sa aproksimacijom gradijenta

Za minimizaciju dva puta diferencijabilne funkcije f možemo koristiti Njutnov iterativni postupak za rešavanje $f'(x) = 0$:

$$x_{k+1} = x_k - (f''(x_k))^{-1} f'(x_k).$$

Možemo koristiti i postupak negativnog gradijenta

$$x_{k+1} = x_k - a_k f'(x_k),$$

gde a_k nalazimo u svakoj iteraciji, recimo, pomoću linijskog pretraživanja. Ako ne raspolažemo gradijentom funkcije f, možemo ga aproksimirati.

Ako želimo rešiti problem minimizacije funkcije f, a poznate su nam samo vrednosti na koje je uticao šum, realizacije F iz (3.1), možemo uraditi sličnu stvar kao u determinističkom slučaju: možemo aproksimirati gradijent.

Postupak koji dobijamo je sličan postupku (3.7): za zadato polazno x_0 generišemo iterativni niz

$$x_{k+1} = x_k - a_k \hat{G}_k, \tag{3.12}$$

gde je $\hat{G}_k$ realizovana vrednost aproksimacije gradijenta koristeći realizacije F vrednosti funkcije f, (3.1).

Kiefer i Wolfowitz su 1952. u [35] posmatrali minimizaciju u jednoj dimenziji i za aproksimaciju gradijenta su koristili

$$\hat{G}_k = \frac{F(x_k + c_k) - F(x_k - c_k)}{2c_k},$$

koristeći za računanje vrednosti F, po formuli (3.1), $2k$-tu i $(2k-1)$-tu realizovanu vrednost funkcije šuma. Dokazali su L^2 konvergenciju iterativnog niza ka minimumu funkcije f, iz koje sledi konvergencija u verovatnoći.

Blum je u [7] pokazao skoro sigurnu konvergenciju iterativnog niza za višedimenzionalnu optimizaciju. Koristi se FD aproksimacija gradijenta: i-ta komponenta vektora gradijenta se aproksimira

$$\hat{G}_k^i = \frac{F(x_k + c_k e_i) - F(x_k)}{c_k}, i = 1, 2, \ldots, n. \tag{3.13}$$

Pritom se koristi $n+1$ realizovana vrednost funkcije f: po jedna u $x_k + c_k e_i$ za i-ti jedinični vektor e_i i jedna za x_k.

3.1.3 Konvergencija FDSA postupka

Da bi dokazao konvergenciju, slično kao Kiefer i Wolfowitz u [35], Blum je na a_k i c_k postavio uslove:

B1 $a_k > 0$ i $c_k > 0$, $k = 0, 1, \ldots$, $a_k \to 0$ za $k \to \infty$, $c_k \to 0$ za $k \to \infty$,

B2 $\sum_{k=0}^{\infty} a_k = \infty$,

B3 $\sum_{k=0}^{\infty} a_k c_k < \infty$,

B4 $\sum_{k=0}^{\infty} \left(\frac{a_k}{c_k}\right)^2 < \infty$.

Jedan mogući izbor nizova je $a_k = 1/(k+1)$ i $c_k = 1/(k+1)^{1/3}$.

Postupak (3.12) koji se dobija koristeći FD aproksimaciju gradijenta (3.13) nazivamo FDSA.

Ako se za aproksimaciju gradijenta koristi centralna diferencna šema, i-ta komponenta gradijenta se računa

$$\hat{G}_k^i = \frac{F(x_k + c_k e_i) - F(x_k - c_k e_i)}{2c_k}, i = 1, 2, \ldots, n, \tag{3.14}$$

kažemo da je postupak dvostrani FDSA. Primetimo da se za dvostrani FDSA funkcija F računa $2n$ puta u svakoj iteraciji.

Isto kao u teoremi 30, označavaćemo sa $\mathscr{F}_k$ sigma algebru generisanu realizovanim vrednostima $x_0, x_1, \ldots, x_k$. Koristimo oznake (3.1). Uslovi pod kojima je Spal dokazao konvergenciju dvostranog FDSA u [55] su:

A1' Niz dužina koraka $\{a_k\}$ i niz $\{c_k\}$ zadovoljavaju B1, B2, B3, B4.

A2' Postoji jedinstveni minimum x^* takav da za svako $\eta > 0$

$$\inf_{\|x-x^*\|>\eta} \|g(x)\| > 0, \quad g(x) = \nabla f(x) \quad \text{i} \quad \inf_{\|x-x^*\|>\eta} (f(x) - f(x^*)) > 0.$$

A3' Postoji $C > 0$ tako da za sve $i = 1, 2, \ldots, n$ i sve $k = 0, 1, 2, \ldots$

$$E\left(\xi_k^{(i+)} - \xi_k^{(i-)} \middle| \mathscr{F}_{k-1}\right) = 0, \; E\left(\left(\xi_k^{(i\pm)}\right)^2 \middle| \mathscr{F}_{k-1}\right) \leq C < \infty.$$

A4' Matrica Hesijana $\nabla^2 H(x) = \partial^2 f / \partial x \partial x^T$ postoji i ima uniformno ograničenu normu za sve $x \in \mathbb{R}^n$.

Teorema 31 *Neka $f : \mathbb{R}^n \to \mathbb{R}$ ima jedinstveni minimum x^*. Neka je za proizvoljno x_0 definisan iterativni niz $\{x_k\}$ po formulama* (3.12), (3.14) *i neka su zadovoljeni uslovi A1', A2', A3', A4'.*

Onda niz x_k konvergira ka x^* skoro sigurno za $k \to \infty$.

Ova teorema je dokazana u [55].

3.1.4 Brzina konvergencije FDSA postupka

Kod FDSA se takođe može razmatrati uticaj izbora koeficijenata a_k i c_k na brzinu konvergencije i važe slična razmatranja kao u glavi 3.1.

Standardni oblik za a_k i c_k je

$$a_k = \frac{a}{(k+1+A)^\alpha} \ \text{ i } \ c_k = \frac{c}{(k+1)^\gamma}, \tag{3.15}$$

gde su $a > 0$, $c > 0$, $\alpha > 0$, $\gamma > 0$ i $A \geq 0$.

Sacks je u radu [51] razmatrao asimptotsku raspodelu iteracija FDSA. Više o daljim rezultatima u ovoj oblasti ima u [55].

Ako su zadovoljeni uslovi teoreme 31 i ako

$$\beta = \alpha - 2\gamma > 0 \ \text{ i } \ 3\gamma - \frac{\alpha}{2} > 0\,,$$

i niz $\{x_k\}$ dobijen dvostranim FDSA postupkom (3.12), (3.14) sa a_k i c_k dobijenim po formulama (3.15) konvergira, onda $k^{\beta/2}(x_k - x^*)$ konvergira u raspodeli ka slučajnoj promenljivi sa normalnom raspodelom $\mathcal{N}(\mu_{FD}, \Sigma_{FD})$.

Drugim rečima, asimptotska stohastička brzina konvergencije niza dobijenog dvostranim FDSA postupkom je $k^{-\beta/2}$. Onda je lako videti da se maksimalna brzina postiže za $\alpha = 1$ i $\gamma = \frac{1}{6}$ i da ona iznosi $k^{-\beta/2} = 1/k^{1/3}$.

To pokazuje da poznavanje gradijenta i korišćenje SA umesto FDSA može da poboljša granicu maksimalne brzine sa $O(1/k^{1/3})$ na $O(1/k^{1/2})$.

U praktičnoj primeni se žele zadržati koraci koji se sporije smanjuju. Da bi se ispunili uslovi za konvergenciju, najniže vrednosti u praksi se uzimaju $\alpha = 0.602$ i $\gamma = 0.101$. Proces nalaženja boljih vrednosti a, A, α, c i γ se može polu automatizovati za određeni problem, vidi Spall [57].

3.1.5 SPSA postupak

SPSA je skraćenica za Simultaneous Perturbation Stochastic Approximation, misli se na stohastičku aproksimaciju (SA) gde se gradijent aproksimira istovremenim perturbacijama (SP).

Metod je uveo Spall u radu [56], prihvatili su ga i dalje razradili Hill i Fu u [31] i mnogi drugi. Ideja je da se u SA metodi (3.12) aproksimacija gradijenta vrši istovremenim perturbacijama umesto pojedinačnim perturbacijama po svakoj koordinati.

Formula aproksimacije gradijenta SPSA postupkom je

$$\hat{G}_k^i = \frac{F(x_k + c_k\Delta_k) - F(x_k - c_k\Delta_k)}{2c_k\Delta_{k,i}}, i = 1,2,\ldots,n, \tag{3.16}$$

gde je $[\Delta_{k,1}, \Delta_{k,2}, \ldots, \Delta_{k,n}]^T$ slučajni vektor nezavisnih slučajnih promenljivih sa istom, dozvoljenom raspodelom. Dozvoljene raspodele ne mogu imati nulu kao realizovanu vrednost skoro sigurno, a moraju biti centrirane oko nule. Takođe, dozvoljene raspodele moraju imati ograničen negativni momenat prvog ili višeg reda. Normalna i uniformna raspodela ne zadovoljavaju ove uslove. Najčešće korišćen primer raspodele koja zadovoljava uslove je Bernulijeva raspodela koja uzima vrednosti -1 i 1 sa verovatnoćom $\frac{1}{2}$.

Kako su imenioci svih komponenti (3.16) isti, možemo pisati

$$\hat{G}_k = \frac{F(x_k + c_k\Delta_k) - F(x_k - c_k\Delta_k)}{2c_k}[\Delta_{k,1}^{-1}, \Delta_{k,2}^{-1}, \ldots, \Delta_{k,n}^{-1}]^T. \tag{3.17}$$

Primetimo da se računanje funkcije F poziva samo dva puta bez obzira na dimenziju sistema n.

SPSA aproksimacija gradijenta je manje precizna od FDSA aproksimacije. Ali, ako se poređenje vrši sa istom kvotom računanja funkcije, tokom iterativnog postupka, SPSA metoda postiže bolje rezultate.

3.1.6 Konvergencija i brzina konvergencije SPSA

Spal je u [56] dokazao skoro sigurnu konvergenciju iterativnog niza ka jedinstvenom minimumu. Dokaz je dat tehnikom običnih diferencijalnih jednačina i najvažnija razlika od dosadašnjih dokaza o konvergenciji je da se zahteva apriori ograničenost iterativnog niza.

Uslovi na koeficijente a_k i c_k, $k = 0,1,\ldots$ su isti kao za dokaz konvergencije FDSA: B1 - B4. Ako se dodaju uslovi

$$\beta = \alpha - 2\gamma > 0 \text{ i } 3\gamma - \frac{\alpha}{2} > 0,$$

i niz $\{x_k\}$ dobijen SPSA postupkom (3.12), (3.16) sa a_k i c_k dobijenim po formulama (3.15) konvergira, onda $k^{\beta/2}(x_k - x^*)$ konvergira u raspodeli ka slučajnoj promenljivi sa normalnom raspodelom $\mathcal{N}(\mu_{SP}, \Sigma_{SP})$.

Isto kao kod FDSA, najbolja brzina konvergencije je $k^{-\beta/2} = k^{-1/3}$.

Ipak, isto kao kod FDSA, u praksi se želi sporije opadanje koeficijenata a_k. Stoga se uzimaju najmanje dozvoljene vrednosti, a to su, isto kao kod FDSA:

$$\alpha = 0.602 \ \text{ i } \ \gamma = 0.101.$$

Izjednačavanje asimptotske srednje kvadratne greške pri oceni parametara sa ovim vrednostima α i γ za postupke FDSA i SPSA dobijamo da količnik

$$\frac{\text{broj računanja funkcije } F \text{ FDSA postupkom}}{\text{broj računanja funkcije } F \text{ SPSA postupkom}} \to \frac{1}{n}.$$

Glava 4

Kvazi Njutnove metode u stohastičkoj optimizaciji

Mnogi rezultati u stohastičkoj optimizaciji su analogoni rezultata iz determinističke optimizacije.

Algoritmi sa Kvazi Njutnovim pravcima kada se dužine koraka određuju linijskim pretraživanjem postižu izvanredne rezultate u optimizaciji. Prirodno, pojavila se želja da se nađe analogon u stohastičkom okruženju.

Posmatramo problem (3.4), odnosno:

$$\min_{x \in D} f(x), \tag{4.1}$$

gde je $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ neprekidno diferencijabilna funkcija ograničena odole, sa gradijentom $g = \nabla f$. Nisu nam na raspolaganju tačne vrednosti funkcije f i g, već vrednosti na koje je uticao šum, kao u poglavlju II, formule (3.1) i (3.2).

Ako se za rešavanje problema (4.1) pokuša primeniti deterministički algoritam u kome se za vrednosti funkcije i gradijenta uzimaju vrednosti na koje je uticao šum, ne možemo garantovati konvergenciju postupka.

U glavama koje slede je izložen originalni doprinos ove disertacije.

4.1 Stohastička optimizacija sa opadajućim pravcima

U odsustvu šuma, odnosno u determinističkom okruženju, SA metod je gradijentni metod sa predefinisanom dužinom koraka. U determinističkom okruženju, kada su pravci negativni gradijenti, linijsko pretraživanje optimalne dužine koraka dalo bi efikasniji postupak od postupka sa predefinisanim dužinama koraka. Ideja imitiranja linijskog pretraživanja objavljena je u radovima Vardija (Wardi) [60, 61]. U tim radovima uslovi za konvergenciju zahtevaju povećavano uzorkovanje za računanje funkcije cilja i njenog gradijenta što značajno poskupljuje metod.

Pregled unapređenja SA metode raznim aproksimacijama gradijenta i Hesijana ima u Spall [54]. Takođe, u tom radu, Spal je opisao metodu u kojoj koristi informacije drugog reda tako što aproksimira Hesijan slično apriksimacijama gradijenta simultanim perturbacijama (SP).

Taj postupak se naziva Adaptive simultaneous perturbation (ASP). ASP se sastoji iz dve paralelne rekurzivne formule: jedna za iteracije x_k i jedna za aproksimacije Hesijana funkcije f:

$$x_{k+1} = x_k - a_k \bar{\bar{H}}_k^{-1} G_k\,, \;\; \bar{\bar{H}}_k = \psi_k(\bar{H}_k) \tag{4.2a}$$

$$\bar{H}_k = \frac{k}{k+1}\bar{H}_{k-1} + \frac{1}{k+1}\hat{H}_k\,, \;\; k = 0,1,\ldots \tag{4.2b}$$

gde su a_k dužine koraka, kao u glavi 3.1, G_k realizovana vrednost gradijenta sa šumom za vrednost x_k, $\psi_k(\cdot)$ funkcija koja matrici Hesijana $n \times n$ pridružuje pozitivno definitnu aproksimaciju istog formata. $\hat{H}_k$ je uobičajena aproksimacija matrice Hesijana.

Jednačina (4.2a) je stohastički analogon kvazi Njutnovih postupaka iz determinističkog slučaja. Jednačina (4.2b) je formula za ažuriranje aproksimacije Hesijana koja prigušuje uticaj šuma. U implementaciji kad raspolaže gradijentom $g = \nabla f$, Spal postupak naziva 2SG, a kad se za aproksimaciju gradijenta koriste simultane preturbacije drugog reda funkcije cilja f, postupak naziva 2SPSA.

U postupku 2SG, Spal koristi aproksimaciju Hesijana

$$\hat{H}_k = \frac{1}{2}\left[\frac{\delta G_k^T}{2c_k\Delta_k} + \left(\frac{\delta G_k^T}{2c_k\Delta_k}\right)^T\right],$$

gde je

$$\delta G_k = G_k(x_k + c_k\Delta_k) - G_k(x_k - c_k\Delta_k)\,,$$

slučajni vektor nezavisnih slučajnih promenljivih sa istom, dozvoljenom raspodelom, kao u glavi 3.1.5.

U radu Bertsekas [4] se posmatra postupak sličan SA postupku i dokazuje da niz normi gradijenata u iteracijama teži nuli.

Druge modifikacije SA postupka u kojima se pokušava popraviti postupak prilagođavanjem pravaca u iteracijama imamo u radovima Xu [66], [65]. Takođe Xu posmatra gradijentni postupak (sa šumom) i adaptivnim dužinama koraka [67].

Glavna mana linijskog pretraživanja u stohastičkom slučaju je da su dužine koraka prevelike za iteracije u blizini rešenja, i da ne daju kvalitetne efekte prigušenja šuma koje ima SA metod i ne garantuju konvergenciju.

Ipak, čak i u stohastičkom okruženju, daleko od stacionarne tačke, koraci dobijeni line search tehnikom su efikasniji od predefinisanih koraka SA metode.

Mogućnost primene postupaka drugog reda je takođe važna, kao što je urađeno sa problemima mašinskog učenja, radovi Byrd i ostali [10, 11, 12].

U radu Krejić i ostali [39], razvijena je ideja da se na početku iterativnog postupka primenjuje linijsko pretraživanje za generisanje dužina koraka (dok smo daleko od rešenja) i prelazi na predefinisane SA korake kad se približimo rešenju i kada oni postanu bolji. Uslovi za konvergenciju nisu zahtevali usrednjavanje povećavanjem uzorka radi smanjenja šuma. Ti rezultati u [38] su unapređeni i testirani na velikom broju test primera sa dve aproksimacije Hesijana. Rezultati su originalni doprinos disertacije i predstavljeni su u nastavku.

Određivanje rednog broja iteracije u kojoj treba preći na predefinisane korake je kritično za kvalitet postupka. Algoritam mora da omogući skoro sigurnu konvergenciju novog postupka, ali i da omogući da kvalitet line search dužina koraka ubrza postupak.

Isto kao u glavi 3.1.2 posmatramo problem nalaženja vrednosti x^* za koju se ostvaruje minimum funkcije

$$\min_{x \in \mathbb{R}^n} f(x) = f(x^*), \tag{4.3}$$

za neprekidno diferencijabilnu funkciju $f : D \subset \mathbb{R}^n \to \mathbb{R}$ ograničenu odole na D. Posmatramo ili probleme bez ograničenja, $D = \mathbb{R}^n$ ili probleme kod kojih D ne igra ulogu.

Pretpostavljamo da su poznate samo vrednosti:

$$F(x) = f(x) + \xi(x) \quad \text{i} \quad G(x) = g(x) + \varepsilon(x), \tag{4.4}$$

gde su $\xi(x)$ i $\varepsilon(x)$ redom slučajna promenljiva i n-dimenzionalni slučajni vektor definisani na prostoru verovatnoće $(\Omega, \mathscr{F}, P)$. Primetimo da $\xi(x)$ i $\varepsilon(x)$ zavise od vrednosti $x \in \mathbb{R}^n$.

Pretpostavljamo da postoji jedinstveno x^* koje je rešenje problema (4.3) za koje je zadovoljeno

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

4.1.1 Linijsko pretraživanje ili SA, motivacija

Iterativni niz za rešavanje problema (4.3) postupkom SA iz glave 3.1.2 definisan u (3.12) glasi

$$x_{k+1} = x_k - a_k G_k, \tag{4.5}$$

gde je G_k aproksimacija gradijenta $g(x_k) = \nabla f(x_k)$, i $a_k > 0$ predefinisan niz koeficijenata koji zadovoljava uslove (3.6) date u glavi 3.1.

Takođe je u glavi 3.1 razmatran uticaj dužine koraka a_k na brzinu konvergencije. Standardni izbor dužine koraka (3.11) je previše "oprezan" i daje male vrednosti za dužine koraka od početne iteracije ako je početna iteracija daleko od rešenja. Tada je SA postupak problematično spor.

Iako asimptotski SA koraci garantuju skoro sigurnu konvergenciju, u praksi se posle nekog konačnog broja koraka mora stati. Ako je početna tačka bila daleko od krajnjeg rešenja, kada se prestane sa iteracijama, poslednja iteracija je još uvek daleko od rešenja.

Jedan pravac za prevazilaženje ovog problema su adaptivne dužine koraka, koje se podešavaju uzimajući informacije iz trenutne iteracije. Vidi [19, 34, 67].

U determinističkom okruženju, dužine koraka se za gradijentni postupak dobijaju linijskim pretraživanjem opisanim u glavi 2.2.6. Postupak linijskog pretraživanja nalazi dužinu koraka blizu optimalne uz zadovoljenost uslova koji garantuju konvergenciju, recimo (2.22a). Da bi se ovaj postupak primenio u stohastičkom okruženju, neophodno je njegovo prilagođavanje.

4.1.2 Stohastički postupak linijskog pretraživanja po opadajućim pravcima

Za zadato $x_0 \in \mathbb{R}^n$ posmatraćemo opšti iterativni postupak

$$x_{k+1} = x_k + a_k d_k, \; k = 0, 1, \ldots,$$

gde su $a_k \in \mathbb{R}$ dužine koraka i $d_k \in \mathbb{R}^n$ pravci za $k = 0, 1, \ldots$.

Negativni gradijenti u okruženju sa šumom $d_k = -G_k$ ne moraju uvek biti opadajući pravci. Iz determinističkog okruženja znamo da postoje bolji izbori pravaca iterativnog postupka od negativnog gradijenta.

Iskorišćavanje informacije drugog reda kao što su Njutnovi ili Kvazi Njutnovi pravci je teško u stohastičkom okruženju. Aproksimacije BFGS pravaca su razmatrane u radu Schraudolph i ostali [53].

Pretpostavljamo da je poznata aproksimacija gradijenta u k-toj iteraciji

$$G_k = G(x_k) = g(x_k) + \varepsilon_k(x_k)\,.$$

Za pravac d_k kažemo da je opadajući u x_k ako

$$G_k^T d_k < 0\,. \tag{4.6}$$

Pritom nam je poznata vrednost funkcije sa šumom

$$F_k = F(x_k) = f(x_k) + \xi_k(x_k)\,.$$

Onda Armijo pravilo (2.22a) postaje

$$F_k(x_k + a_k d_k) \leq F_k + c_1 a_k G_k^T d_k, \tag{4.7}$$

gde je d_k posmatrani opadajući pravac u x_k.

Ovde ćemo razmatrati kombinaciju dve metode za određivanje dužine koraka iterativnog postupka za opadajuće pravce. U prvoj fazi ćemo koristiti linijsko pretraživanje, da bismo u jednom momentu prešli na "sigurne" SA dužine koraka. Određivanje koraka iterativnog postupka u kome treba napraviti taj prelazak je najveći izazov u implementaciji ovog postupka.

Ako posmatramo relativnu grešku ocene gradijenta

$$r = \frac{g(x_k) - G_k}{\|g(x_k)\|},$$

pretpostavićemo da je mala, $\|r\| \approx 0$ i dok je $\|G_k\| > C$, za unapred odabranu konstantu C. Za veliko $\|G_k\|$, možemo smatrati da smo daleko od rešenja. Daleko od rešenja šum nema dominantan uticaj na normu gradijenta i opadanje funkcije. To su iteracije u kojima ćemo koristiti linijsko pretraživanje koje oponaša determinističko linijsko pretraživanje.

Kada smo blizu rešenja, kad je $\|x - x^*\| < \varepsilon$, za neko malo $\varepsilon > 0$, više ne možemo pretpostavljati da linijsko pretraživanje, koje zadovoljava Armijo uslov (4.7), daje zadovoljavajuće korake. Naime, šum tada ima značajan uticaj na promenu F_k, isto i šum koji utiče na G_k, a koji je odabran različitim uzorkom. U blizini rešenja veličina šuma je dominantna. Tada naš algoritam prelazi na korake SA metode po opadajućim pravcima koji zadovoljavaju uslove (3.6).

Opisaćemo sada algoritam Stohastičke aproksimacije po opadajućim pravcima, DSLS (Descent Stochastic Line Search).

Algoritam 4 DSLS

Ulaz: $x_0 \in \mathbb{R}^n$, $c_1 \in (0,1)$, $C, \underline{\delta}(C) > 0$ i $\{a_k\} \in \mathbb{R}$ za koje važi (3.6).
Postavi $k = 0$, $phase = 1$.

Korak 1 Odaberi opadajući pravac (koji zadovoljava (4.6)).

Korak 2 Odaberi korak a_k.

Korak 3 Odredi novu iteraciju $x_{k+1} = x_k + a_k d_k$.

Korak 4 Postavi $k = k+1$ pređi na korak 1.

Daćemo dodatnu razradu koraka 2, koji je ključni za algoritam.

Algoritam 5 Odaberi korak a_k.

Korak 2

Korak 2.1 Ako $phase = 1$ pređi na korak 2.2, inače pređi na korak 2.3

Korak 2.2 Ako $\|G_k\| \geq C$ odaberi $a_k > \underline{\delta}(C)$ tako da zadovoljava (4.7). Pređi na korak 3. Inače postavi $phase = 2$ i pređi na korak 2.3.

Korak 2.3 Odaberi a_k iz predefinisanog ulaza a_k.

4.1.3 Konvergencija DSLS

Pokažimo prvo konvergenciju SA metode sa opadajućim pravcima. Neka je za dato x_0 definisan iterativni niz

$$x_{k+1} = x_k + a_k d_k, \; k = 0, 1, \ldots, \tag{4.8}$$

gde je d_k niz opadajućih pravaca koji zadovoljava (4.6) i $\{a_k\}$ niz dužina koraka tipa $\frac{1}{k}$, (zadovoljava (3.6)). Problem sličan ovom izučavaju Bertsekas i Tsitsiklis [4], gde se iterativni niz dobija rekurzivnom formulom

$$x_{k+1} = x_k + a_k(s_k + w_k).$$

Uslovi za konvergenciju u [4] sadrže pretpostavku da s_k i gradijent zadovoljavaju sledeće veze:

$$c_1\|g(x_k)\|^2 \leq -g(x_k)^T s_k, \ \ \|s_k\| \leq c_2(1+\|g(x_k)\|), \ \ k = 0,1,\ldots$$

Ovde ćemo dokazati skoro sigurnu konvergenciju koristeći pretpostavke A1, A3, A4 za konvergenciju SA postupka iz glave 3.1.1 uz još neke pretpostavke.

Označimo sa $\mathscr{F}_k, k = 0,1,\ldots$ σ-algebru generisanu iteracijama $x_0, x_1, \ldots, x_k$. Navedimo potrebne pretpostavke.

C1 Niz dužina koraka $\{a_k\}$ je tipa $\frac{1}{k}$, odnosno, zadovoljava (3.6):

$$a_k > 0, k \in \mathbb{N}, \ \sum_{k=0}^{\infty} a_k = \infty, \ \sum_{k=0}^{\infty} a_k^2 < \infty.$$

C2 Niz šumova sa generisanim sigma algebrama $(\varepsilon_k, \mathscr{F}_k)$ je niz martingal razlika za neopadajući niz sigma algebri $\mathscr{F}_k$ takav da važi

$$E(\varepsilon_k|\mathscr{F}_{k-1}) = 0, \ E(\|\varepsilon_k\|^2) < \infty.$$

C3 Funkcija g i uslovni drugi momenat realizovanog šuma su ograničeni:

$$\|g(x)\|^2 + E(\|\varepsilon_k\|^2|\mathscr{F}_{k-1}) < c(1+v(x)), k = 0,1,\ldots,$$

za neko $c > 0$.

C4 Za sve pravce $d_k, k = 0,1,\ldots$ važi

$$(x_k - x^*)^T E[d_k|\mathscr{F}_k] \leq -c_2\|x_k - x^*\| \ \text{ skoro sigurno},$$

za neko $c_2 > 0$.

C5 Postoji $c_3 > 0$ tako da za sve $k = 0, 1, \ldots$

$$\|d_k\| \le c_3 \|G_k\| \text{ skoro sigurno.}$$

Pretpostavke C1, C2 i C3 su iste kao redom A1, A3 i A4 u teoremi konvergencije SA postupka iz glave 3.1.1.

Pretpostavke C4 i C5 zamenjuju pretpostavku A2 iz glave 3.1.1.

Pretpostavka C4 ograničava uticaj šuma i ekvivalentna je sa C4 iz Spall [54].

Ako $d_k = -G_k$, onda je C4 zadovoljeno ako postoji konstanta $c_2 > 0$ takvo da

$$(x_k - x^*)^T g_k \ge c_2 \|x_k - x^*\| \text{ skoro sigurno za sve } k = 0, 1, \ldots.$$

Pretpostavka C5 povezuje realizovane vrednosti gradijenta sa šumom i opadajuće pravce. Naravno, ako stavimo $d_k = -G_k$ dobijamo da je C5 zadovoljeno za proizvoljnu konstantu $c_3 \ge 1$.

Sledi teorema o konvergenciji SA postupka po opadajućim pravcima (4.8).

Teorema 32 *Neka je za dato x_0 generisan iterativni niz (4.8) po opadajućim pravcima d_k. Neka su zadovoljeni uslovi C1 - C5. Onda niz x_k konvergira ka x^*, rešenju problema* (4.3).

Dokaz. Iz $x_{k+1} = x_k + a_k d_k$ i pretpostavki C4 i C5 imamo

$$\begin{aligned}
E[\|x_{k+1} - x^*\|^2 | \mathscr{F}_k] &= E[\|x_k + a_k d_k - x^*\|^2 | \mathscr{F}_k] \\
&= \|x_k - x^*\|^2 + 2a_k (x_k - x^*)^T E[d_k | \mathscr{F}_k] + a_k^2 E[\|d_k\|^2 | \mathscr{F}_k] \\
&\le \|x_k - x^*\|^2 - 2c_2 a_k \|x_k - x^*\| + c_3^2 a_k^2 E[\|G_k\|^2 | \mathscr{F}_k].
\end{aligned}$$

Zbog $G_k = g_k + \varepsilon_k$ pretpostavke C2 i C3 impliciraju

$$\begin{aligned}
E[\|G_k\|^2 | \mathscr{F}_k] &= E[\|g_k + \varepsilon_k\|^2 | \mathscr{F}_k] \\
&= \|g_k\|^2 + 2g_k^T E[\varepsilon_k | \mathscr{F}_k] + E[\|\varepsilon_k\|^2 | \mathscr{F}_k] \\
&= \|g_k\|^2 + E[\|\varepsilon_k\|^2 | \mathscr{F}_k] \\
&< c(1 + \|x_k - x^*\|^2)\,.
\end{aligned}$$

$$\begin{aligned}
E[\|x_{k+1} - x^*\|^2 | \mathscr{F}_k] &\le \|x_k - x^*\|^2 - 2c_2 a_k \|x_k - x^*\| + c_3^2 a_k^2 E[\|G_k\|^2 | \mathscr{F}_k] \\
&< \|x_k - x^*\|^2 - 2c_2 a_k \|x_k - x^*\| + c c_3^2 a_k^2 (1 + \|x_k - x^*\|^2) \\
&= (1 + c c_3^2 a_k^2) \|x_k - x^*\|^2 - 2c_2 a_k \|x_k - x^*\| + c c_3^2 a_k^2
\end{aligned}$$

Uzimamo $U_k = \|x_k - x^*\|^2$, $\beta_k = cc_3^2 a_k^2$, $\xi_k = cc_3^2 a_k^2$ i $\zeta_k = 2c_2 a_k \|x_k - x^*\|$, pritom znamo da iz $\sum_{k=0}^{\infty} a_k^2 < \infty$ sledi $\sum_{k=0}^{\infty} \beta_k < \infty$ i $\sum_{k=0}^{\infty} \xi_k < \infty$.

Teorema 18 iz glave 1.1.7 daje da U_k konvergira skoro sigurno ka slučajnoj promenljivi i da $\sum_{k=0}^{\infty} \zeta_k < \infty$. Onda $a_k \|x_k - x^*\| \to 0$ skoro sigurno za $k \to \infty$.

Pošto je $\sum_{k=0}^{\infty} a_k = \infty$ sledi

$$\|x_k - x^*\| \to 0 \text{ skoro sigurno kad } x \to \infty\,.$$

Posledica je da $x_k \to x^*$ skoro sigurno kad $x \to \infty$. ■

Dalje, neka je $\{x_k\}$ niz generisan algoritmom DSLS.

Pretpostavke kojima ćemo pokazati konvergenciju za metodu sa linijskim pretraživanjem uključuju i Lipšic neprekidnost gradijenta funkcije cilja. Tako da dodajemo pretpostavku

C6 Gradijent $g(x) = \nabla f(x)$ je Lipšic neprekidan, to jest, postoji $L > 0$ tako da

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

Treba nam i dodatna pretpostavka za šum:

C7 Realizacije šuma su ograničene i postoji $M > 0$ takvo da

$$\|\xi_k(x)\| \leq M, \ \|\varepsilon_k(x)\| \leq M \text{ skoro sigurno}$$

za sve $k = 0, 1, \ldots$ i $x \in D$.

DSLS metod koristi opadajuće pravce i linijsko pretraživanje, zato imamo i pretpostavke koje su uobičajene za opadajuće pravce u determinističkoj optimizaciji ([21]).

C8 Postoji $\delta > 0$ takvo da za sve $k = 0, 1, \ldots$

$$G_k^T d_k \leq -\delta \|G_k\| \|d_k\| \text{ skoro sigurno.}$$

C9 Postoji $\underline{\Delta} \in (0, \Delta)$ takvo da za sve $k = 0, 1, \ldots$ važi

$$\|d_k\| \geq \underline{\Delta} \text{ skoro sigurno.}$$

Sledeća teorema pokazuje da je algoritam DSLS dobro definisan, a naredna posle nje pokazuje da DSLS algoritam generiše konačan broj koraka linijskim pretraživanjem.

TEOREMA 33 *Neka važe pretpostavke C5 - C8. Neka*

$$C \geq \frac{M + 2\sqrt{2ML} + 1}{\delta(1 - c_1)}.$$

Onda skoro sigurno postoji $\underline{\delta}(C) > 0$ *kojim je algoritam DSLS dobro definisan.*

Dokaz. Označimo sa f_k i g_k redom vrednosti funkcije cilja f i njenog gradijenta g u vrednostima $x = x_k$. Neka je $\alpha > 0$ i $d \in \mathbb{R}^n$ proizvoljno. Onda

$$\begin{aligned} f(x_k + \alpha d) &= f_k + \alpha g(x_k + t\alpha d)^T d \\ &= f_k + \alpha g(x_k + t\alpha d)^T d + \alpha g_k^T d - \alpha g_k^T d \\ &= f_k + \alpha g_k^T d + \alpha (g(x_k + t\alpha d) - g_k)^T d \\ &\leq f_k + \alpha g_k^T d + \alpha \|g(x_k + t\alpha d) - g_k\| \cdot \|d\|, \end{aligned}$$

za neko $t \in (0,1)$. Pretpostavka C6 i $t \in (0,1)$ daju

$$f(x_k + \alpha d) \leq f_k + \alpha g_k^T d + \alpha^2 L \|d\|^2.$$

Zato što su f i g dobijene sa šumom

$$F_k(x_k + \alpha d_k) = f(x_k + \alpha d_k) + \tilde{\xi}_k,$$

$$F_k = f_k + \xi_k \quad \text{and} \quad G_k = g_k + \varepsilon_k,$$

gde je korišćena oznaka $\tilde{\xi}_k = \xi_k(x_k + \alpha d_k)$.

Uzimajući $d = d_k$ imamo

$$\begin{aligned} F(x_k + \alpha d_k) &= f(x_k + \alpha d_k) + \tilde{\xi}_k \\ &\leq f_k + \alpha g_k^T d_k + \alpha^2 L \|d_k\|^2 + \tilde{\xi}_k \\ &= F_k + \alpha G_k^T d_k - \alpha \varepsilon_k^T d_k + \alpha^2 L \|d_k\|^2 + \tilde{\xi}_k - \xi_k \\ &\leq F_k + \alpha G_k^T d_k + \alpha M \|d_k\| + \alpha^2 L \|d_k\|^2 + 2M \end{aligned} \tag{4.9}$$

skoro sigurno, za $k = 0, 1, \ldots$.

Sada treba da dokažemo da skoro sigurno postoji $\underline{\delta}(C) > 0$ i $\overline{\alpha} > \underline{\delta}(C)$ tako da za $\alpha \in (\underline{\delta}(C), \overline{\alpha})$ važi

$$F_k + \alpha G_k^T d_k + \alpha M \|d_k\| + \alpha^2 L \|d_k\|^2 + 2M \ \leq \ F_k + c_1 \alpha G_k^T d_k, \tag{4.10}$$

što je ekvivalentno sa

$$\alpha(1-c_1) G_k^T d_k + \alpha M \|d_k\| + \alpha^2 L \|d_k\|^2 + 2M \leq 0.$$

Iz pretpostavke C8 sledi

$$\begin{aligned} &\alpha(1-c_1) G_k^T d_k + \alpha M \|d_k\| + \alpha^2 L \|d_k\|^2 + 2M \\ &\leq \alpha^2 L \|d_k\|^2 - \delta\alpha(1-c_1)\|G_k\|\|d_k\| + \alpha M \|d_k\| + 2M, \end{aligned}$$

pa će se dokaz dobiti iz analize kvadratne funkcije

$$\phi(\alpha) = \alpha^2 L \|d_k\|^2 - \delta\alpha(1-c_1)\|G_k\|\|d_k\| + \alpha M \|d_k\| + 2M.$$

Definišimo

$$\begin{aligned} A_\phi &= L\|d_k\|^2 > 0, \\ B_\phi &= -\delta(1-c_1)\|G_k\|\|d_k\| + M\|d_k\| < 0, \\ C_\phi &= 2M > 0. \end{aligned}$$

Želimo pokazati da je $B_\phi^2 - 4A_\phi C_\phi > 0$. Posmatrajmo nule kvadratne funkcije

$$\psi(u) = \delta^2(1-c_1)^2 u^2 - 2\delta M(1-c_1)u + M^2 - 8ML.$$

Postoje $u_1 < u_2 \in \mathbb{R}$ takvi da $\psi(u_1) = \psi(u_2) = 0$, i $\psi(u) > 0$ za sve $u > u_2$, gde je

$$u_2 = \frac{2\delta M(1-c_1) + \sqrt{32\delta^2(1-c_1)^2 ML}}{2\delta^2(1-c_1)^2} = \frac{M + 2\sqrt{2ML}}{\delta(1-c_1)}.$$

Kako je $\|G_k\| \geq C > u_2$ zaključujemo da je $B_\phi^2 - 4A_\phi C_\phi > 0$ zadovoljeno i stoga funkcija $\phi(\alpha)$ ima dve realne nule

$$\alpha_1 = \frac{-B_\phi - \sqrt{B_\phi^2 - 4A_\phi C_\phi}}{2A_\phi} \text{ and } \alpha_2 = \frac{-B_\phi + \sqrt{B_\phi^2 - 4A_\phi C_\phi}}{2A_\phi}$$

i važi (4.10) za sve $\alpha \in (\alpha_1, \alpha_2)$. Primetimo i da je $0 < \alpha_1 < -\frac{B_\phi}{2A_\phi} < \alpha_2$.

Pokažimo da je (4.10) ispunjeno za skup vrednosti α koji je uniformno ograničen odole.

Dovoljno je da nađemo donju granicu $\underline{\alpha} > 0$ nezavisnu od k, zato $-\frac{B_\phi}{2A_\phi} \geq \underline{\alpha}$.

Iz pretpostavke C5 imamo

$$-\frac{B_\phi}{2A_\phi} = \frac{\delta(1-c_1)\|G_k\| - M}{2L\|d_k\|} \geq \frac{\delta(1-c_1)\|G_k\| - M}{2Lc_3\|G_k\|}.$$

Pošto je $\|G_k\| \geq C > 0$,

$$\frac{\delta(1-c_1)\|G_k\| - M}{2Lc_3\|G_k\|} = \frac{\delta(1-c_1) - M/\|G_k\|}{2Lc_3} \geq \frac{\delta(1-c_1) - M/C}{2Lc_3},$$

i iz $C \geq \frac{M+2\sqrt{2ML}+1}{\delta(1-c_1)}$ imamo

$$\frac{\delta(1-c_1) - M/C}{2Lc_3} \geq \frac{\delta(1-c_1)(2\sqrt{2ML}+1)}{2Lc_3(M+2\sqrt{2ML}+1)}.$$

Stoga možemo uzeti

$$\underline{\alpha} = \frac{\delta(1-c_1)(2\sqrt{2ML}+1)}{2Lc_3(M+2\sqrt{2ML}+1)}.$$

Zaključujemo da za $\underline{\delta}(C) = \max\{\alpha_1, \underline{\alpha}\}$ skoro sigurno postoji $a_k \in (\underline{\delta}(C), \alpha_2)$ takvo da (4.10) važi i da je skup vrednosti a_k skoro sigurno ograničen odole. ∎

Pokažimo sada da Algoritam DSLS sigurno prelazi na SA korake.

TEOREMA 34 *Neka važe pretpostavke C6 - C9. Neka je*

$$C \geq \max\{\frac{2M+1}{\underline{\alpha} c_1 \delta \underline{\Delta}}, \frac{M+2\sqrt{2ML}+1}{\delta(1-c_1)}\},$$

$$\underline{\alpha} = \frac{\delta(1-c_1)(2\sqrt{2ML}+1)}{2Lc_3(M+2\sqrt{2ML}+1)}.$$

Neka je $\{x_k\}$ niz generisan Algoritmom DSLS. Neka je $\{x_j\}, j \in J$ podniz za koji važi

$$\|G_j\| \geq C. \tag{4.11}$$

Onda je $\{x_j\}$ skoro sigurno konačan.

Dokaz. Pretpostavimo suprotno: neka je niz $\{x_j\}$ beskonačan. Ako je (4.11) zadovoljeno, onda

$$\|G_j\| \geq \frac{M + 2\sqrt{2ML} + 1}{\delta(1 - c_1)}$$

i teorema 33 tvrdi da je za sve $j \in J$ sledeća iteracija x_{j+1} dobijena skoro sigurno linijskim pretraživanjem tako da

$$F_j(x_{j+1}) \leq F_j(x_j) + c_1 a_j G_j^T d_j, \quad x_{j+1} = x_j + a_j d_j, \; a_j > \underline{\delta}(C) \geq \underline{\alpha}.$$

Dalje, neka su i vrednosti koje slede dobijene skoro sigurno linijskim pretraživanjem i neka

$$F_i(x_{i+1}) \leq F_i(x_i) + c_1 a_i G_i^T d_i, \quad a_i > \underline{\alpha}, \; i = 0, 1, \ldots, j.$$

Kako je

$$F_i(x_{i+1}) = f(x_{i+1}) + \tilde{\xi}_i, \quad F_i(x_i) = f(x_i) + \xi_i,$$

$\xi_i - \tilde{\xi}_i \leq 2M$ skoro sigurno, $a_i > \underline{\alpha}$, $\|G_i\| \geq C$ imamo

$$\begin{aligned} f(x_{i+1}) &\leq f(x_i) + c_1 a_i G_i^T d_i + \xi_i - \tilde{\xi}_i \\ &< f(x_i) - c_1 \delta \underline{\alpha} C \underline{\Delta} + 2M \end{aligned} \tag{4.12}$$

skoro sigurno. Uzmimo $K = c_1 \delta \underline{\alpha} C \underline{\Delta} - 2M > 0$,

$$f(x_{i+1}) < f(x_i) - K, \; i = 0, 1, 2, \ldots, j.$$

Sabirajući gornje nejednakosti za sve $j \in J$ imamo

$$\sum_{i=0}^{j} f(x_{i+1}) < \sum_{i=0}^{j} f(x_i) - \sum_{i=0}^{j} K$$

$$f(x_{j+1}) - f(x_0) < -\sum_{i=0}^{j} K,$$

odakle

$$\sum_{i=0}^{j} K < f(x_0) - f(x_{j+1}) \leq K_1$$

pošto je f ograničena odole. Kako je $K > 0$, imamo

$$j + 1 < K_1 / K = K_2$$

za proizvoljno $j \in J$. Ovo je u suprotnosti sa pretpostavkom da je niz $\{x_j\},\ j \in J$ beskonačan, odakle sledi tvrđenje naše teoreme. ∎

Dokažimo sada glavnu teoremu o konvergenciji DSLS algoritma koristeći dokazane teoreme.

TEOREMA 35 *Neka važe pretpostavke C1 - C9. Neka*

$$
\begin{aligned}
C &\geq \max\{\frac{2M+1}{\underline{\alpha} c_1 \delta \underline{\Delta}}, \frac{M+2\sqrt{2ML}+1}{\delta(1-c_1)}\},\\
\underline{\alpha} &= \frac{\delta(1-c_1)(2\sqrt{2ML}+1)}{2Lc_3(M+2\sqrt{2ML}+1)}.
\end{aligned}
$$

Neka je $\{x_k\}$ niz generisan Algoritmom DSLS. Onda x_k konvergira skoro sigurno ka x^*, rešenju problema (4.3).

Dokaz.

Iz teoreme 34 imamo da skoro sigurno postoji konačno mnogo koraka koji se generišu linijskim pretraživanjem. Onda Algoritam DSLS generiše beskonačno iteracija koje zadovoljavaju uslove C1 - C5 i onda teorema 32 daje da $x_k \to x^*$ skoro sigurno. ∎

4.1.4 Implementacija DSLS postupka

Ovde ćemo se baviti implementacijom Algoritma DSLS na rešavanje problema (4.3) uz poznavanje vrednosti funkcije f i gradijenta g sa šumom, redom F i G, odnosno, uz poznavanje njihovih realizacija F_k i G_k. Pri tome ćemo za računanje F_k i G_k koristiti male uzorke veličine p jednako raspoređenih nezavisnih slučajnih promenljivih, odnosno vektora, sa istom raspodelom kao F_k i G_k redom. Za realizacije F_k i G_k ćemo koristiti aritmetičke sredine:

$$
F_k = F(x_k) = \frac{1}{p}\sum_{i=1}^{p} F(x_k, \xi_{k^i}),\ G_k = \nabla F(x_k) = \frac{1}{p}\sum_{i=1}^{p} G(x_k, \varepsilon_{k^i}), \tag{4.13}
$$

za neku malu vrednost p.

Da bismo izvršili implementaciju DSLS, prvo ćemo se dotaći nekoliko bitnih tačaka.

Dokazana teorema 35 o konvergeniciji DSLS postupka oslanja se na poznavanje konstante C koja određuje iteraciju u kojoj će se preći sa računanja dužine koraka linijskim pretraživanjem na predefinisane korake tipa $\frac{1}{k}$. Konstanta C zavisi od Lipšicove konstante L i granice šuma M, jasno je da se ne može precizno odrediti.

Pravci d_k iz (4.8) treba da su opadajući u ovom postupku i bilo bi dobro kad bismo mogli koristiti pravce koji u sebi sadrže i informaciju drugog reda, a ne samo pravce negativnog gradijenta.

Dužine koraka u (4.8) takođe imaju slobodu biranja u okviru skupa dozvoljenih nizova tipa $\frac{1}{k}$.

Često se u stohastičkoj optimizaciji kao izlazni kriterijum koristi kvota računanja funkcije. Linijsko pretraživanje zasnovano na interpolaciji, Nocedal and Wright [47], Dennis and Schnabel [20], u proceduri linijskog pretraživanja vodi račun o broju izračunavanja funkcije cilja. Popularne implementacije (recimo Matlab) ove metoda linijskog pretraživanja imaju u sebi ugrađen deo koji vodi evidenciju o broju izračunatih funkcija cilja.

Mi ćemo postaviti kvotu $maxfcalcls$ na broj računanja funkcije cilja u linijskom pretraživanju za trenutni pravac. Ako se za tu kvotu ne pronađe zadovoljavajući novi korak, zaključićemo da linijsko pretraživanje više nije uspešno i to će nam biti iteracija počev od koje ćemo preći na predefinisane SA korake.

Praktično, mi ćemo koristiti Algoritam 2 Line Search iz glave 2.2.9 sa unapred zadatom kvotom računanja funkcije. Ako nam Line Search algoritam vrati $retcode = 2$, što znači da nije uspeo naći zadovoljavajuću tačku i da je potrošio $maxfcalcls$ računanja funkcije cilja, za iterativni postupak je to znak da smo blizu rešenja i da treba da pređemo na predefinisane korake.

Line Search iz glave 2.2.9 ima u sebi ugrađenu i zaštitu od greške zaokruživanja koja može dovesti do zaglavljivanja iterativnog postupka, kada se generišu jako kratki koraci, vraća $retcode = 1$.

Takođe, greška zaokruživanja retko, a šum često, u stohastičkom okruženju dovode do sledeće greške: Računski se dobije negativna diskriminanta u nalaženju minimuma kubne interpolacije aproksimacije funkcije duž pravca pretraživanja. Ovu situaciju ćemo obraditi isto kao i situaciju koja dovodi do $retcode = 1$ iako će ovde povratni kod biti $retcode = 3$. Obrada je sledeća:

Stohastički iterativni postupak kada Line Search vrati kodove 1 ili 3 će se ponašati kao da su to posledice slučajnosti, pa će ostati u prethodnoj iteraciji.

Očekuje se da će ga sledeća realizacija šuma izvesti iz trenutne tačke.

Kada se pređe na predefinisane korake (SA korake) koristiće se standardni oblik

$$a_k = \frac{a}{(k+1+A)^\alpha}\,,\ \ k = 0, 1, \ldots \tag{4.14}$$

čije parametre ćemo fino podesiti sa $a = 1$, $A = 0$, $\alpha = 0.602$, kao što je preporučeno u [55].

Pseudo kod algoritma DSLS je dat u Algoritmu 6.

Primena kvazi Njutnovih pravaca koji u sebi nose informaciju drugog reda, zahvaljujući aproksimaciji Hesijana B_k u k-toj iteraciji, značajno može ubrzati iterativni postupak. Testiraćemo dve aproksimacije Hesijana: BFGS i SR1, obe će pravac dati formulom

$$d_k = -B_k^{-1} G_k.$$

Primetimo da kada se koriste negativni gradijenti ista formula važi ako se uzme da je matrica B_k jedinična matrica u svim iteracijama.

Algoritam DSLS je dat modularno, u smislu da važi za sva tri slučaja, sa tim što se za proceduru updateQN uzima algoritam updateBFGS za BFGS pravce, updateSR1 za SR1 pravce, a da se linija algoritma updateQN ignoriše za pravce negativnih gradijenata.

Za dato x_k, x_{k+1} BFGS formula za ažuriranje pravca je

$$B_{k+1} = B_k - \frac{B_k \delta_k \delta_k^T B_k}{\delta_k^T B_k \delta_k} + \frac{\Delta_k \Delta_k^T}{\Delta_k^T \delta_k}, \tag{4.15}$$

gde su

$$\delta_k = x_{k+1} - x_k,\ \Delta_k = G_{k+1} - G_k,$$

a SR1 formula je

$$B_{k+1} = B_k + \frac{(\Delta_k - B_k \delta_k)(\Delta_k - B_k \delta_k)^T}{(\Delta_k - B_k \delta_k)^T \delta_k}. \tag{4.16}$$

Pseudo kod algoritama koji vrše odgovarajuća ažuriranja je dat u Algoritmu 7 i Algoritmu 8 redom. Obe formule sa dobrim linijskim pretraživanjem daju super linearnu konvergenciju.

U stohastičkom okruženju je za oba ažuriranja važna veličina

$$\Delta_k = G(x_{k+1}, \varepsilon_{k+1}) - G(x_k, \varepsilon_k)$$

koja se može značajno razlikovati od željene nepoznate determinističke veličine zbog uticaja šuma. Jedan mogući prilaz ovom problemu je predložen u Spall [54]: Ažuriranje se vrši sa težinama koje bi trebalo da prigušuju šum.

Mi ćemo prihvatiti prilaz koji je uspešno testiran u [53]. Ideja je da se razlika gradijenata Δ_k izračuna koristeći isti šum. Na taj način nismo u mogućnosti da recikliramo gradijent izračunat u prethodnoj iteraciji, pa se duplira broj računanja gradijenta. Naša testiranja su pokazala da su se dodatna računanja gradijenta isplatila. Tako da u oba ažuriranja, BFGS i SR1 veličinu Δ_k dobijamo po formuli

$$\Delta_k = G(x_{k+1}, \varepsilon_k) - G(x_k, \varepsilon_k). \tag{4.17}$$

Ulazni parametri Algoritma 6 DSLS su početna iteracija x_0, definicija funkcije cilja F, definicija njenog gradijent G, maksimalan broj iteracija $itnlimit$, maksimalan broj računanja funkcije $maxfcalc$, granica gradijenta za izlazni kriterijum $gradtol$.

Izlazni parametri su y vrednost poslednje iteracije, G_y vrednost gradijenta u poslednjoj iteraciji, $termcode$ kod koji objašnjava šta je bio prvi izlazni kriterijum koji je iteracija zadovoljila. Vrednosti za $termcode$ su

$0 =$ izračunato je $itnlimit$ iteracija,

$1 =$ u poslednjoj iteraciji je dostignut izlazni kriterijum $\|G_k\| \leq gradtol$,

$2 =$ potrošena je kvota $maxfcalc$ za računanje funkcija i gradijenta.

Kao što je već rečeno, parametar $maxfcalcls$ koji određuje kvotu računanje funkcije cilja u Line Search proceduri, zapravo određuje tačku prelaska na SA korake. Parameter $near$ u DSLS algoritmu je $false$ dok se koraci nalaze primenom linijskog pretraživanja. Jednom kada postane $true$ jer je u Line Search potrošena kvota $maxfcalcls$ računanja funkcije cilja, ostaje do kraja $true$, a to odgovara $phase = 2$ u Koraku 2.2. algoritma iz glave 4.1.2.

U odsustvu šuma BFGS aproksimacija Hesijana generiše opadajuće pravce i ima samoispravljajuću osobinu: kroz iteracije popravlja lošu aproksimaciju. U stohastičkom okruženju moramo dodati meru opreza: preskakanje ažuriranja aproksimacije Hesijana ako je $|\Delta_k \delta_k| < safeguard$ za BFGS, odnosno ako je $|\Delta_k \delta_k| < safeguard|\delta|\,|B\delta|$ za SR1 aproksimaciju. Veličina $safeguard$ parametra se podešava na $eps^{1/4} \approx 1.2E-4$, gde je eps mašinska preciznost double IEEE preciznosti.

Algoritam 6 DSLS

Input: $x_0 \in \mathbb{R}^n$, $F : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^n$, $itnlimit \in \mathbb{Z}$, $maxfcalcls \in \mathbb{Z}$, $gradtol \in \mathbb{R}$
Output: $y \in \mathbb{R}^n$, $G_y \in \mathbb{R}^n$, $termcode \in \mathbb{Z}$

```
B := I;                                              ▷ Za prvu iteraciju QN
termcode := 0;
[G_x, z_x] := G(x);                                  ▷ zapamti šum z_x
near := false;                                       ▷ staviti true za SA korak
if ¬near then F_x := F(x); end if
for itncount := 1 to itnlimit do
    d := −B^{-1} G_x;
    if ¬near then
        [y, F_y, retcode, λ, fcalcls] :=
                Line Search (x, F_x, F(·), G_x, −G_x, maxfcalcls);
        if retcode = 2 then                          ▷ fcalcls ≥ maxfcalcls
            near := true;
            λ := a / (itncount + 1 + A)^α;
        else
            if (retcode = 1)||(retcode = 3) then λ := λ_minfix; end if
        end if
    else
        λ := a / (itncount + 1 + A)^α;
    end if
    y := x + λ ∗ d;
    [G_y, z_y] := G(y);                              ▷ zapamti šum z_y
    if |G_y| ≤ gradtol then
        termcode := 1; return
    else
        if number_of_function_calculations > maxfcalc then
            termcode := 2; return
        end if
    end if
    updateQN; x := y; G_x := G_y; z_x := z_y;
    if ¬near then F_x := F_y; end if
end for
return
```

Algoritam 7 updateBFGS (koristi globalne promenljive iz DSLS)

$G'_y := G(y, z_x)$; ▷ Gradijent sa šumom z_x
$\delta := y - x$; $\Delta := G'_y - G_x$;
if $|\Delta^T \delta| \geq safeguard$ **then** ▷ Bez deljenja nulom
  if $itncount = 1$ **then**
    $B := \dfrac{\Delta\Delta^T}{\Delta^T \delta} I$;
  else
    $B := B - \dfrac{B\delta\delta^T B}{\delta^T B\delta} + \dfrac{\Delta\Delta^T}{\Delta^T \delta}$;
  end if
end if
return

Algoritam 8 updateSR1 (koristi globalne promenljive iz DSLS)

$G'_y := G(y, z_x)$; ▷ Gradijent sa šumom z_x
$\delta := y - x$; $\Delta := G'_y - G_x$;
if $|\Delta^T \delta| \geq safeguard|\delta|\,|B\delta|$ **then** ▷ Bez deljenja nulom
  $B := B + \dfrac{(\Delta - B\delta)(\Delta - B\delta)^T}{(\Delta - B\delta)^T \delta}$;
end if
return

Glava 5

Fiksni Njutnov metod za model ekvilibriuma

Velika je lista problema u kojima se može primeniti stohastička optimizacija. Otkako su zakoni ekonomije iskazani formulama i definisan problem ekvilibriuma, matematička optimizacija se primenjuje za rešavanje problema u ekonomiji.

Razvoj računarske tehnike je omogućio primenu simulacija i stohastičke analize kojima se podaci mogu preciznije modelirati.

U ovom poglavlju ćemo razmotriti primenu Fiksnog Njutnovog metoda na model ekvilibriuma. Prvo ćemo definisati iterativni postupak prilagođen stohastičkoj prirodi problema i razmotriti uslove za konvergenciju. Zatim ćemo analizirati primenu tog postupka na model ekvilibriuma.

Rezultat koji sledi u ovoj glavi je originalni doprinos i objavljen je u [37].

5.1 Problem

Posmatramo problem

$$F(x,W) = 0, \tag{5.1}$$

gde je $x \in \mathbb{R}^n$ nepoznati vektor, W je $m-$dimenzionalni vektor parametara i $F(\cdot,W) : \mathbb{R}^n \to \mathbb{R}^n$ za svako W.

Najjednostavniji slučaj problema ovog tipa se pojavljuje kada je $W \in \mathbb{R}^m$ konstantan vektor. Mnogi problemi u ekonomiji se mogu postaviti u obliku (5.1). U praktičnim problemima se često ne mogu tačno odrediti vrednosti parametara W, već je W slučajni vektor.

Kada je F nelinearna, i kad je n veliko, rešavanje (5.1) je komplikovano. Da bi se pojednostavilo, često se W zamenjuje konstantnim vektorom, najčešće vektorom očekivanih vrednosti. To pojednostavljeno modelovanje uvodi nove nesigurnosti u model. Jer, važne informacije koje se mogu dobiti iz realnih podataka se gube i model postaje manje kvalitetan.

Prirodno se javlja potreba za nalaženja dovoljno dobrih procedura za rešavanje (5.1) za opštiji slučaj W od konstantnog.

Ako se u problemu (5.1) W oceni slučajnim vektorom sa raspodelom nad nekim prostorom verovatnoće $(\Omega, \mathscr{F}, P)$, taj model će sigurno sadržati više informacija od modela sa konstatnim W. Problem (5.1) tada postaje stohastički problem, a njegovo rešenje slučajni vektor $\hat{x}^*$. Analitičko rešenje nije poznato pa se pojavljuje potreba za pristupačnim numeričkim rešenjem.

Rešavanje stohastičkih nelinearnih sistema numerički je složeno. Za konstantne vrednosti W postoji niz metoda: recimo Marcotte i ostali [44], Esteban-Bravo [24, 25]. To su metode koje rešavaju deterministički problem (5.1) zasnovane na metodama unutrašnje tačke (interrior point) ili homotopiji. Ovde ćemo predložiti drugačiji pristup rešavanju, sa ciljem da ne izgubimo kvalitet modela.

Ako je W slučajni vektor $W : \Omega \to \bar{\Omega} \subseteq \mathbb{R}^m$ nad posmatranim prostorom verovatnoće $(\Omega, \mathscr{F}, P)$ sa poznatom raspodelom, čiji realizovani prost slučajni uzorak je $(w^1, w^2, \ldots, w^N)$, onda se može rešavati niz problema

$$F(x, w^i) = 0, \; i = 1, 2, \ldots, N, \tag{5.2}$$

za koje je niz rešenja $(x^{1*}, \ldots, x^{N*}) \in \mathbb{R}^{n \times N}$. Za dovoljno veliko N taj niz rešenja treba da omogući bolji uvid u ponašanje rešenja $\hat{x}^*$ problema (5.1).

Rešavanje velikog broja nelinearnih problema može biti vremenski zahtevno. Zato predlažemo numeričku proceduru koja se zasniva na fiksnim Njutnovim pravcima i daje rešenje problema (5.1) uz prihvatljivu računarsku složenost.

Za problem (5.1) ćemo postaviti standardni niz pretpostavki. Neka je $D \subseteq \mathbb{R}^n$ domen nad kojim važi:

(D1) Za sve $w \in \bar{\Omega}$ postoji $x^{w*} \in D$ za koje je $F(x^{w*}, w) = 0$.

(D2) Za sve $w \in \bar{\Omega}$ matrica Jakobijana $F'(x^{w*}, w)$ je nesingularna.

(D3) Za sve $x, y \in D$ i $w \in \bar{\Omega}$ postoji $\gamma > 0$ za koje je

$$\|F'(x, w) - F'(y, w)\| \leq \gamma \|x - y\| .$$

(D4) Za sve $w_1, w_2 \in \bar{\Omega}$ i $x \in D$ postoji $\gamma_W > 0$ za koje

$$\|F'(x, w_i) - F'(x, w_j)\| \leq \gamma_W \|w_i - w_j\|.$$

Matrica Jakobijana $F'(x,w)$ za

$$F(x,w) = \begin{bmatrix} F_1(x_1,\ldots,x_n,w_1,\ldots,w_m) \\ \vdots \\ F_n(x_1,\ldots,x_n,w_1,\ldots,w_m) \end{bmatrix} \text{ je } F'(x,w) = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_n} \\ & \ddots & \\ \frac{\partial F_n}{\partial x_1} & \cdots & \frac{\partial F_n}{\partial x_n} \end{bmatrix}.$$

Njutnova metoda je jedan izbor za rešavanje sistema (5.2). To je pouzdana metoda koja daje kvadratnu konvergenciju za dobru početnu vrednost. Njutnov iterativni niz za dato $x^{i,0}$ definisan jednakostima

$$F'(x^{i,k},w^i)\,s^{i,k} = -F(x^{i,k},w^i), \quad x^{i,k+1} = x^{i,k} + s^{i,k}, \; k = 1,2,\ldots \tag{5.3}$$

će konvergirati rešenju x^{i*} problema (5.2) ako $F(x,w^i), i = 1,2,\ldots N$ zadovoljava pretpostavke D1 - D3. Ipak, ta procedura je računski skupa za veliko N i srednje veliku dimenziju problema n.

Ako se pogledaju Jakobijani $F'(x^{i,k},w^i), i = 1,2,\ldots N$ može se videti da će imati sličnu strukturu za razne realizacije w^i. Iterativni postupak bi to mogao iskoristiti. Prirodno se nameće upotreba procedure zasnovane na Fiksnom Njutnovom postupku. Ideja ovog postupka je da se Jakobijan izračuna jednom i da se ne menja tokom iteracija. Fiksni Njutnov postupak je konvergentan ako je početna vrednost dovoljno blizu rešenja. Predlažemo Algoritam FNM (Algoritam 9) za rešavanje problema (5.2).

Algoritam 9 FNM

Korak 0 Neka je dato $x^0 \in \mathbb{R}^n$ i $(w^1,\ldots,w^N)$. Izračunati srednju vrednost $\bar{w} = \sum_{i=1}^{N} w^i/N$ i rešiti $F(x,\bar{w}) = 0$ Njutnovom metodom. Označimo rešenje $x^{\bar{w}}$ i definišimo $A = F'(x^{\bar{w}},\bar{w})$.

Korak 1 Za $i = 1,\ldots,N$

Korak 1i Postavi $x^{i,0} = x^{\bar{w}}$

Korak 2i Ponavljaj do konvergencije

$$As^{i,k} = -F(x^{i,k},w^i),\; x^{i,k+1} = x^{i,k} + s^{i,k},\; k = k+1$$

Korak 3i Postavi $x^{i*} = x^{i,k}$

Napomena. Matrica korišćena u Algoritmu FNM $A = F'(x^{\bar{w}}, \bar{w})$ je jedan mogući izbor konstantne matrice iz koraka 2i. Većina računskih operacija u (5.3) se troši na računanje matrice Jakobijana u iteracijama za sve elemente uzorka w^i a potom na rešavanje odgovarajućih linearnih sistema u iterativnom koraku. Stoga je glavna prednost predloženog algoritma to što se svi sistemi u iteracijama rešavaju sa istom matricom sistema A. Pri tome A treba da je dovoljno dobra aproksimacija tačnog Jakobijana.

Što se tiče uštede u rešavanju sistema, može se koristiti faktorisana matrica iz prvog koraka u svakom koraku 2i donoseći dodatnu uštedu na broju računskih operacija.

Matrica A bi trebala biti dovoljno dobra aproksimacija $F'(x, w^i)$, za x dovoljno blizu x^{i*}. Testiranja koja smo izvršili nisu pokazala veliku razliku kada se $F'(x, w^i)$ aproksimira vrednošću $A = F'(x^0, w^j)$, za neko drugo $j \in \{1, 2, \ldots, N\}$.

Algoritam 9 kao rezultat daje niz $(x^{1*}, \ldots, x^{N*})$ ako je korak 2i dobro definisan, odnosno, ako iterativni postupak 2i za svako $i = 1, 2, \ldots, N$ završi sa rešenjem x^{i*}. Dokazaćemo da je to tačno da pod pretpostavkama D1 - D4.

Bitna prednost ovog algoritma je što se može parelelizovati ([2]).

Teorema koja sledi se zasniva na Banahovom principu kontrakcije. Uslovi za konvergenciju se mogu shvatiti kao kombinacija uslova o udaljenosti početne tačke od rešenja i varijanse parametara W. Uslovi možda izgledaju strogi, ali su numerički rezultati pokazali veliku efikasnost algoritma.

Označimo sa $\mathscr{B} = \mathscr{B}(x, \delta) = \{y \in \mathbb{R}^n : \|x - y\| \leq \delta\}$ zatvorenu loptu oko x poluprečnika r. Pod pretpostavkom da važe D1, D2, D3, lako se vidi da je Njutnov metod za rešavanje $F(x, \bar{w}) = 0$ lokalno konvergentan. Dakle, postoji $\delta_1 > 0$ takvo da za $x^0 \in \mathscr{B}(x^{\bar{w}*}, \delta_1)$ Korak 0 jeste dobro definisan i da daje niz koji konvergira. Neka je $x^{\bar{w}}$ vrednost dobijena posle s iteracija Njutnovog metoda za s koje daje $x^{\bar{w}}$ dovoljno blizu $x^{\bar{w}*}$. Onda je matrica $F'(x^{\bar{w}}, \bar{w})$ invertibilna i postoji $M > 0$ takvo da $\|F'(x^{\bar{w}}, \bar{w})^{-1}\| \leq M$.

Sledeća teorema daje konvergenciju Algoritma FNM.

TEOREMA 36 *Neka F zadovoljava pretpostavke D1 - D4. Ako za $\varepsilon_W = \mathrm{diam}(\bar{\Omega})$ i $M = \|F'(x^{\bar{w}}, \bar{w})^{-1}\|$ postoji $\delta > 0$ takvo da je $\alpha = M(\gamma\delta + \gamma_W \varepsilon_W) < 1$ i da je $\|F'(x^{\bar{w}}, \bar{w})^{-1} F(x^{\bar{w}}, w^i)\| \leq \delta(1 - \alpha)$, onda za sve $w^i \in \bar{\Omega}$ niz $\{x^{w^i,k}\}_{k=0}^{\infty}, i = 1, 2, \ldots, N$ definisan u Koraku 1 Algoritma FNM konvergira linearno ka rešenju $F(x, w^i) = 0$.*

Dokaz. Neka $G(x) = x - F'(x^{\bar{w}}, \bar{w})^{-1} F(x, w^i)$. Onda za sve $x \in \mathscr{B}$

$$\begin{aligned}
\|G'(x)\| &= \|I - F'(x^{\bar{w}}, \bar{w})^{-1} F(x, w^i)\| \\
&\leq \|F'(x^{\bar{w}}, \bar{w})^{-1}\| \big(\|F'(x^{\bar{w}}, \bar{w}) - F'(x, \bar{w})\| + \\
&\quad + \|F'(x, \bar{w}) - F'(x, w^i)\|\big) \\
&\leq M(\gamma \|x^{\bar{w}} - x\| + \gamma_W \|\bar{w} - w^i\|) \\
&\leq M(\gamma\delta + \gamma_W \varepsilon_W) = \alpha < 1.
\end{aligned}$$

Na osnovu teoreme o srednjoj vrednosti dobijamo

$$\|G(x) - G(y)\| \leq \sup_{t \in [0,1]} \|G'(x + t(y - x))\| \cdot \|x - y\| \leq \alpha \|x - y\|, \quad \forall x, y \in \mathscr{B}.$$

Sledi da je G kontrakcija na $\mathscr{B}$. Pored toga

$$\|G(x) - x^{\bar{w}}\| \leq \|G(x) - G(x^{\bar{w}})\| + \|G(x^{\bar{w}}) - x^{\bar{w}}\| \leq \alpha\delta + \delta(1 - \alpha) = \delta\,,$$

G preslikava $\mathscr{B}$ na samog sebe. Na osnovu Banahovog principa kontrakcije sledi da postoji jedinstvena nepokretna tačka u $\mathscr{B}$ koja je rešenje $F(x, w^i) = 0$ nad $\mathscr{B}$.

■

5.2 Modeli ekvilibriuma

5.2.1 Neoklasična ekonomija

Model neoklasične ekonomije je jedan od najstarijih i najviše korišćenih ekonomskih modela. Kreirao ga je Walras, kasnije su ga proširili i formalizovali Arrow i Debreu (vidi Esteban-Bravo [25], Scarf [52]). Model neoklasične ekonomije je često analiziran (vidi de la Fuente, [18], Judd [33]). Jedan pregled modela ekvilibriuma se može naći u Esteban-Bravo [24]. Najčešći pristupi rešavanju problema ekvilibriuma [24] su: primena simpleksa u konstruktivnom dokazu Brouwerove teoreme o nepokretnoj tački, Njutnov metod, Smaleov metod, metode unutrašnje tačke. Takođe, metod homotopije imamo analiziran u Eaves, [23].

Neka je $\mathbb{R}^n_+$ skup n-torki nenegativnih realnih brojeva. Posmatraćemo model koji sadrži n roba iz $X \subset \mathbb{R}^n_+$ i m agenata. Agenti imaju funkcije korisnosti u^j i početno zaduženje ω^j za $j = 1,2,\ldots,m$. Budžet svakog agenta je $\pi\omega^j$, gde je $\pi \in \mathbb{R}^n_+$ vektor cena robe. Izbor funkcije korisnosti definiše model.

Posmatraćemo tri izbora funkcije korisnosti.

- Cobb-Douglas: $u(x) = x_1^{a_1} \cdots x_n^{a_n}, \quad \sum_{i=1}^{n} a_i = 1, \; a_i \geq 0$
- fiksne proporcije: $u(x) = \min\left\{\frac{x_1}{a_1}, \cdots, \frac{x_n}{a_n}\right\}, \; a_i > 0$
- CES: $u(x) = \left(\sum_{i=1}^{n} a_i^{1/b} x_i^{(b-1)/b}\right)^{b/(b-1)}, \quad a_i > 0, \; b \leq 1$

U svakom od ova tri slučaja skup parametara $a_1, a_2, \ldots, a_n, b$ treba da se oceni. Modeli posmatrani u [24] podrazumevaju da su parametri konstantni, dok mi posmatramo opštiji slučaj u kome su parametri slučajne promenljive.

Cilj svakog agenta je da maksimizuje funkciju korisnosti, tj. da pronađe

$$x^j(\pi) = \arg \max_{\pi x \leq \pi\omega^j} u^j(x), \quad j = 1,2,\ldots,m.$$

Za tri funkcije korisnosti koje posmatramo maksimumi su:

- Cobb-Douglas: $x_i^j(\pi) = \dfrac{\pi \cdot \omega^j}{\pi_i} a_i^j, \; i = 1,2,\ldots,n$
- fiksne proporcije: $x_i^j(\pi) = \dfrac{\pi \cdot \omega^j}{\pi \cdot a^j} a_i^j, \; i = 1,2,\ldots,n$
- CES: $x_i^j(\pi) = \dfrac{\pi \cdot \omega^j}{\pi_i^{b^j} \sum_{k=1}^{n} \pi_k^{1-b^j} a_k^j} a_i^j, \; i = 1,2,\ldots,n.$

Ako se funkcija viškova definiše

$$\xi(\pi) = \sum_{j=1}^{m} x^j(\pi) - \sum_{j=1}^{m} \omega^j,$$

pozitivno π^* za koje je

$$\xi(\pi^*) = 0 \tag{5.4}$$

je **ekvilibrium**.

5.2.2 Model prostornog ekvilibriuma

Modeli prostornog ekvilibriuma su klasa problema ekvilibriuma gde su ponuda, potražnja i cene prevoza nelinearne i odvojive, Marcotte i ostali [44]. Karakteristična osobina ovih modela je uračunavanje cena skladištenja i prevoza za svaku transakciju između proizvođača i potrošača.

Model koji posmatramo sadrži jednu ili više roba, skup proizvođača i skup potrošača. Koristićemo sledeći zapis:

- Y_i - ponuda i-tog proizvođača, $i \in I = \{1, \dots, m\}$
- Z_i - tražnja j-tog potrošača, $j \in J = \{1, \dots, n\}$
- x_{ij} - količina robe transportovana (trgovana) između i-tog proizvođača i j-tog potrošača, $i \in I,\ j \in J$
- $S_i(Y_i)$ - inverzna funkcija ponude i-tog proizvođača
- $I_j(Z_j)$ - inverzna funkcija potražnje j-tog potrošača
- $c_{ij}(x_{ij})$ - cena transporta jedinice robe između i-tog proizvođača i j-tog potrošača.

Ekvilibrium je trojka $\{(Y_i)_{i\in I}, (Z_j)_{j\in J}, (x_{ij})_{i\in I,\, j\in J}\}$ koja zadovoljava sistem nejednačina

$$
\begin{aligned}
S_i(Y_i) + c_{ij}(x_{ij}) &\begin{cases} = I_j(Z_j), & \text{if } x_{ij} > 0 \\ \geq I_j(Z_j), & \text{if } x_{ij} = 0 \end{cases} \\
Y_i &= \sum_{j\in J} x_{ij}, \quad \forall i \in I \\
Z_j &= \sum_{i\in I} x_{ij}, \quad \forall j \in J \\
x_{ij} &\geq 0, \quad \forall i, j \in I \times J.
\end{aligned} \tag{5.5}
$$

Ovaj model se posmatra uz pretpostavke:

M1 Ponuda, potražnja i cena transporta su nenegativne neprekidno diferencijabilne funkcije (klase $C^1(\mathbb{R}_+)$). Ponuda i cena prevoza su neopadajuće funkcije, potražnja je nerastuća funkcija.

M2 $$\lim_{Y_i \to \infty} S_i(Y_i) = +\infty, \quad \lim_{Z_i \to \infty} I_j(Z_j) = 0$$

Ako važe M1 i M2, onda su uslovi ekvilibriuma (5.5) ekvivalentni KKT uslovima za problem minimizacije

$$\min_{x \geq 0} G(x) = \sum_{i \in I} \int_0^{\sum_{j \in J} x_{ij}} S(t)\, dt + \sum_{i \in I} \sum_{j \in J} \int_0^{x_{ij}} c_{ij}(t)\, dt - \sum_{j \in J} \int_0^{\sum_{i \in I} x_{ij}} I(t)\, dt\,.$$

Ponuda, potražnja i cena transporta su parametri modela koji se zadaju, dok su x_{ij} nepoznate koje računamo da bismo našli ekvilibrium.

Ako je neka količina transporta $x_{ij} = 0$, odgovarajuća nejednakost u (5.5) je striktna. Štaviše, možemo, kao u [44], pretpostaviti da važe hipoteze stroge komplementarnosti:

$$x_{ij}^* = 0 \Leftrightarrow S_i(Y_i^*) + c_{ij}(x_{ij}^*) - I_j(Z_j^*) \neq 0$$

za rešenje (x^*, Y^*, Z^*).

Ako stroga komplementarnost ne važi, rešenje je degenerisano. Najčešći slučajevi degenerisanog rešenja su kada su funkcije troškova transporta $c_{ij}(x_{ij})$ konstantne.

Dakle, ako pretpostavimo da hipoteza komplementarnosti (koja daje nedegenerativnost) važi, treba nam strategija za određivanje skupa indeksa $\mathscr{P}$ sa nenula vrednostima transporta. Potom nam od (5.5) ostaje da rešimo sistem jednačina.

U našim primerima smo pretpostavili da su svi $x_{ij} > 0$, odnosno $\mathscr{P} = I \times J$.

Ako označimo sa x vektor čije su komponente $x_{ij}, i \in I, j \in J$ i sa W parametre modela dobijamo da je (5.5) ekvivalentno sa $F(x, W) = 0$. Na ovaj sistem jednačina ćemo primeniti FNM algoritam. Rezultati su predstavljeni u glavi 7.

Deo III

Numerički rezultati

Glava 6

Testiranje DSLS algoritma

Posmatramo problem optimizacije bez ograničenja

$$\text{minimize } f(x),$$

za funkciju cilja $f : \mathbb{R}^n \to \mathbb{R}$, x je n-dimenzionalni vektor $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$.

Pretpostavićemo da raspolažemo samo sa vrednostima funkcije $f(x)$ sa šumom i gradijenta $\nabla f(x) = g(x)$ sa šumom.

Za $x \in \mathbb{R}^n$, neka su $\xi(x)$ i $\varepsilon(x)$ redom slučajna promenljiva i slučajni vektor definisani u prostoru verovatnoće $(\Omega, \mathscr{F}, P)$. Vrednosti funkcije i gradijenta sa šumom su

$$F(x) = f(x) + \xi(x) \quad \text{i} \quad G(x) = g(x) + \varepsilon(x),$$

gde su ξ and ε slučajni šum.

Mi ćemo u ovom istraživanju isprobati dve vrste šuma sa tri veličine standardnog odstupanja šuma. Za izbor funkcije $f : \mathbb{R}^n \to \mathbb{R}$ na koju ćemo dodavati šum posegnućemo u biblioteke standardnih test problema za probleme minimizacije bez ograničenja.

Funkcija cilja f najčešće predstavlja sumu kvadrata odstupanja koja se dobija u metodi najmanjih kvadrata za nelinearni model:

$$\text{minimize } f(x) = \sum_{i=1}^{m} f_i^2(x), \tag{6.1}$$

gde su funkcije $f_1, f_2, \ldots, f_m : \mathbb{R}^n \to \mathbb{R}$.

Ako su za neki problem vrednosti n i m proizvoljne, onda ćemo pored vrednosti n i m koju smo uzeli u zagradi napisati dopustive vrednosti.

6.1 Spisak problema za stohastičku optimizaciju

Problemi koje ćemo ovde navesti su klasični problemi za testiranje algoritama za optimizaciju. Neki od njih su veštački kreirani da bi onemogućili nedovoljno robustne algoritme da ih reše.

Skoro svi problemi su bili zadavani za testiranje determinističke optimizacije i na njih je dodat šum, najčešće kao slučajna promenljiva sa normalnom raspodelom i očekivanjem nula.

U nekim problemima šum je zavisio od argumenta funkcije, u nekim je veličina šuma bila nezavisna od argumenta.

Problemi i način dodavanja šuma su uzeti slično kao u drugim radovima iz ove oblasti.

Osim n i m vrednosti koje su gore opisane, za svaki problem ćemo navesti i način dodavanja šuma $\xi(x)$, početnu iteraciju x^0 i optimum x^* ako je poznat.

1. Spall's polynomial, [54]

$n = 4$

$$f(x) = x^T A^T Ax + 0.1\sum_{i=1}^{n}(Ax)_i^3 + 0.01\sum_{i=1}^{n}(Ax)_i^4$$

gde je A gornja trougaona matrica popunjena jedinicama.

Šum:

$$\xi(x) = [x^T, 1]\,\hat{z}_1, \text{ gde je } \hat{z}_1 \sim \mathcal{N}(0, \sigma^2 I_{(n+1)\times(n+1)}),$$

$$\varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = 0.2(1,1,\ldots,1)$

Optimum: $x^* = (0,0,\ldots 0)$, $f(x^*) = 0$

2. QuadAppr, [42]

$n = 2$

$$f(x) = 200\,\phi_b(x_1,x_2)/\phi_b(10,10)$$

gde je

$$\phi_b(x_1,x_2) = \exp\left(0.1\sqrt{x_1^2 + bx_2^2}\right) + \exp\left(-0.1\sqrt{x_1^2 + bx_2^2}\right) - 2.$$

Šum:

$$\xi(x) = [x^T, 1]\, \hat{z}_1, \text{ gde je } \hat{z}_1 \sim \mathcal{N}(0, \sigma^2 I_{(n+1)\times(n+1)}),$$

$$\varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = 0.5(1,1)$
Optimum: $x^* = (0,0)$, $f(x^*) = 0$ $(b = 1)$

3. Fletcher-Powell helical valley function, [28, 46]

$n = 3$

$$m = 3$$

$$f_1(x) = 10[x_3 - 10\,\theta(x_1, x_2)]$$

$$f_2(x) = 10(\sqrt{x_1^2 + x_2^2} - 1)$$

$$f_3(x) = x_3$$

gde je

$$\theta(x_1, x_2) = \begin{cases} \frac{1}{2\pi} \arctan \frac{x_2}{x_1}, & x_1 > 0 \\ \frac{1}{2} + \frac{1}{2\pi} \arctan \frac{x_2}{x_1}, & x_1 < 0. \end{cases}$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2),\ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = (-1, 0, 0)$
Optimal: $x^* = (1, 0, 0)$, $f(x^*) = 0$ $(b = 1)$.

4. Biggs EXP6 function, [5, 46]

$n = 6$

$$m = 13\ (m \geq n)$$

$$f_i(x) = x_3 \exp(-t_i x_1) - x_4 \exp(-t_i x_2) + x_6 \exp(-t_i x_5) - y_i,$$

gde je

$$t_i = i/10,\ y_i = \exp(-t_i) - 5\exp(-10t_i) + 3\exp(-4t_i),$$

za $i = 1,2,\ldots,m$.

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2),\ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (1,2,1,1,1,1)$

Optimum:

lokalni: $f(x^*) = 5.65565 \cdot 10^{-3}$

globalni: $x^* = (1,10,1,5,4,3)$, $f(x^*) = 0$.

5. Gaussian function, [46]

$n = 3$

$$m = 15$$

$$f_i(x) = x_1 \exp\left(\frac{-x_2(t_i - x_3)^2}{2}\right) - y_i,$$

gde je

$$t_i = (8-i)/2$$

za sve $i = 1,2,\ldots,m$ i

i	1,15	2,14	3,13	4,12	5,11	6,10	7,9	8
y_i	0.0009	0.0044	0.0175	0.0540	0.1295	0.2420	0.3521	0.3989

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2),\ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (0.4,1,0)$

Optimum: $f(x^*) = 1.12793 \cdot 10^{-8}$.

6. Box three-dimensional function, [8, 46]

$n = 3$

$$m = 10\ (m \geq n)$$

$$f_i(x) = \exp(-t_i x_1) - \exp(-t_i x_2) - x_3\left(\exp(-t_i) - \exp(-10t_i)\right),$$

gde je

$$t_i = i/10$$

za sve $i = 1, 2, \ldots, m$.

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = (0, 10, 5)$
Optimum: $f(x^*) = 0$ za $(1, 10, 1)$, $(10, 1, -1)$ i za ($x_1 = x_2$ i $x_3 = 0$).

7. Variably dimensioned function, [46]

$n = 4$

$$m = n + 2 = 6$$

$$f_i(x) = x_i - 1, i = 1, 2, \ldots, n,$$

$$f_{n+1}(x) = \sum_{j=1}^{n} j(x_j - 1)$$

$$f_{n+2}(x) = \left(\sum_{j=1}^{n} j(x_j - 1)\right)^2.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = (1 - \frac{1}{n}, 1 - \frac{2}{n}, \ldots, 0)$.
Optimum: $x^* = (1, 1, \ldots, 1)$, $f(x^*) = 0$.

8. Watson function, [36, 46]

$n = 4$ $(2 \leq n \leq 31)$

$$m = 31$$

$$f_i(x) = \sum_{j=2}^{n}(j-1)x_j t_i^{j-2} - \left(\sum_{j=1}^{n} x_j t_i^{j-1}\right)^2 - 1, i = 1,2,\dots,29,$$

$$f_{30}(x) = x_1,\ f_{31}(x) = x_2 - x_1^2 - 1.$$

Šum:

$$\xi(x) = \ z,\ \text{gde je}\ z \sim \mathscr{N}(0,\sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je}\ \hat{z} \sim \mathscr{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (0,0,\dots,0)$.
Optimum: $f(x^*) = 2.4384 \cdot 10^{-6}$.

9. Penalty Function #1, [29, 46]

$n = 10$ $(n \in \mathbb{N})$

$$m = n+1 = 11$$

$$f_i(x) = \sqrt{a}\,(x_i - 1), i = 1,2,\dots,n,$$

$$f_{n+1}(x) = -\frac{1}{4} + \sum_{j=1}^{n} x_j^2,$$

$$\text{gde je}\ a = 10^{-5}.$$

Šum:

$$\xi(x) = \ z,\ \text{gde je}\ z \sim \mathscr{N}(0,\sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je}\ \hat{z} \sim \mathscr{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (1,2,\dots,n)$.
Optimum: $f(x^*) = 7.08765 \cdot 10^{-5}$.

10. Penalty function #2, [29, 46]

$n = 4$ ($n \in \mathbb{N}$)

$$m = 2n = 8$$
$$f_1(x) = x_1 - 0.2$$
$$f_i(x) = \sqrt{a}\left(\exp\frac{x_i}{10} + \exp\frac{x_{i-1}}{10} - y_i\right), i = 2,3,\ldots,n,$$
$$f_i(x) = \sqrt{a}\left(\exp\frac{x_{i-n+1}}{10} + \exp\frac{-1}{10}\right), i = n+1, n+2, \ldots, 2n-1,$$
$$f_{2n}(x) = -1 + \sum_{j=1}^{n}(n-j+1)x_j^2,$$
$$\text{gde je } a = 10^{-5}, y_i = \exp\frac{i}{10} + \exp\frac{i-1}{10}.$$

Šum:

$\xi(x) = z$, gde je $z \sim \mathcal{N}(0,\sigma^2)$, $\varepsilon(x) = \hat{z}$, gde je $\hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n})$.

Start: $x^0 = 0.5\,(1,1,\ldots,1)$.
Optimum: $f(x^*) = 9.37629\cdot 10^{-6}$.

11. Gulf research and development function, [16, 46]

$n = 3$ ($n \le m \le 100$)

$$m = 99,\ (n \le m \le 100)$$
$$f_i(x) = \exp\left(-\frac{|y_i - x_2|^{x_3}}{x_1}\right) - t_i,\ i = 1,2,\ldots,n,$$
$$\text{gde je } t_i = 1/100,\ y_i = 25 + (-50\ln(t_i))^{2/3}.$$

Šum:

$\xi(x) = z$, gde je $z \sim \mathcal{N}(0,\sigma^2)$, $\varepsilon(x) = \hat{z}$, gde je $\hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n})$.

Start: $x^0 = (40, 20, 1.2)$.
Optimum: $x^* = (50, 25, 1.5)$, $f(x^*) = 0$.

12. Trigonometric function, [46, 58]

$n = 10$ $(n \in \mathbb{N})$

$$m = n = 10$$

$$f_i(x) = n - \sum_{j=1}^{n} \cos x_j + i(1 - \cos x_i) - \sin x_i,\ i = 1, 2, \ldots, n.$$

Šum:

$$\xi(x) = \ z,\ \text{gde je } z \sim \mathcal{N}(0, \sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = \frac{1}{n}(1, 1, \ldots, 1)$.
Optimum: $f(x^*) = 0$.

13. Extended Rosenbrock parabolic valley function, [46, 58]

$n = 2$ $(n = 2k,\ k \in \mathbb{N})$

$$m = n$$

$$f_{2i-1}(x) = 10(x_{2i} - x_{i-1}^2),\ f_{2i}(x) = 1 - x_{2i-1},\ i = 1, 2, \ldots, n/2.$$

Šum:

$$\xi(x) = \ z,\ \text{gde je } z \sim \mathcal{N}(0, \sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = (-1.2, 1, -1.2, 1, \ldots)$.
Optimum: $x^* = (1, 1, \ldots, 1)$, $f(x^*) = 0$.

14. Extended Powel singular quartic function, [46, 58]

$n = 4$ $(n = 4k,\ k \in \mathbb{N})$

$$m = n$$

$$f_{4i-3}(x) = x_{4i-3} + 10x_{4i-2},\ f_{4i-2}(x) = 5^{1/2}(x_{4i-1} - x_{4i}),$$

$$f_{4i-1}(x) = (x_{4i-2} - 2x_{4i-1})^2,\ f_{4i}(x) = 10^{1/2}(x_{4i-3} - x_{4i})^2,$$
$$i = 1,2,\dots,n/4.$$

Šum:

$$\xi(x) = z,\ \text{gde je } z \sim \mathcal{N}(0,\sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (3,-1,0,1,3,-1,0,1,\dots)$.
Optimum: $x^* = (0,0,\dots,0)$, $f(x^*) = 0$.

15. Beale function, [3, 46]

$n = 2$

$$m = 3$$
$$f_i(x) = y_i - x_1(1 - x_2^i),\ i = 1,2,3,$$
$$y_1 = 1.5,\ y_2 = 2.25,\ y_3 = 2.625.$$

Šum:

$$\xi(x) = z,\ \text{gde je } z \sim \mathcal{N}(0,\sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (1,1)$.
Optimum: $x^* = (3,0.5)$, $f(x^*) = 0$.

16. Wood function, [15, 46]

$n = 4$

$$m = 6$$
$$f_1(x) = 10(x_2 - x_1^2),\ f_2(x) = 1 - x_1,\ f_3(x) = 90^{1/2}(x_4 - x_3^2),$$
$$f_4(x) = 1 - x_3,\ f_5(x) = 10^{1/2}(x_2 + x_4 - 2),\ f_6(x) = 10^{-1/2}(x_2 - x_4).$$

Šum:

$$\xi(x) = z,\ \text{gde je } z \sim \mathcal{N}(0,\sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (-3,-1,-3,-1)$.
Optimum: $x^* = (1,1,1,1)$, $f(x^*) = 0$.

17. Chebyquad function [27, 46]

$n = 10\,(n \in \mathbb{N})$

$$m = n\,(m \geq n)$$

$$f_i(x) = \frac{1}{n}\sum_{j=1}^{n} T_i(x_j) - \int_0^1 T_i(x)dx,$$

gde je T_i i-ti Čebiševljev polinom pomeren na interval $[0,1]$, za koji važi

$$\int_0^1 T_i(x)dx = \begin{cases} 0, & i = 2k+1 \\ \frac{-1}{i^2-1}, & i = 2k, \end{cases} \quad k \in \mathbb{N}.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (1,2,\ldots,n)/(n+1)$.

Optimum: $f(x^*) = 0$ za $m = n, 1 \leq n \leq 7$ i za $n = 9$.
$f(x^*) = 3.51687E-3$ za $m = n = 8$, $f(x^*) = 6.50395E-3$ za $m = n = 10$.

18. Leon cubic valley function [9, 41]

$n = 2$

$$f(x) = 100(x_2 - x_1^3)^2 + (1-x_1)^2.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (-1.2,-1)$.

Optimum: $x^* = (1,1)$, $f(x^*) = 0$.

19. Gregory and Karney tridiagonal matrix function [9, 30]

$n = 4\,(n \in \mathbb{N})$

$$f(x) = x^T Ax,\ \text{gde je } A = \begin{bmatrix} 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 2 \end{bmatrix}$$

Šum:

$$\xi(x) = \ z,\ \text{gde je } z \sim \mathcal{N}(0, \sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = (0, 0, \ldots, 0)$.
Optimum: $x^* = (n, n-1, ..., 1)$, $f(x^*) = -n = -4$,

20. Hilbert matrix function [9]

$n = 4\,(n \in \mathbb{N})$

$$f(x) = x^T Ax,\ \text{gde je } A = \left[a_{i,j}\right],\ a_{i,j} = \frac{1}{i+j-1},\ i, j = 1, 2, \ldots, n.$$

Šum:

$$\xi(x) = \ z,\ \text{gde je } z \sim \mathcal{N}(0, \sigma^2),\ \varepsilon(x) = \hat{z},\ \text{gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^0 = (1, 1, \ldots, 1)$.
Optimum: $x^* = (0, 0, \ldots, 0)$, $f(x^*) = 0$,

21. De Jong function F1, [45], [17]

$n = 3$

$$f(x) = \sum_{i=1}^{3} x_i^2$$

Šum:

$$\xi(x) = \ z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (-5.12, 0, 5.12)$.
Ograničenje: $-5.12 \leq x_i \leq 5.12$, $i = 1,2,3$.
Optimum: $x^* = (0,0,0)$, $f(x^*) = 0$,

22. De Jong function F2, [45], [17]

$n = 2$

$$f(x) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2$$

Šum:

$$\xi(x) = \ z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (-2.048, 2.048)$. Ograničenje: $-2.048 \leq x_i \leq 2.048$, $i = 1,2$.
Optimum: $x^* = (1,1)$, $f(x^*) = 0$,

23. The De Jong function F3, (discontinuous), [45], [17]

$n = 5$

$$f(x) = \sum_{i=1}^{n} \lfloor x_i \rfloor$$

Šum:

$$\xi(x) = \ z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (-5.12, -2.56, 0, 2.56, 5.12)$.
Ograničenje: $-5.12 \leq x_i \leq 5.12$, $i = 1,2,\ldots,5$.
Optimum: $x^* = -5.12(1,1,1,1,1)$, $f(x^*) = -30$.

24. The De Jong function F4, (Gaussian noise), [45], [17]

$n = 30$

$$f(x) = \sum_{i=1}^{n} i\, x_i^4 + \zeta,$$

Originalni šum: $\zeta \sim \mathscr{N}(0,1)$

Šum:

$$\xi(x) = \; z, \text{ gde je } z \sim \mathscr{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathscr{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (-1.2800, -1.1917, \ldots, 1.2800)$.
Ograničenje: $-1.28 \le x_i \le 1.28$, $i = 1, 2, \ldots, 30$.
Optimum: $x^* = (0, 0, \ldots, 0)$, $f(x^*) = 0$.

25. The De Jong function F5, [45], [17]

$n = 2$

$$f(x) = \left(\frac{1}{K} + \sum_{j=1}^{25} \frac{1}{f_j(x)} \right)^{-1},$$

$$K = 500, \quad f_j(x) = c_j + \sum_{i=1}^{2} (x_i - a_{i,j})^6,$$

$$[a_{i,j}] = \begin{bmatrix} -32 & -16 & 0 & 16 & 32 & -32 & -16 & \cdots & 0 & 16 & 32 \\ -32 & -32 & -32 & -32 & -32 & -16 & -16 & \cdots & 32 & 32 & 32 \end{bmatrix},$$

c_j je aproksimacija funkcije između $a_{1,j}$ i $a_{2,j}$.

Šum:

$$\xi(x) = \; z, \text{ gde je } z \sim \mathscr{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathscr{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^0 = (-32.01, -32.02)$.
Ograničenje: $-65.536 \le x_i \le 65.536$, $i = 1, 2$.
Optimum: $x^* = (-32.00, -32.00)$, $f(x^*) = 0$.

26. The Schaffer Function F6 [48]

$n = 2$

$$f(x) = 0.5 + \frac{\sin^2 \sqrt{x_1^2 + x_2^2} - 0.5}{1 + 0.001(x_1^2 + x_2^2)^2}.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathscr{N}(0, \sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathscr{N}(0, \sigma^2 I_{n \times n}).$$

Start: $x^0 = (-5.0, 10.0)$.
Ograničenje: $-10 \leq x_i \leq 10$, $i = 1, 2$.
Optimum: $x^* = (0, 0)$, $f(x^*) = 0$.

27. The Schaffer function F7, [62]

$n = 2$

$$f(x) = \sqrt{x_1^2 + x_2^2} \left(1.0 + \sin^2 \left(50.0 (x_1^2 + x_2^2)^{1/10}\right)\right).$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathscr{N}(0, \sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathscr{N}(0, \sigma^2 I_{n \times n}).$$

Start: $x^0 = (-5.0, 10.0)$.
Optimum: $x^* = (0, 0)$, $f(x^*) = 0$.

28. Goldstein price polynomial, [40]

$n = 5$

$$f(x) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2, \text{ gde je } -2.048 \leq x_i \leq 2.048, \ i = 1, 2.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathscr{N}(0, \sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathscr{N}(0, \sigma^2 I_{n \times n}).$$

Start: $x^0 = (-2.048, 2.048)$.
Optimum: $x^* = (1,1)$, $f(x^*) = 0$,

29. Branin RCOS function

$n = 2$

$$f(x) = \left(x_2 - \frac{5}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos x_1 + 10\,.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2),\ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^* = (-1, 1)$.
Optimum: $x^0 = (-\pi, 12.275), (\pi, 2.275), (9.42478, 2.475)$, $f(x^*) = 0.397887$.

30. The Shekel SQRN5 function [45]

$n = 4$

$$f(x) = -\sum_{i=1}^{m}\left(c_i + \sum_{j=1}^{n}(x_j - a_{i,j})^2\right)^{-1},$$

gde je

$$m = 5$$

$$[a_{i,j}] = \begin{bmatrix} 4.0 & 4.0 & 4.0 & 4.0 \\ 1.0 & 1.0 & 1.0 & 1.0 \\ 8.0 & 8.0 & 8.0 & 8.0 \\ 6.0 & 6.0 & 6.0 & 6.0 \\ 3.0 & 7.0 & 3.0 & 7.0 \end{bmatrix}, c = [0.1, 0.2, 0.2, 0.4, 0.6]^T.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2),\ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^* = (1, 3, 5, 6)$.
Optimum: $x^0 = (4, 4, 4, 4)$, $f(x^*) = -10.1527$.

31. The Shekel SQRN7 function

$n = 4$

$$f(x) = -\sum_{i=1}^{m}\left(c_i + \sum_{j=1}^{n}(x_j - a_{i,j})^2\right)^{-1},$$

gde je

$$m = 7$$

$$[a_{i,j}] = \begin{bmatrix} 4.0 & 4.0 & 4.0 & 4.0 \\ 1.0 & 1.0 & 1.0 & 1.0 \\ 8.0 & 8.0 & 8.0 & 8.0 \\ 6.0 & 6.0 & 6.0 & 6.0 \\ 3.0 & 7.0 & 3.0 & 7.0 \\ 2.0 & 9.0 & 2.0 & 9.0 \\ 5.0 & 5.0 & 3.0 & 3.0 \end{bmatrix}, c = [0.1, 0.2, 0.2, 0.4, 0.6, 0.6, 0.3]^T.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^* = (1, 3, 5, 6)$.
Optimum: $x^0 = (4, 4, 4, 4)$, $f(x^*) = -10.4023$.

32. The Shekel SQRN10 function

$n = 4$

$$f(x) = -\sum_{i=1}^{m}\left(c_i + \sum_{j=1}^{n}(x_j - a_{i,j})^2\right)^{-1},$$

gde je

$$m = 10$$

$$[a_{i,j}] = \begin{bmatrix} 4.0 & 4.0 & 4.0 & 4.0 \\ 1.0 & 1.0 & 1.0 & 1.0 \\ 8.0 & 8.0 & 8.0 & 8.0 \\ 6.0 & 6.0 & 6.0 & 6.0 \\ 3.0 & 7.0 & 3.0 & 7.0 \\ 2.0 & 9.0 & 2.0 & 9.0 \\ 5.0 & 5.0 & 3.0 & 3.0 \\ 8.0 & 1.0 & 8.0 & 1.0 \\ 6.0 & 2.0 & 6.0 & 2.0 \\ 7.0 & 3.6 & 7.0 & 3.6 \end{bmatrix}, c = [0.1, 0.2, 0.2, 0.4, 0.6, 0.6, 0.3, 0.7, 0.5, 0.5]^T.$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = (1,3,5,6)$.
Optimum: $x^0 = (4,4,4,4)$, $f(x^*) = -10.5358$.

33. The six-hump camel-back polynomial

$n = 4$

$$f(x) = (4.0 - 2.1x_1^2 + x_1^4/3.0)\, x_1^2 + x_1 * x_2 + 4.0(x_2^2 - 1.0)x_2^2$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \ \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = (-1.5, 0.5)$.
Optimum: $x^0 = \pm(-0.0898, 0.7126)$, $f(x^*) = -1.0316$.

34. The Shubert function

$n = 2$

$$f(x) = \left(\sum_{i=1}^{5} i\cos((i+1)x_1 + i)\right)\left(\sum_{i=1}^{5} i\cos((i+1)x_2 + i)\right)$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = (0.5, 1.0)$.
Optimum: $x^0 = \pm(0,0)$, $f(x^*) = 19.8758$.

35. The Stuckman function, [59]

$n = 2$

$$f(X) = -\begin{cases} \lfloor (\lfloor m_1 \rfloor + 1/2)(\sin(a_1)/a_1) \rfloor, & 0 \le x_1 \le b \\ \lfloor (\lfloor m_2 \rfloor + 1/2)(\sin(a_2)/a_2) \rfloor, & b < x_2 \le 10 \end{cases}$$

gde je $a_1 = |x_1 - xr_{1,1}| + |x_2 - xr_{2,1}|$, $a_2 = |x_1 - xr_{1,2}| + |x_2 - xr_{2,2}|$.

Konstante koje određuju problem su slučajne promenljive koje uzimaju vrednost slučajnog broja iz intervala:

$$\begin{array}{rll} b & random & \in [0,10] \\ m_j & random & \in [0,100] \\ xr_{1,1} & random & \in [0,b] \\ xr_{1,2} & random & \in [b,10] \\ xr_{i,j} & random & \in [0,10] \text{ u preostala dva slučaja} \end{array}$$

Ograničenje: $0 \le X \le 10$.
Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = (0.5, 1.0)$.
Optimum: $x^* = \begin{cases} x = (r_{1,1}, r_{1,2}), & m_2 < m_1 \\ x = (r_{1,2}, r_{2,2}), & \text{inače} \end{cases}$

36. The Easom function

$n = 2$

$$f(x) = -\cos x_1 \cos x_2 \exp(-(x_1 - \pi)^2 - (x_2 - \pi)^2)$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = \pi(0.5, 1.0)$.
Ograničenje: $-100 \le x_i \le 100$, $i = 1,2$.
Optimum: $x^0 = \pi(1,1)$, $f(x^*) = -1$.

37. The Bohachevsky function #1 [32]

$n = 2$

$$f(x) = x_1^2 - 0.3\cos(3\pi x_1) + 2x_2^2 - 0.4\cos(4\pi x_2) + 0.7$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = \pi(0.5, 1.0)$.
Ograničenje: $-100 \le x_i \le 100$, $i = 1,2$.
Optimum: $x^0 = \pi(0,0)$, $f(x^*) = 0$.

38. The Bohachevsky function #2 [32]

$n = 2$

$$f(x) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1)\,0.4\cos(4\pi x_2) + 0.3$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = \pi(0.6, 1.3)$.
Ograničenje: $-100 \le x_i \le 100$, $i = 1,2$.
Optimum: $x^0 = \pi(0,0)$, $f(x^*) = 0$.

39. The Bohachevsky function #3 [32]

$n = 2$

$$f(x) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1 + 4\pi x_2) + 0.3$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^* = \pi(0.5, 1.0)$.
Ograničenje: $-100 \leq x_i \leq 100$, $i = 1, 2$.
Optimum: $x^0 = \pi(0, 0)$, $f(x^*) = 0$.

40. The Colville Polynomial [32]

$n = 4$

$$f(x) = 100(x_1 - x_2^2)^2 + (1 - x_1)^2 + 90(x_4 - x_3^2)^2 + (1 - x_3)^2 +$$
$$+ 10.1((x_2 - 1)^2 + (x_1 - 1)^2) + 19.8(x_2 - 1)(x_4 - 1)$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n\times n}).$$

Start: $x^* = \pi(0.5, 1.0, -0.5, -1.0,)$.
Ograničenje: $-10 \leq x_i \leq 10$, $i = 1, 2, 3, 4$.
Optimum: $x^0 = \pi(1, 1, 1, 1)$, $f(x^*) = 0$.

41. The Powell 3D function

$n = 3$

$$f(x) = 3 - (1 + (x_1 - x_2)^2 - \sin(0.5\,\pi\, x_2\, x_3))^{-1} - \exp(-((x_1 + 2x_2 + x_3)/x_2)^2)$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = (0,1,2)$.
Optimum: $x^0 = (1,1,1)$, $f(x^*) = 1$.

42. The Himmelblau function, [32]

$n = 2$

$$f(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = (-1.3, 2.7)$.
Ograničenje: $-5 \le x_i \le 5$, $i = 1,2$.
Optimum: $x^0 = (3,2)$, $f(x^*) = 0$.

43. Strictly convex 1

$n = 10$

$$f(x) = \sum_{i=1}^{n} (\exp(x_i) - x_i)$$

Šum:

$$\xi(x) = z, \text{ gde je } z \sim \mathcal{N}(0,\sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0,\sigma^2 I_{n\times n}).$$

Start: $x^* = (1,2,\ldots,10)/10$.
Optimum: $x^0 = (0,0,\ldots,0)$, $f(x^*) = 10$.

44. Strictly convex 2

$n = 10$

$$f(x) = \sum_{i=1}^{n} \frac{i}{10}(\exp(x_i) - x_i)$$

Šum:

$$\xi(x) = \; z, \text{ gde je } z \sim \mathcal{N}(0, \sigma^2), \; \varepsilon(x) = \hat{z}, \text{ gde je } \hat{z} \sim \mathcal{N}(0, \sigma^2 I_{n \times n}).$$

Start: $x^* = (1, 1, \ldots, 1)$.
Optimum: $x^0 = (0, 0, \ldots, 0)$, $f(x^*) = 5.5$.

6.2 Numerički rezultati algoritma DSLS

Ovde ćemo prikazati efikasnost algoritma DSLS. Ovaj algoritam, koji u prvoj fazi dužine koraka određuje linijskim pretraživanjem (obeležavaćemo sa LS), poredićemo sa SA algoritmom koji od početka koristi predefinisane dužine koraka.

Pseudo kod algoritma DSLS je dat u Algoritmu 6. Dužine koraka u drugoj fazi DSLS i dužine koraka u SA algoritmu određujemo po formuli (4.14):

$$a_k = \frac{a}{(k+1+A)^\alpha}, \;\; k = 0, 1, \ldots$$

sa $a = 1$, $A = 0$, $\alpha = 0.602$, k redni broj iteracije.

Druga stvar je da ćemo testirati efikanost algoritama koji za iterativne pravce uzimaju: negativni gradijent (G), BFGS pravac i SR1 pravac.

U sva tri slučaja pravac se određuje po formuli

$$d_k = -B_k^{-1} G_k,$$

pri čemu je za pravac negativnog gradijenta matrica B_k jedinična matrica u svim iteracijama, a za BFGS i SR1 pravce se, počev od jedinične, ažurira po formulama redom (4.15) i (4.16).

Kada to iskombinujemo, imamo 6 različitih algoritama. Označavaćemo ih:

SA-G	SA-BFGS	SA-SR1
LS-G	LS-BFGS	LS-SR1

Problemi na kojima ćemo testirati ove metode dati su u glavi 6.1 i zavise od parametra σ, standardne devijacije šuma. Testirali smo više vrednosti za standardnu devijaciju šuma. Daćemo rezultate za vrednosti $\sigma = 1, \sigma = 0.2$ i $\sigma = 0.04$.

Za svaki eksperiment smo uradili $N = 50$ nezavisnih iterativnih postupaka polazeći od iste početne tačke. Da bismo u svim eksperimentima imali isti broj računanja funkcije f i gradijenta g, postavili smo $itnlimit = \infty$. Maksimalni broj računanja funkcije je postavljen na $maxfcalc = 200n$, gde je n dimenzija problema. Računanje gradijenta je brojano kao n računanja funkcije cilja.

Parametar $maxfcalcls = 4$ je korišćen za kriterijum prelaska sa linijskog pretraživanja na predefinisane SA korake. Koristili smo $p = 3$ veličinu uzorka za određivanje funkcije cilja i gradijenta kao srednju vrednost, formula (4.13).

Izlazni kriterijum za sve algoritme je bio

$$gradtol = \min\{\sqrt{n}\sigma, 1\}$$

ili maksimalni broj računanja funkcije i gradijenta.

$$maxfcalc = 200n.$$

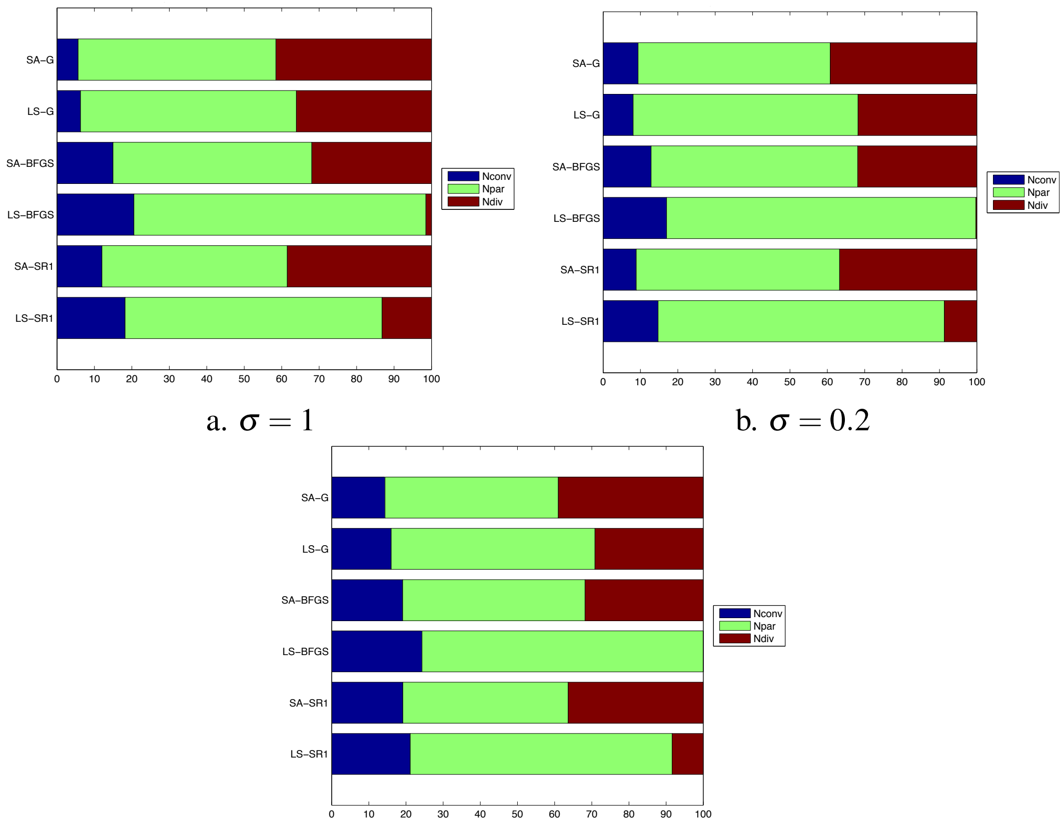
Kada se postupak zaustavljao sa zadovoljenim kriterijumom $\|G_{end}\| \leq gradtol$ smatrali smo da je iterativni postupak bio uspešan. Broj uspešnih završetaka postupka smo označavali $Nconv$. Ako je na kraju postupka

$$\|G_{end}\| > 200\sqrt{n}$$

proglašavali smo da je postupak divergirao. Broj iterativnih postupaka koji su divergirali označavali smo $Ndiv$.

Konačno, kada se iterativni postupak zaustavljao zbog potrošene kvote dozvoljenih računanja funkcije i gradijenta, smatrali smo da je postupak bio delimično uspešan. Broj delimično uspešnih završetaka iterativnog procesa obeležavamo $Npar$. Pretpostavlja se da tada $maxfcalc$ nije bilo dovoljno veliko da bi se postigla konvergencija, ali da su se vrednosti funkcije ipak smanjile kroz iteracije, te da je algoritam ipak napravio neki napredak.

Sve u svemu, imali smo 2200 iterativnih postupaka po svakom algoritmu i na slici 6.1 imamo prikazane procente $Nconv$, $Npar$ i $Ndiv$ za sve tri veličine šuma.



a. $\sigma = 1$ b. $\sigma = 0.2$

c. $\sigma = 0.04$

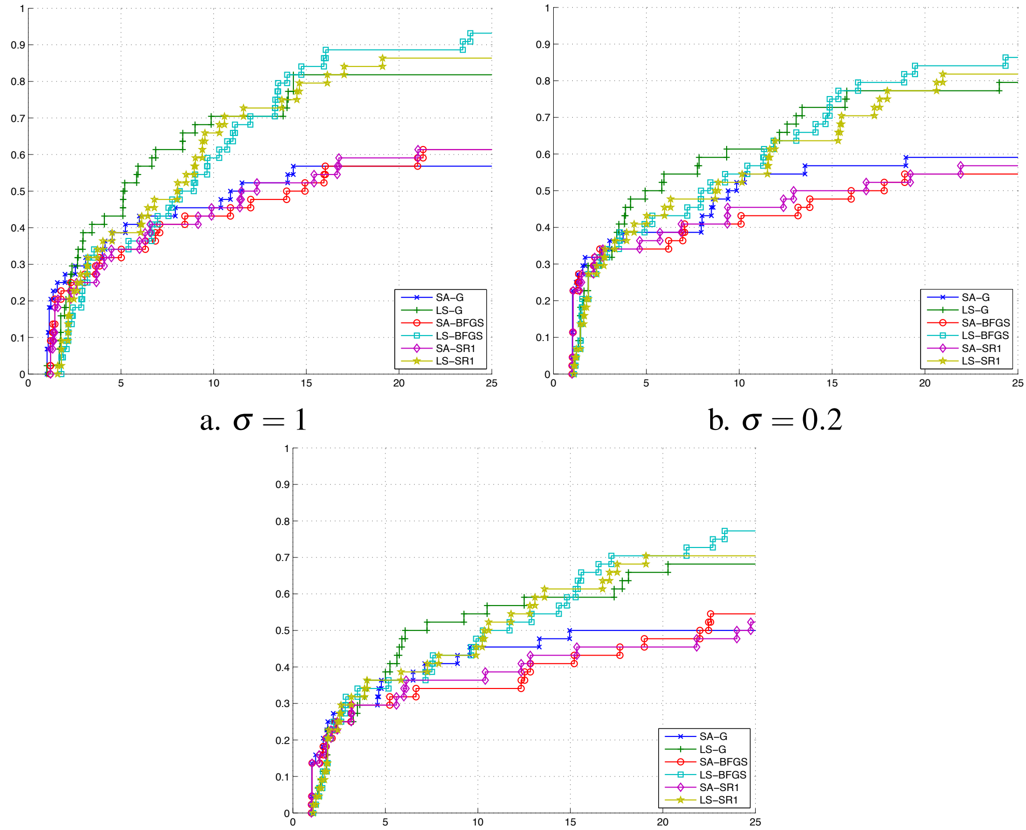
Slika 6.1: Procenti broja uspešnih, delimično uspešnih i divergentnih

Rezultati sa slike 1 jasno pokazuju da Algoritam DSLS ima značajno manji broj divergencija u odnosu na SA za sve nivoe šuma.

Dalje, pravci koji koriste informaciju drugog reda, BFGS i SR1 su bolji od pravca negativnog gradijenta i za SA i za DSLS algoritam. DSLS sa BFGS pravcima je najbolji za sva tri nivoa šuma.

Na slici 2 imamo prikaz rezultata pomoću profila performansi, (Performance profiles), Doolan, More [22].

Mera performanse je prosečan broj računanja funkcije za konvergentne i delimično konvergentne iterativne postupke, normalizovan u odnosu na dimenziju



a. $\sigma = 1$ b. $\sigma = 0.2$

c. $\sigma = 0.04$

Slika 6.2: Performans profili za svih 6 metoda

problema:

$$fc_{ij} = \frac{1}{|Ncon_{ij}UNpar_{ij}|} \sum_{r \in Ncon_{ij}UNpar_{ij}} \frac{fcalc_{ij}^{r}}{n_j},$$

gde je $fcalc_{ij}^{r}$ broj računanja funkcije (i gradijenta) za rešavanje problema j dimentije n_j metodom i u r-tom iterativnom postupku. $Nconv_{ij}$ i $Npar_{ij}$ su redom broj konvergencija i broj delimično uspešnih iterativnih postupaka pri rešavanju problema j metodom i.

Performans profili jasno pokazuju da su DSLS postupci bolji od SA postupaka. Razlika je veća za veću standardnu devijaciju šuma σ.

Ako se posmatraju DSLS postupci, BFGS i SR1 pravci, koji sadrže i informaciju drugog reda, bolji su od pravaca negativnog gradijenta.

Zaključujemo da broj računanja funkcije (zajedno sa računanjima gradijenta) u oba poređenja daje prednost algoritmu koji smo predložili.

Glava 7

Primena algoritma FNM na modele ekvilibriuma

U neoklasičnoj ekonomiji ravnotežnu cenu, odnosno ekvilibrium, dobijamo nalaženjem pozitivnog rešenja π^* sistema (5.4):

$$\xi(\pi^*) = 0\,,$$

za funkciju viškova $\xi(\pi)$ definisanu

$$\xi(\pi) = \sum_{j=1}^{m} x^j(\pi) - \sum_{j=1}^{m} \omega^j,$$

gde su $x^j(\pi)$ vrednosti u kojima funkcije korisnosti dostižu maksimume

$$x^j(\pi) = \arg \max_{\pi x \leq \pi\omega^j} u^j(x)\,, \quad j = 1, 2, \ldots, m$$

za početna zaduženja ω^j.

Zbog homogenosti stepena nula, ravnotežne cene možemo tražiti na pozitivnom simpleksu

$$\Delta_+ = \left\{ \pi \in \mathbb{R}^n_+ \,\middle|\, \sum_{i=1}^{n} \pi_i = 1 \right\},$$

tako što ćemo poslednju jednačinu zameniti jednačinom

$$\sum_{i=1}^{n} \pi_i = 1.$$

Dobijamo problem

$$\bar{\xi}(\pi) = 0\,.$$

7.1 Statistika parametara

U našim numeričkim eksperimentima rešavaćemo ekvilibrium za više vrednosti parametara. Parametre dobijamo kao pseudo slučajne brojeve raspoređene po normalnoj raspodeli. Za srednju vrednost normalne raspodele uzimamo početnu vrednost iz standardnih test problema, a standardna devijacija je zadata.

U prvoj fazi algoritma dobijamo $x^{\bar{w}}$ za srednju vrednost uzorka parametara $\bar{w}$. Zbog nelinearnosti sistema dobijeno rešenje je aproksimacija srednje vrednosti uzorka rešenja. Potom u drugoj fazi dobijamo uzorak rešenja čiji elementi daju detaljniju sliku i omogućavaju statističku analizu rešenja.

Za uzorak rešenja smo prikazali deskriptivne statističke podatke i izvršili testiranje na normalnost. Koristili smo Anderson-Darling test ([1]) sa pragom značajnosti $\alpha = 5\%$. Anderson-Darling statistika je

$$A^2 = -\frac{1}{N}\left\{\sum_{i=1}^{N}(2i-1)\left[\ln z_i + \ln(1-z_{N+1-i})\right]\right\} - N\,,$$

gde je N veličina uzorka a z_i je i-ta vrednost sortiranog uzorka. Mi smo koristili modifikovanu Anderson-Darling statistiku

$$A_m^2 = A^2\left(1 + \frac{0.75}{N} + \frac{2.25}{N^2}\right)\,,$$

čija kritična vrednost za prag značajnosti $\alpha = 0.05$ iznosi 0.752, a za $\alpha = 0.01$ iznosi 1.035.

Na slikama smo prikazali vrednosti statistike spljoštenost (kurtosis) i zakrivljenost (skewness) po komponentama uzorka ekvilibriuma $\pi^{*i}, i = 1,\ldots,N$. Ove vrednosti opisuju oblik gustine raspodele slučajne veličine koja odgovara raspodeli komponente uzorka. Formule za kurtosis i skewness su

$$\beta_1^2 = \mu_3^2/\mu_2^3\,,\ \beta_2 = \mu_4/\mu_2^2$$

redom, gde je

$$\mu_k = \frac{1}{N}\sum_{i=1}^{N}(x^i - m_1)^k\,,\ m_1 = \frac{1}{N}\sum_{i=1}^{N}(x^i)\,.$$

Za normalnu raspodelu kurtosis je 3 a skewness 0.

Vrednosti β_1^2, β_2, histogram, statistički moment m_1 i μ_k, vrednost Anderson-Darling statistike i druge deskriptivne statistike sadrže važne podatke o ponašanju ekvilibriuma kad na parametre deluje Gausovski šum.

7.1.1 Primer 1

Posmatramo neoklasičnu ekonomiju sa dve robe i dva agenta sa početnim zaduženjima $w^1 = (3,1)$ i $w^2 = (1,2)$.

Funkcija korisnosti prvog agenta je Cobb-Douglas funkcija

$$u^1(\pi) = \pi_1^{a_1^1} \pi_2^{1-a_1^1}.$$

Funkcija korisnosti drugog agenta je funkcija fiksne proporcije

$$u^2(\pi) = \min\left\{\frac{\pi_1}{a_1^2}, \frac{\pi_2}{a_2^2}\right\}.$$

Funkcije potrebe i tražnje su

$$\xi_1(\pi) = a_1^1 \frac{3\pi_1 + \pi_2}{\pi_1} + a_1^2 \frac{\pi_1 + 2\pi_2}{a_1^2\pi_1 + a_2^2\pi_2} - 4 \tag{7.1}$$

$$\xi_2(\pi) = (1 - a_1^1)\frac{3\pi_1 + \pi_2}{\pi_2} + a_2^2 \frac{\pi_1 + 2\pi_2}{a_1^2\pi_1 + a_1^2\pi_2} - 3\,. \tag{7.2}$$

Zamenimo (7.2) sa

$$\xi_2(\pi) = \pi_1 + \pi_2 - 1 \tag{7.3}$$

i od sistema (5.4) dobijamo

$$\bar{\xi}(\pi) = 0\,.$$

	NM	FNM
broj iteracija	2954	4553
broj izračunavanja funkcije	2954	4553
broj izračunavanja Jacobijana	2954	6

Tabela 7.1: Rešavanje ekvilibriuma za neoklasičnu ekonomiju (7.1), (7.3) sa 2 robe i 2 agenta, veličine uzorka parametara $N = 500$

Početna vrednost Njutnovog iterativnog postupka (NM) i FNM je $\pi_0 = (0.1, 0.9)$ za sve vrednosti uzorka. Izlazni kriterijum za oba algoritma NM i FNM je $\|\bar{\xi}\| < \varepsilon = 10^{-6}$.

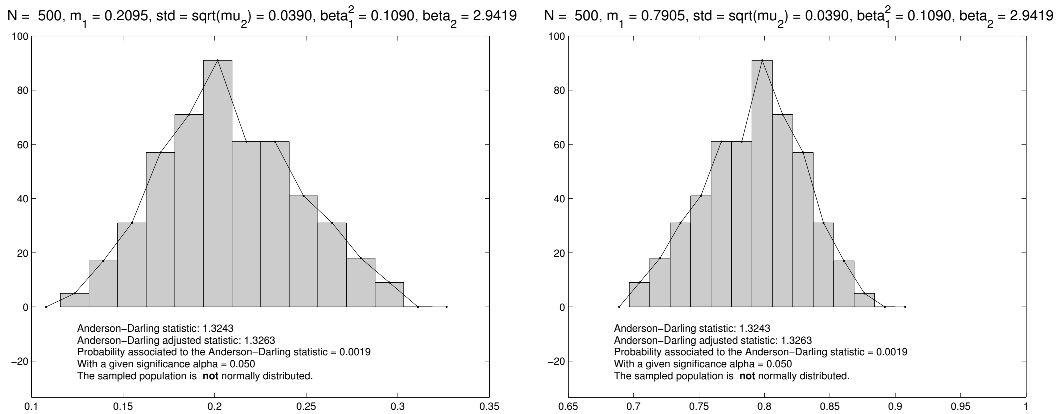
Za vrednost parametara $a_1^1 = 0.4$, $a_1^2 = 2$, $a_2^2 = 3$, NM daje rešenje $\pi^* = (0.2087, 0.7913)$ u 6 iteracija.

Potom kreiramo uzorak od $N = 500$ vrednosti za parametre, sa normalnim raspodelama: $a_1^1 : \mathcal{N}(0.4, 0.05)$, $a_1^2 : \mathcal{N}(2.0, 0.05)$, $a_2^2 : \mathcal{N}(3.0, 0.05)$.

Neka W označava uređenu trojku parametara $W = (a_1^1, a_1^2, a_2^2)$ i $w^i, i = 1, \ldots, N$ neka je uzorak parametara.

Primenimo oba postupka: NM i FNM na uzorak parametara. Tako dobijamo uzorak rešenja π^{*i}, $i = 1, \ldots, N$.

U tabeli 7.1 poredimo ukupni broj iteracija, broj računanja funkcije i Jakobijana za postupke NM i FNM. Pri tome je u FNM korišćen Jakobijan $A = \xi'(x^{\bar{w}}, \bar{w})$ iz rešenja $x^{\bar{w}}$ NM metoda za srednju vrednost parametara $\bar{w}$. Prosečan broj iteracija za ovih 500 iterativnih postupaka je 5.91 za NM i 9.11 za FNM. Ipak, vreme potrebno za iteracija, CPU, koje iznosi 0.156 za NM i 0.098 za FNM jasnije oslikava prednost FNM.



Slika 7.1: Histogram i deskriptivna statistika uzorka rešenja π_1^* (levo) i π_2^* (desno) za uzorak parametara veličine $N = 500$

Na slici 7.1 vidimo histogram i deskriptivne statističke podatke uzorka rešenja za komponente rešenja π_1^* i π_2^*. Prirodno je što smo dobili simetrične rezultate, jer su π_1^* i π_2^* zavisni ($\pi_1 + \pi_2 = 1$).

Problem (5.4) definiše neprekidno preslikavanje skupa parametara u skup rešenja. To preslikavanje definiše slučajnu promenljivu rešenja koja nastaje transformacijom slučajne promenljive uzorka parametara. Iz histograma rešenja možemo zaključiti da je slučajna promenljiva rešenja blago nagnuta ($\beta_1^2 = 0.1090$, a za simetrične raspodele, kao što je normalna raspodela $\beta_1^2 = 0$).

Parametar spljoštenosti uzorka rešenja je $\beta_2 = 2.9419$, blizu $\beta_2 = 3$ za normalnu raspodelu.

Ipak, pored toga što su ova dva parametra blizu vrednosti parametara normalne raspodele, modifikovana Anderson-Darling statistika $A_m^2 = 1.3263$ je prešla kritičnu vrednost, pa se stoga odbacuje hipoteza o normalnoj raspodeli ($P < 0.0019$).

7.1.2 Primer 2

Posmatramo neoklasičnu ekonomiju sa $n = 3$ robe i $m = 4$ agenta sa CES funkcijom korisnosti. Funkcije potrebe i tražnje su

$$\xi_j(\pi) = \sum_{i=1}^{4} \left[\frac{\sum_{k=1}^{3} \omega_k^i \pi_k}{\pi_j^{b^i} \sum_{k=1}^{3} \alpha_k^i \pi_k^{1-b^i}} \alpha_j^i - \omega_j^i \right], \text{ for } j = 1,2,3. \tag{7.4}$$

U ovom primeru parametri i početna zaduženja su dati u matricama

$$\alpha := \alpha_0 = \begin{bmatrix} 0.1 & 0.7 & 0.2 \\ 0.1 & 0.4 & 0.5 \\ 0.2 & 0.3 & 0.5 \\ 0.9 & 0.05 & 0.05 \end{bmatrix}, \omega := \omega_0 = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 2 & 0 & 3 \\ 1 & 1 & 2 \end{bmatrix}, b := b_0 = \begin{bmatrix} 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \end{bmatrix},$$

gde i-ta vrsta predstavlja i-tog agenta. Element u i-toj vrsti i j-toj koloni matrice α odnosno ω je α_j^i odnosno ω_j^i iz formule (7.4).

Zamenom poslednje jednačine iz (5.4) jednačinom

$$\xi_3(\pi) = \pi_1 + \pi_2 + \pi_3 - 1 \tag{7.5}$$

dobijamo $\bar{\xi}$.

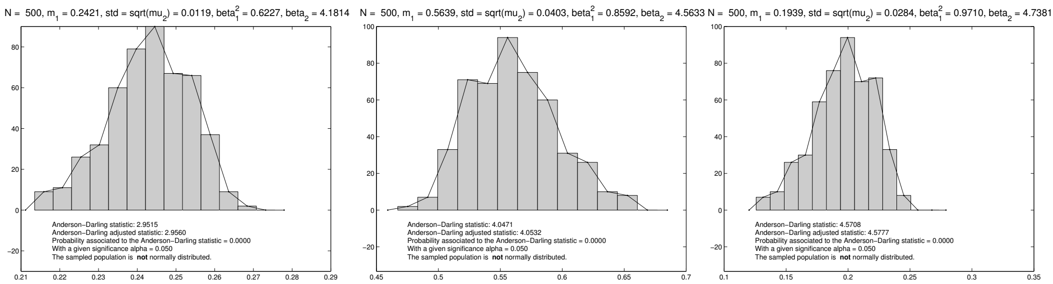
Uzimajuće početnu vrednost $\pi_0 = (0.4, 0.2, 0.4)$, Njutnov postupak sa izlaznim kriterijumom $\|\bar{\xi}\| < \varepsilon = 10^{-6}$ u 7 iteracija daje rešenje $\pi^* = (0.2441, 0.5566, 0.1993)$.

Uzorak koji koristimo u ovom primeru dobijamo dodajući na komponente vektora b: $b^i, i = 1,2,3,4$ $N = 500$ vrednosti šuma raspoređene po normalnoj raspodeli: $b^i : \mathcal{N}(0.5, 0.1), i = 1,2,3,4$.

Isto kao u primeru 1, u tabeli 7.2 dajemo ukupan broj iteracija, broj računanja funkcije i Jakobijana za oba metoda NM i FNM. Prosečan broj iteracija je 6.60

	NM	FNM
broj iteracija	3301	4581
broj izračunavanja funkcije	3301	4581
broj izračunavanja Jacobijana	3301	7

Tabela 7.2: Rešavanje ekvilibriuma neoklasične ekonomije (5.4), (7.4), (7.5), sa CES funkcijom korisnosti, 3 robe i 4 agenta, uzorak veličine $N = 500$ vrednosti parametra



Slika 7.2: Histogram i deskriptivna statistika za uzorak rešenja π_1^* (levo), π_2^* (sredina) i π_3^* (desno) za uzorak veličine $N = 500$ vrednosti (5.4), (7.4), (7.5)

za NM i 9.16 za FNM, ali vreme računanja (CPU time) je 1.267 za NM i 0.820 za FNM.

Deskriptivna statistika je data u tabeli 7.2.

Vidimo iz histograma da je uzorak asimetričan, $\beta_1^2 > 0.6$ i da je spljoštenost (kurtosis) daleko od 3, spljoštenosti za normalnu raspodelu. Pored toga, AD test sa greškom manjom od $P < 10^{-5}$ tvrdi da raspodela uzorka nije normalna raspodela.

7.1.3 Primer 3

Posmatramo prostorni model ekvilibriuma (5.5) iz [44], nelinearan, sa jednom robom, sa dva snabdevača i dva potrošača.

Pretpostavljamo da su funkcije ponude i tražnje

$$S_i(Y_i) = u_i + \left(\frac{Y_i}{v_i}\right)^2,$$

$$I_j(Z_j) = \frac{1}{\theta_j} \ln\left(\frac{2000}{Z_j}\right).$$

Pretpostavljmo i da su funkcije troškova linearne:

$$c_{ij}(x_{ij}) = \gamma_{ij} x_{ij} \,.$$

Parametri u_i, v_j, θ_j i γ_{ij} su redom iz intervala $[3,5]$, $[15,20]$, $[0.1,0.5]$, $[5,10]$.

Početna vrednost za NM i FNM je bila $x_0 = \begin{bmatrix} 20 & 20 \\ 20 & 20 \end{bmatrix}$ a izlazni kriterijum je bio $\|F\| < \varepsilon = 10^{-6}$.

Za vrednosti parametara

$$u = [4.00, 4.00]^T, v = [17.50, 17.50]^T, \theta = [0.30, 0.30]^T, \gamma = \begin{bmatrix} 7.50 & 7.50 \\ 7.50 & 7.50 \end{bmatrix},$$

rešenje dobijeno u 4 iteracije Njutnove metode je $x^* = \begin{bmatrix} 2.182 & 2.182 \\ 2.182 & 2.182 \end{bmatrix}$.

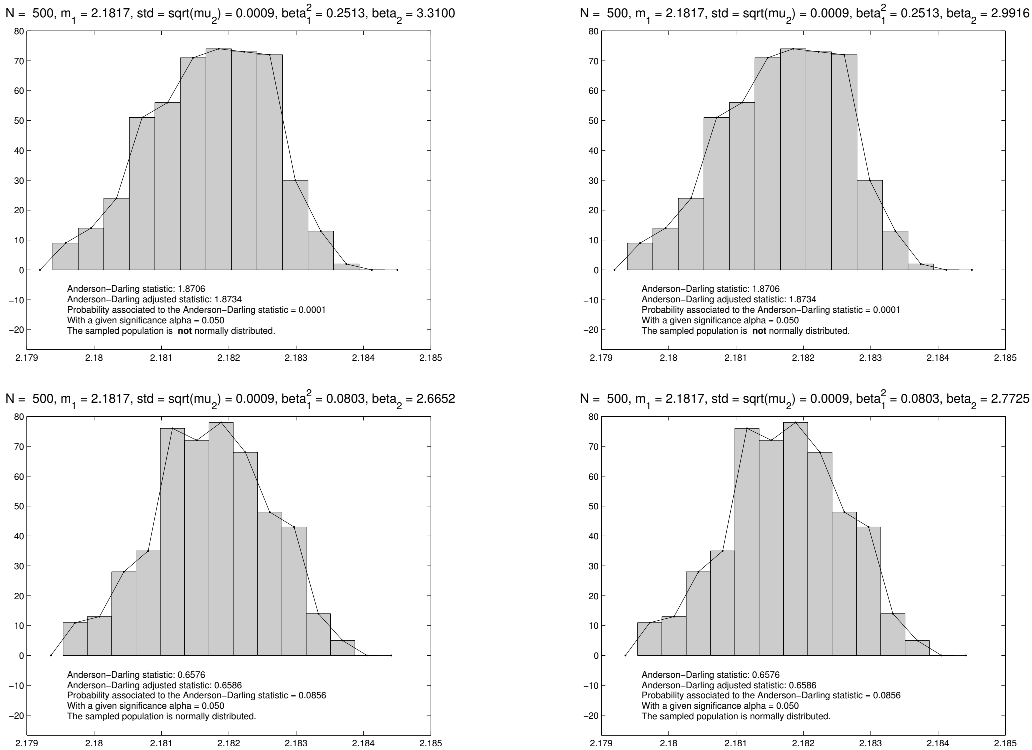
Zatim koristimo uzorak veličine $N = 500$ nezavisnih slučajnih vrednosti, za v: v_1 i v_2, raspoređenih po normalnoj raspodeli $v_{1,2} : \mathcal{N}(17.5, 1)$. Ostali parametri su kao gore. Njutnovom metodom smo rešavali problem ekvilibriuma. Potom smo, isto kao u primeru 1, fiksnim Njutnovim postupkom rešavali ekvilibriume, koristeći Jakobijan iz Njutnove metode primenjene za srednje vrednosti parametara. Podaci o iteracijama, izračunavanjima funkcije i Jakobijana su u tabeli 7.3.

	NM	FNM
broj iteracija	5465	3227
broj izračunavanja funkcije	5465	3227
broj izračunavanja Jacobijana	5465	4

Tabela 7.3: Nalaženje SPAT ekvilibriuma (5.5), sa dva snabdevača i dva potrošača, uzorak $N = 500$ vrednosti

Prosečan broj iteracija je 10.93 za NM i 6.45 za FNM i CPU vremenom 1.272 za NM i 0.289 za FNM. Na slici 7.3 imamo histogram i deskriptivnu statistiku za uzorak rešenja.

U dva od 4 slučaja nulta hipoteza o normalnosti nije odbačena. Ako bismo želeli povećati uzorak radi ponavljanja testova, FNM bi se pokazao još efikasnijim od NM.



Slika 7.3: Histogram i deskriptivna statistika za uzorak rešenja $x^*_{1,1}$ (gore levo), $x^*_{1,2}$ (gore desno), $x^*_{2,1}$ (dole levo) i $x^*_{2,2}$ (dole desno) za uzorak veličine $N = 500$ vrednosti parametara problema (5.5)

7.2 CPU vreme i veće dimenzije problema

U tabeli 7.1 i tabeli 7.2 se vidi prednost fiksnog Njutnovog postupka. CPU vreme koje vidimo je vreme korišćeno za račun ekvilibriuma za sve iteracije svih N elemenata uzorka parametara, mereno u sekundama.

Povećavanje dimenzije problema pokazuje još veću prednost FNM. Testirali smo primer 2 i primer 3 za razne vrednosti n i m. Za primer 2 smo generisali matricu α_0 slučajnim vrednostima čiji zbir po vrstama iznosi 1, $w_0 = 2 + \mathscr{U}(0,1)$, $b = [0.9; 0.9; ...; 0.9]$. Sa $\mathscr{U}(0,1)$ smo označili uniformnu raspodelu na intervalu $[0,1]$. Testirano je na uzorku veličine $N = 500$, prikazali smo broj iteracija i CPU[1] vreme u tabeli 7.4.

[1]Intel(R) Core(TM)2Duo CPU E8400 @ 3.00GHz

	NM iter	NM CPU	FNM iter	FNM CPU
$m=4, n=3$	2134	0.773	3156	0.555
$m=8, n=6$	2893	1.383	3200	0.606
$m=16, n=12$	3072	3.554	3949	0.959
$m=32, n=24$	2514	14.592	4144	1.988

Tabela 7.4: Primer 2

Za primer 3 parametre smo zadali: $u_i = 4$, $v_j = 17.5 + 0.5 \cdot \mathscr{N}(0,1)$, $\theta_j = 0.3$ i $\gamma_{ij} = 7.5$ za sve i i j. $\mathscr{N}(0,1)$ označava nezavisne, normalno raspoređene vrednosti sa očekivanjem 0 i standardnom devijacijom 1. Početna vrednost je bila

$$x_0 = \begin{bmatrix} 20 & \cdots & 20 \\ \vdots & \ddots & \vdots \\ 20 & \cdots & 20 \end{bmatrix},$$

a izlazni kriterijum $\|F\| < \varepsilon = 10^{-6}$. Rezultati su dati u tabeli 7.5, isto kao pre: ukupni broj iteracija i CPU vreme u sekundama.

	NM iter	NM CPU	FNM iter	FNM CPU
$m=2, n=2$	5248	1.212	2989	0.263
$m=4, n=4$	5902	1.892	3719	0.334
$m=6, n=6$	5521	2.496	3988	0.375
$m=8, n=8$	5751	4.034	3997	0.397
$m=10, n=10$	5950	7.408	3996	0.546

Tabela 7.5: Primer 3

7.3 Zaključak

Problemi ekvilibriuma se često postavljaju kao sistemi nelinearnih jednačina koje zavise od parametra. Rešavanje većeg broja sistema sa simuliranim vrednostima parametara daje više informacija o ekvilibriumu.

Predstavili smo model u kome parametri uzimaju vrednosti po zakonu raspodele slučajne promenljive. Svaki od sistema za vrednosti parametara iz uzorka smo rešavali fiksnim Njutnovim postupkom. Na taj način smo dobili uzorak rešenja relativno jeftinim postupkom.

Upotreba viška informacija može da nam prikaže deskriptivno statističke podatke o uzorku rešenja i odgovori na pitanje da li uzorak rešenja ima normalnu raspodelu. U našem slučaju, uzorak nema normalnu raspodelu pa se ne mogu primenjivati formule za intervale poverenja i parametarske testove za vrednosti ekvilibriuma.

Problem formula za intervale poverenja ravnotežnih cena je i dalje otvoren. Stohastičke simulacije (Monte-Carlo) nam mogu dati neke informacije.

Rešavanje velikog broja sistema zahteva više CPU vremena koje može biti ograničeno. Na našim primerima vidimo da se upotreba sličnosti u podacima može iskoristiti da bi se dobila pristupačna metoda. U primerima 1 i 2 se povećao broj računanja funkcije u iterativnom postupku, ali se "skupi" Jakobijan skoro ni ne računa. Kada se poveća dimenzija sistema, ta prednost se drastično odražava na procesorsko vreme.

Literatura

[1] ANDERSON, T. W., AND DARLING, D. A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics 23*, 2 (06 1952), 193–212.

[2] ARNAL, J., MIGALLÓN, H., MIGALLÓN, V., AND PENADÉS, J. Parallel Newton iterative methods based on incomplete lu factorizations for solving nonlinear systems. In *High Performance Computing for Computational Science - VECPAR 2004*, M. Daydé, J. Dongarra, V. Hernández, and J. Palma, Eds., vol. 3402 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2005, pp. 716–729.

[3] BEALE, E. On an iterative method of finding a local minumum of a function of more than one variable. Tech. Rep. No. 25, Statistical Techniques Research Group, Princeton Univ., Princeton, N.J., 1958.

[4] BERTSEKAS, D. P., AND TSITSIKLIS, J. N. Gradient convergence in gradient methods with errors. *SIAM Journal on Optimization 10*, 3 (2000), 627–642.

[5] BIGGS, M. Minimization algorithms making use of non-quadratic properties of the objective function. *IMA Journal of Applied Mathematics 8*, 3 (1971), 315–327.

[6] BLUM, J. R. Approximation methods which converge with probability one. *Ann. Math. Statist. 25*, 2 (06 1954), 382–386.

[7] BLUM, J. R. Multidimensional stochastic approximation methods. *Ann. Math. Statist. 25*, 4 (12 1954), 737–744.

[8] BOX, M. A comparison of several current optimization methods, and the use of transformations in constrained problems. *The Computer Journal 9*, 1 (1966), 67–77.

[9] BRENT, R. *Algorithms for Minimization Without Derivatives*. Prentice-Hall series in automatic computation. Prentice-Hall, 1972.

[10] BYRD, R., CHIN, G., NEVEITT, W., AND NOCEDAL, J. On the use of stochastic hessian information in optimization methods for machine learning. *SIAM Journal on Optimization 21*, 3 (2011), 977–995.

[11] BYRD, R. H., CHIN, G. M., NEVEITT, W., AND NOCEDAL, J. On the use of stochastic hessian information in unconstrained optimization. Tech. rep., arXiv:1401.7020 [math.OC], 2010.

[12] BYRD, R. H., CHIN, G. M., NOCEDAL, J., AND WU, Y. Sample size selection in optimization methods for machine learning. *Math. Program. 134*, 1 (Aug. 2012), 127–155.

[13] CHEN, H. *Stochastic Approximation and Its Applications*. Nonconvex Optimization and Its Applications. Springer, 2002.

[14] CHUNG, K. *A Course in Probability Theory*. Academic Press, 2001.

[15] COLVILLE, A.R. A comparative study of nonlinear programming codes. Rep. 320-2949, IBM New York Scientific Center, 1968.

[16] COX R.A. Comparison of the performance of seven optimization algorithms on twelve unconstrained optimization problems. ref. 1335cn04. *Gulf Research and Development Company, Pittsburg, Jan. 1969.* (1969).

[17] DE JONG, K. A. *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. PhD thesis, University of Michigan, Ann Arbor, MI, USA, 1975. AAI7609381.

[18] DE LA FUENTE, A. *Mathematical Methods and Models for Economists*. Cambridge University Press, New York, NY, USA, 2000.

[19] DELYON, B., AND JUDITSKY, A. Accelerated stochastic approximation. *SIAM Journal on Optimization 3*, 4 (1993), 868–881.

[20] DENNIS, J., AND SCHNABEL, R. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1996.

[21] DINIZ-EHRHARDT, M. A., MARTÍNEZ, J. M., AND RAYDAN, M. A derivative-free nonmonotone line-search technique for unconstrained optimization. *J. Comput. Appl. Math. 219*, 2 (Sept. 2008), 383–397.

[22] DOLAN, E. D., AND MORÉ, J. J. Benchmarking optimization software with performance profiles. *Mathematical programming 91*, 2 (2002), 201–213.

[23] EAVES, B. C., AND SCHMEDDERS, K. General equilibrium models and homotopy methods. *Journal of Economic Dynamics and Control 23*, 9 (1999), 1249–1279.

[24] ESTEBAN-BRAVO, M. Computing equilibria in general equilibrium models via interior-pointmethods. *Comput. Econ. 23*, 2 (Mar. 2004), 147–171.

[25] ESTEBAN-BRAVO, M. An interior-point algorithm for computing equilibria in economies with incomplete asset markets. *Journal of Economic Dynamics and Control 32*, 3 (2008), 677–694.

[26] FABIAN, V. On asymptotic normality in stochastic approximation. *Ann. Math. Statist. 39*, 4 (08 1968), 1327–1332.

[27] FLETCHER, R. Function minimization without evaluating derivatives—a review. *The Computer Journal 8*, 1 (1965), 33–41.

[28] FLETCHER, R., AND POWELL, M. J. A rapidly convergent descent method for minimization. *The Computer Journal 6*, 2 (1963), 163–168.

[29] GILL P.E., MURRAY W., AND PITFIELD R.A. *The implementation of two revised quasi Newton algorithms*. National Physical Laboratory Report NAC-11, 1972.

[30] GREGORY, R. T., AND KARNEY, D. L. *A collection of matrices for testing computational algorithms*. Wiley-Interscience, New York, 1969.

[31] HILL, S. D., AND FU, M. C. Simulation optimization via simultaneous perturbation stochastic approximation. In *Proceedings of the 26th conference on Winter simulation* (1994), Society for Computer Simulation International, pp. 1461–1464.

[32] JAMIL, M., AND YANG, X.-S. A literature survey of benchmark functions for global optimisation problems. *International Journal of Mathematical Modelling and Numerical Optimisation 4*, 2 (2013), 150–194.

[33] JUDD, K. L. *Numerical methods in economics*. MIT press, 1998.

[34] KESTEN, H. Accelerated stochastic approximation. *The Annals of Mathematical Statistics 29*, 1 (03 1958), 41–59.

[35] KIEFER, J., AND WOLFOWITZ, J. Stochastic estimation of the maximum of a regression function. *Ann. Math. Statist. 23*, 3 (09 1952), 462–466.

[36] KOWALIK, J., AND OSBORNE, M. *Methods for unconstrained optimization problems*. Mathematical Linguistics and Automatic Language Processing. American Elsevier Pub. Co., 1968.

[37] KREJIĆ, N., LUŽANIN, Z., AND OVCIN, Z. Stochastic Newton-like methods for computing equilibria in general equilibrium models. *Computational & Applied Mathematics 30* (00 2011), 127 – 149.

[38] KREJIĆ, N., LUŽANIN, Z., OVCIN, Z., AND STOJKOVSKA, I. Descent direction method with line search for unconstrained optimization in noisy environment. *Optimization Methods and Software 30*, 6 (2015), 1164–1184.

[39] KREJIĆ, N., LUŽANIN, Z., AND STOJKOVSKA, I. A gradient method for unconstrained optimization in noisy environment. *Appl. Numer. Math. 70* (Aug. 2013), 1–21.

[40] LAGUNA, M., AND MARTÍ, R. Experimental testing of advanced scatter search designs for global optimization of multimodal functions. *Journal of Global Optimization 33*, 2 (2005), 235–255.

[41] LEON A. A comparison of eight known optimizing procedures. *Proceedings of the Symposium on Recent Advances in Optimization Techniques* (1966).

[42] LEVY, M., TROSSET, M., AND KINCAID, R. Quasi-Newton methods for stochastic optimization. In *Proceedings of the 4th International Symposium on Uncertainty Modelling and Analysis* (Washington, DC, USA, 2003), ISUMA ’03, IEEE Computer Society, pp. 304–.

[43] LOÈVE, M. On almost sure convergence. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (Berkeley, Calif., 1951), University of California Press, pp. 279–303.

[44] MARCOTTE, P., MARQUIS, G., AND ZUBIETA, L. A Newton-sor method for spatial price equilibrium. *Transportation Science 26*, 1 (1992), 36–47.

[45] MICHALEWICZ, Z. *Genetic Algorithms + Data Structures = Evolution Programs (3rd Ed.)*. Springer-Verlag, London, UK, UK, 1996.

[46] MORÉ, J., GARBOW, B., AND HILLSTROM, K. Testing unconstrained optimization software. *ACM Trans. Math. Softw. 7*, 1 (Mar. 1981), 17–41.

[47] NOCEDAL, J., AND WRIGHT, S. J. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2006.

[48] RAHNAMAYAN, S., TIZHOOSH, H. R., AND SALAMA, M. M. A novel population initialization method for accelerating evolutionary algorithms. *Computers & Mathematics with Applications 53*, 10 (2007), 1605–1614.

[49] ROBBINS, H., AND MONRO, S. A stochastic approximation method. *The Annals of Mathematical Statistics 22*, 3 (09 1951), 400–407.

[50] ROBBINS, H., AND SIEGMUND, D. A convergence theorem for non negative almost supermartingales and some applications. In *Herbert Robbins Selected Papers*, T. Lai and D. Siegmund, Eds. Springer New York, 1985, pp. 111–135.

[51] SACKS, J. Asymptotic distribution of stochastic approximation procedures. *Ann. Math. Statist. 29*, 2 (06 1958), 373–405.

[52] SCARF, H. E. The Computation of Equilibrium Prices: An Exposition. Cowles Foundation Discussion Papers 473, Cowles Foundation for Research in Economics, Yale University, 1977.

[53] SCHRAUDOLPH, N. N., YU, J., AND GUNTER, S. A stochastic quasi-Newton method for online convex optimization. In *Proceedings of the 11th International Conference Artificial Intelligence and Statistics (AISTATS)* (2007), pp. 433–440.

[54] SPALL, J. Adaptive stochastic approximation by the simultaneous perturbation method. *Automatic Control, IEEE Transactions on 45*, 10 (2000), 1839–1853.

[55] SPALL, J. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*. Wiley Series in Discrete Mathematics and Optimization. Wiley, 2005.

[56] SPALL, J. C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *Automatic Control, IEEE Transactions on 37*, 3 (1992), 332–341.

[57] SPALL, J. C. An overview of the simultaneous perturbation method for efficient optimization. *Johns Hopkins APL Technical Digest 19*, 4 (1998), 482–492.

[58] SPEDICATO, E. Computational experience with quasi-Newton algorithms for minimization problems of moderately large size. *Rep. CISE-N-175, Segrate (Milano)* (1975).

[59] STUCKMAN, B. E. A global search method for optimizing nonlinear systems. *IEEE Transactions on Systems, Man, and Cybernetics 18*, 6 (Nov 1988), 965–977.

[60] WARDI, Y. A stochastic steepest-descent algorithm. *J. Optim. Theory Appl. 59*, 2 (Nov. 1988), 307–323.

[61] WARDI, Y. Stochastic algorithms with armijo stepsizes for minimization of functions. *J. Optim. Theory Appl. 64*, 2 (Feb. 1990), 399–417.

[62] WHITLEY, D., RANA, S., DZUBERA, J., AND MATHIAS, K. E. Evaluating evolutionary algorithms. *Artificial Intelligence 85*, 1–2 (1996), 245 – 276.

[63] WOLFOWITZ, J. On the stochastic approximation method of Robbins and Monro. *The Annals of Mathematical Statistics 23*, 3 (1952), pp. 457–461.

[64] WOLPERT, D. H., AND MACREADY, W. G. No free lunch theorems for optimization. *Evolutionary Computation, IEEE Transactions on 1*, 1 (1997), 67–82.

[65] XU, Z. A combined direction stochastic approximation algorithm. *Optimization Letters 4*, 1 (2010), 117–129.

[66] XU, Z., AND DAI, Y.-H. A stochastic approximation frame algorithm with adaptive directions. *Numer. Math. Theory Methods Appl 1*, 4 (2008), 460–474.

[67] XU, Z., AND DAI, Y.-H. New stochastic approximation algorithms with adaptive step sizes. *Optimization Letters 6*, 8 (2012), 1831–1846.

[68] YOUSEFIAN, F., NEDIĆ, A., AND SHANBHAG, U. V. On stochastic gradient and subgradient methods with adaptive steplength sequences. *Automatica 48*, 1 (Jan. 2012), 56–67.

[69] ZOUTENDIJK, G. Nonlinear programming, computational methods. In *Integer and Nonlinear Programming*, J. Abadie, Ed. North–Holland, Amsterdam, 1970, pp. 37–86.

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCES
KEY WORDS DOCUMENTATION

Accession number:
ANO
Identification number:
INO
Document type: Monograph type
DT
Type of record: Printed text
TR
Contents code: PhD dissertation
CC
Author: Zoran Ovcin
AU
Mentor: Prof. Dr. Nataša Krejić
MN

Title: Quasi Newton Methods for Stochastic Programming Problems
TI

Language of text: Serbian
LT
Language of abstract: English/Serbian
LA
Country of publication: Republic of Serbia
CP
Locality of publication: Vojvodina
LP
Publication year: 2016
PY

Publisher: Author's reprint
PU
Publication place: Novi Sad, Faculty of Sciences, Trg Dositeja Obradovića 4
PP
Physical description: 7/170/69/5/0/8/0
(chapters/pages/literature/tables/pictures/graphics/appendices)
PD
Scientific field: Mathematics
SF
Scientific discipline: Numerical mathematics
SD
Subject / Key words: Nonlinear optimization, Quasi-Newton, stochastic optimization.
SKW

UC:
Holding data: Library of the Department of Mathematics and Informatics, Novi Sad
HD
Note:
N
Abstract: The problem under consideration is unconstrained minimization problem. The problem in deterministic case is often solved with Quasi Newton methods. In noisy environment, which is considered, new approach for step length along descent direction is used. The new approach combines line search and stochastic approximation method using good characteristics of both enabling better efficiency. The convergence is proved. New step length is tested with three descent directions. Many standard test problems show the efficiency of the method. Also, a new, affordable procedure based on application of the fixed Newton method for a sequence of equilibrium problems generated by simulation is introduced. The convergence conditions of the method are derived. The numerical results show a clear difference in the quality of information obtained by solving a sequence of problems if compared with the single equilibrium problem. In the first part general theoretical introduction is given. In the second part a survey of results from scientific community is given together with original results. The third part contains many numerical tests of new methods that show its efficiency.

AB

Accepted by Scientific Board on: September 23, 2010
ASB
Defended:
DE
Thesis defend board:
President: Zorana Lužanin, PhD, Full Professor, Faculty of Sciences, University of Novi Sad

Member: Nataša Krejić, PhD, Full Professor, Faculty of Sciences, University of Novi Sad

Member: Zorica Uzelac, PhD, Full Professor, Faculty of Technical Sciences, University of Novi Sad

Member: Irena Stojkovska, PhD, Associate Professor, Faculty of Natural Sciences and Mathematics, Ss. Cyril and Methodius University in Skopje, Republic of Macedonia

DB

UNIVERZITET U NOVOM SADU
PRIRODNO–MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj:
RBR
Identifikacioni broj:
IBR
Tip dokumentacije: Monografska dokumentacija
TD
Tip zapisa: Tekstualni štampani materijal
TZ
Vrsta rada: Doktorska disertacija
VR
Autor: Zoran Ovcin
AU
Mentor: Prof. dr Nataša Krejić
MN

Naslov rada: Kvazi Njutnovi postupci za probleme stohastičkog programiranja
NR

Jezik publikacije: engleski
JP
Jezik izvoda: engleski/srpski
JI
Zemlja publikovanja: Republika Srbija
ZP
Uže geografsko područje: Vojvodina
UGP
Godina: 2016.
GO

Izdavač: Autorski reprint
IZ
Mesto i adresa: Novi Sad, Prirodno–matematički fakultet, Trg Dositeja Obradovića 4
MA
Fizički opis rada: 7/170/69/5/0/8/0
(broj poglavlja/strana/lit. citata/tabela/slika/grafika/priloga)
FO
Naučna oblast: Matematika
NO
Naučna disciplina: Numerička matematika
ND
Predmetna odrednica/Ključne reči: Nelinearna optimizacija, Kvazi Njutnove metode, stohastička optimizacija.
PO

UDK:
Čuva se: u biblioteci Departmana za matematiku i informatiku, Novi Sad
ČU
Važna napomena:
VN
Izvod: Posmatra se problem minimizacije bez ograničenja. U determinističkom slučaju ti problemi se uspešno rešavaju iterativnim Kvazi Njutnovim postupcima. Ovde se istražuje stohastički slučaj, kada su poznate vrednosti funkcije cilja i njenog gradijenta na koje je uticao šum. Koristi se novi način određivanja dužina koraka, koji kombinuje metod linijskog pretraživanja i metod stohastičke aproksimacije tako da zadrži dobre osobine oba pristupa i obezbedi veću efikasnost postupka. Metod je testiran u kombinaciji sa više načina izbora pravca u iterativnom postupku. Dokazana je konvergencija novog postupka i testiranjem na velikom broju standardnih test problema pokazana njegova efikasnost. Takođe se za rešavanje problema ekvilibriuma u Neoklasičnoj ekonomiji predlaže i dokazuje konvergencija jednog Fiksnog Njutnovog postupka. U zadatku nalaženja rešenja za niz problema kojima se preciznije modelira slučajni sistem, ovaj Fiksni Njutnov postupak ostvaruje veliku uštedu CPU vremena u odnosu na Njutnov metod. U prvom delu teze je dat opšti teoretski uvod. U drugom delu je dat

pregled relevantnih rezultata iz posmatranih oblasti zajedno sa dva originalna rezultata. U trećem delu su dati rezultati numeričkih testova.
IZ

Datum prihvatanja teme od strane NN Veća: 23. IX 2010.
DP
Datum odbrane:
DO
Članovi komisije:
Predsednik: dr Zorana Lužanin, redovni profesor, Prirodno–matematički fakultet, Univerzitet u Novom Sadu

Član: dr Nataša Krejić, redovni profesor, Prirodno–matematički fakultet, Univerzitet u Novom Sadu

Član: dr Zorica Uzelac, redovni profesor, Fakultet tehničkih nauka, Univerzitet u Novom Sadu

Član: dr Irena Stojkovska, vanredni profesor, Prirodno–matematički fakultet, Univerzitet Sveti Kiril i Metodije u Skoplju, Makedonija

Član:
KO

Biografija

Rođen sam 8. III 1966. godine u Somboru. Osnovnu i srednju školu sam završio u Somboru kao vukovac. Odslužio sam vojni rok i od 1986. godine pohađao nastavu na Prirodno–matematičkom fakultetu u Novom Sadu, Matematika, smer Informatika.

Diplomirao sam 1990. godine sa prosekom 8.94. Iste godine sam upisao Magistarske studije Matematike i odbranio seminarske radove položio sve ispite i sa ocenom 10. Magistarsku tezu pod naslovom "Varijacioni princip u fazi metričkim prostorima" sam odbranio 1995. godine.

Od 1991. do 2014. godine sam radio kao asistent na svim predmetima Matematike na Fakultetu tehničkih nauka u Novom Sadu. Od 2014. radim kao predavač strukovnih studija na Fakultetu tehničkih nauka u Novom Sadu.

Učestvovao sam na više naučnih projekata Ministarstva za nauku Srbije. Koautor sam više naučnih radova, učestvovao sam na nekoliko međunarodnih konferencija, koautor sam nekoliko zbirki zadataka.

Novi Sad, Maj 2016. Zoran Ovcin

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

Zoran Ovcin

Kvazi Njutnovi postupci za probleme stohastičkog programiranja

doktorska disertacija

Novi Sad, 2016. godina

Sadržaj

Zahvalnica

Koristim priliku da se zahvalim porodici i prijateljima što su me podržali i bili uz mene dok sam pravio ovu tezu.

Posebno zadovoljstvo mi je da se zahvalim Mentoru, Nataši Krejić, na ogromnom strpljenju i podršci tokom duge izrade ove teze.

Zoran Ovcin

Izvod

Posmatra se problem minimizacije bez ograničenja. U determinističkom slučaju ti problemi se uspešno rešavaju iterativnim Kvazi Njutnovim postupcima. Ovde se istražuje stohastički slučaj, kada su poznate vrednosti funkcije cilja i njenog gradijenta na koje je uticao šum. Koristi se novi način određivanja dužina koraka, koji kombinuje metod linijskog pretraživanja i metod stohastičke aproksimacije tako da zadrži dobre osobine oba pristupa i obezbedi veću efikasnost postupka. Metod je testiran u kombinaciji sa više načina izbora pravca u iterativnom postupku. Dokazana je konvergencija novog postupka i testiranjem na velikom broju standardnih test problema pokazana njegova efikasnost. Takođe se za rešavanje problema ekvilibriuma u Neoklasičnoj ekonomiji predlaže i dokazuje konvergencija jednog Fiksnog Njutnovog postupka. U zadatku nalaženja rešenja za niz problema kojima se preciznije modelira slučajni sistem, ovaj Fiksni Njutnov postupak ostvaruje veliku uštedu CPU vremena u odnosu na Njutnov metod. U prvom delu teze je dat opšti teoretski uvod. U drugom delu je dat pregled relevantnih rezultata iz posmatranih oblasti zajedno sa dva originalna rezultata. U trećem delu su dati rezultati numeričkih testova.

Abstract

The problem under consideration is unconstrained minimization problem. The problem in deterministic case is often solved with Quasi Newton methods. In noisy environment, which is considered, new approach for step length along descent direction is used. The new approach combines line search and stochastic approximation method using good characteristics of both enabling better efficiency. The convergence is proved. New step length is tested with three descent directions. Many standard test problems show the efficiency of the method. Also, a new, affordable procedure based on application of the fixed Newton method for a sequence of equilibrium problems generated by simulation is introduced. The convergence conditions of the method are derived. The numerical results show a clear difference in the quality of information obtained by solving a sequence of problems if compared with the single equilibrium problem. In the first part general theoretical introduction is given. In the second part a survey of results from scientific community is given together with original results. The third part contains many numerical tests of new methods that show its efficiency.