



UNIVERZITET U NOVOM SADU
PRIRODNO-MATEMATIČKI FAKULTET
DEPARTMAN ZA
MATEMATIKU I INFORMATIKU



Andrea Rožnjik

**Optimizacija problema sa stohastičkim
ograničenjima tipa jednakosti – kazneni
metodi sa promenljivom veličinom uzorka**

- doktorska disertacija -

Novi Sad, 2018.

Predgovor

U mnogim realnim problemima javlja se neizvesnost (na primer pojava koja zavisi od ishoda u budućnosti, ali može da potiče i od grešaka merenja). Neizvesnosti tog tipa se često modeliraju pomoću slučajnih promenljivih i tako formiraju probleme optimizacije u kojima je funkcija cilja i/ili dopustiv skup stohastička veličina. Kako je veoma teško pristupiti rešavanju takvog stohastičkog problema, često se slučajnost eliminiše tako što se umesto stohastičke veličine posmatra njeno očekivanje. Time se dobija problem optimizacije s funkcijom cilja i/ili funkcijama ograničenja u obliku matematičkog očekivanja, te je on deterministički problem. Zbog neretke nedostupnosti pronalaznja analitičkog oblika matematičkog očekivanja, pethodni deterministički problem se rešava stohastičkim postupkom, stoga se i naziva problemom stohastičkog programiranja.

Poslednjih nekoliko decenija su problemi stohastičkog programiranja sve zastupljeniji u naučnom istraživanju. Pokazano je da se aproksimiranjem matematičkog očekivanja uzoračkim očekivanjem može doći do rešenja problema stohastičkog programiranja, bilo da su u pitanju problemi bez ograničenja ili sa ograničenjima. Radi postizanja dobre aproksimacije matematičkog očekivanja, potrebno je raditi s veoma velikim uzorcima, što dovodi do velikog broja izračunavanja funkcija. U slučaju kada je evaluacija funkcija skupa, broj izračunavanja funkcija predstavlja dominantni trošak postupka optimizacije, te se teži njegovom redukovanju. Pokazalo se da su metodi s promenljivom veličinom uzorka efikasni u redukciji troškova kada su u pitanju problemi stohastičkog programiranja bez ograničenja. Suštinska ideja tih postupaka je da se tokom iterativnog postupka koriste podskupovi velikog uzorka umesto njega samog. Velika klasa problema podrazumeva i stohastička ograničenja, te je zbog toga značajno istražiti mogućnost primene i adaptacije metoda s promenljivom veličinom uzorka na probleme stohastičkog programiranja s ograničenjima.

U disertaciji je stavljen akcenat na razmatranje problema stohastičkog programiranja s ograničenjima tipa jednakosti. Formirana su dva iterativna postupka za njegovo rešavanje. Osnovna ideja tih postupaka je aproksimiranje

matematičkog očekivanja na osnovu metoda s promenljivom veličinom uzorka uz upotrebu adaptivnog ažuriranja veličine uzorka s ciljem da se dođe do rešenja posmatranog problema, odnosno njegove aproksimacije, sa što manjim brojem izračunavanja vrednosti funkcija. S obzirom na to da je u pitanju problem s ograničenjima, taj pristup je kombinovan s kaznenim postupkom. Kao osnovna strategija iterativnog postupka korišćeno je linijsko pretraživanje. Za oba postupka su sprovedena teorijska analiza i numeričko testiranje.

Disertacija je organizovana u osam poglavlja. Uvodno poglavlje sadrži kratak pregled definicija i teorema iz oblasti teorije verovatnoće, funkcionalne analize i linearne algebre, kao i listu oznaka. U drugom poglavlju se osvrćemo na optimizaciju determinističkih problema, s posebnim naglaskom na probleme s ograničenjima. Treće poglavlje je posvećeno stohastičkoj optimizaciji, s najviše detalja o SAA (Sample Average Approximation) i VSS (Variable Sample Size) metodama. Originalni doprinos disertacije je izložen u narednim poglavljima. U četvrtom poglavlju je prikazan postupak za rešavanje problema stohastičkog programiranja koji je sveden na rešavanje SAA aproksimacije problema. Pokazana je konvergencija podniza iteracija tog algoritma ka rešenju SAA aproksimacije. Kako je u pitanju SAA aproksimacija, u ovom poglavlju se računa s ograničenim uzorkom koji je definisan pre sprovođenja iterativnog postupka za optimizaciju. Rešavanje samog problema stohastičkog programiranja je rađeno u petom poglavlju. Korišćen je niz SAA ocena i adaptivni metod za određivanje veličine uzorka. Neograničen uzorak obezbeđuje da se dobije skoro sigurna konvergencija ka rešenju problema stohastičkog programiranja. U šestom poglavlju su ilustrovani numerički rezultati koji potvrđuju efikasnost oba algoritma, kao i smanjenje troškova u poređenju s ostalim relevantnim postupcima. Kad je u pitanju ograničen uzorak, relevantni postupci su SAA postupak sa maksimalnim uzorkom i heuristički postupak (unapred zadata šema uvećanja uzorka do maksimalnog uzorka). U slučaju algoritma s neograničenim uzorkom poredilo se s heurističkim postupkom s neograničenim uzorkom. Zaključna razmatranja su formulisana u sedmom poglavlju, dok su, radi potpunosti, u dodatku navedeni razmatrani test problemi i prikazani grafici s rezultatima numeričkog testiranja koji potvrđuju reprezentativnost grafika odabranih za šesto poglavlje.

Ovom prilikom bih da zahvalim mentoru Nataši Krklec Jerinkić na korisnim savetima i izuzetnoj pomoći u izradi disertacije. Posebno sam zahvalna profesorki Nataši Krejić na savetima, sugestijama i primedbama. Profesorki Sanji Rapajić dugujem zahvalnost za sugestije kojima je doprinela poboljšanju kvaliteta disertacije. Zahvaljujem i profesoru Zoranu Ovcinu za primedbe i predloge. Veliko hvala porodici i kolegama na pruženoj podršci i bodrenju.

Sadržaj

Predgovor	i
1. Pregled pratećih definicija i teorema	1
1.1. Linearna algebra i funkcionalna analiza	1
1.2. Teorija verovatnoće i statistika	4
2. Deterministička optimizacija	13
2.1. Optimizacija problema bez ograničenja	13
2.1.1. Uslovi optimalnosti	14
2.1.2. Metodi rešavanja	15
2.2. Optimizacija problema s ograničenjima	20
2.2.1. Uslovi optimalnosti	21
2.2.2. Metodi rešavanja	23
3. Stohastička optimizacija	27
3.1. Problem stohastičkog programiranja	27
3.2. Metodi rešavanja problema stohastičkog programiranja	31
3.2.1. SA metod	31
3.2.2. SAA metod	35
3.2.3. Metodi s promenljivom veličinom uzorka	41
3.3. Problemi s ograničenjima u stohastičkom programiranju	44
4. Kazneni postupak s promenljivom veličinom uzorka za rešavanje SAA problema	47
4.1. Algoritam	48
4.2. Analiza konvergenije	54
5. Kazneni postupak s promenljivom veličinom uzorka za rešavanje problema s ograničenjima u formi matematičkog očekivanja	65
5.1. Algoritam	66

5.2. Analiza konvergencije	73
6. Numerički rezultati	89
6.1. Implementacija algoritama	89
6.1.1. Ponašanje veličine uzorka, njegove donje granice i kaznenog parametra	92
6.2. Kazneni VSS postupak za rešavanje SAA problema	94
6.3. Kazneni VSS postupak za rešavanje problema stohastičkog programiranja	97
6.3.1. Analiza rasta veličine uzorka	99
7. Zaključak	103
8. Dodatak	105
Literatura	119
Biografija	125
Ključna dokumentacijska informacija	127

Glava 1

Pregled pratećih definicija i teorema

Istraživanje u okviru disertacije posvećeno je optimizaciji problema sa stohastičkim ograničenjima. To znači da se u tim problemima javlja slučajna promenljiva. Često se problemi sa stohastičkim funkcijama prevode u probleme s funkcijama u obliku matematičkog očekivanja, čime se dobijaju deterministički problemi. Međutim, ti problemi se rešavaju stohastičkim postupcima optimizacije. Tako su istraživanja numeričkih postupaka za rešavanje problema stohastičkog programiranja, pored elemenata analize i linearne algebre, prožeta teorijom verovatnoće, statistike i funkcionalne analize.

S ciljem da obezbedimo lakše praćenje analize rezultata prikazanih u disertaciji ovo poglavlje je posvećeno kratkom pregledu definicija i teorema iz navedenih oblasti, koristeći se s [1,2,23,41,44,55,63].

1.1 Linearna algebra i funkcionalna analiza

Prvo navodimo oznake za prostore koje ćemo spominjati u disertaciji. Skup prirodnih brojeva je označen sa \mathbb{N} , a proširen s nulom označen je sa \mathbb{N}_0 . Skup realnih brojeva označavamo sa \mathbb{R} , dok skup pozitivnih realnih brojeva sa \mathbb{R}_+ . Slova m, n, p nam predstavljaju dimenzije elemenata posmatranih prostora, tako da su $m, n, p \in \mathbb{N}$. Tako nam je \mathbb{R}^n skup realnih vektora dužine n , a $\mathbb{R}^{m \times n}$ skup realnih matrica tipa $m \times n$.

Vektor $x \in \mathbb{R}^n$, s komponentama x_1, \dots, x_n , je vektor kolone i zapisujemo ga sa $x = (x_1, \dots, x_n)^T$, pri čemu je sa \cdot^T označena operacija transponovanja. Relacija poretka je definisana po komponentama, te se poređenje vektora vrši po komponentama. Tako na primer $x \leq y$, gde je $x, y \in \mathbb{R}^n$, $x = (x_1, \dots, x_n)^T$,

$y = (y_1, \dots, y_n)^T$, znači da je $x_i \leq y_i$ za svako $i = 1, \dots, n$. Prostor \mathbb{R}^n je snabdeven skalarnim proizvodom

$$x^T y = \sum_{i=1}^n x_i y_i, \quad x, y \in \mathbb{R}^n,$$

a samim tim i Euklidskom normom

$$\|x\| = \sqrt{x^T x},$$

i u njemu važi Koši-Švarcova nejednakost

$$|x^T y| \leq \|x\| \cdot \|y\| \quad \text{za svako } x, y \in \mathbb{R}^n.$$

Teorema 1.1. Norma $x \mapsto \|x\|$ je neprekidna funkcija na \mathbb{R}^n , a skalarni proizvod $(x, y) \mapsto x^T y$ je neprekidna funkcija na $\mathbb{R}^n \times \mathbb{R}^n$.

Vektori $v_1, \dots, v_m \in \mathbb{R}^n$ su linearno nezavisni ako samo za skalare $\alpha_1 = \dots = \alpha_m = 0$ važi jednakost $\alpha_1 v_1 + \dots + \alpha_m v_m = 0$.

Neka je $A \in \mathbb{R}^{m \times n}$. Rang matrice A , $r(A)$, je maksimalan broj linearno nezavisnih vektora vrsta (ili kolona) matrice A . Matrica A ima pun rang ako je $r(A) = \min\{m, n\}$.

Posmatrajmo sada kvadratnu matricu A reda n , odnosno neka je $A \in \mathbb{R}^{n \times n}$. Za matricu A kažemo da je regularna ako je punog ranga. Ako je A regularna matrica, tada postoji inverzna matrica A^{-1} takva da je $A^{-1}A = AA^{-1} = I$, gde je I jedinična matrica. Skalar λ je karakteristični koren matrice A ako postoji $x \in \mathbb{R}^n$, $x \neq 0$ takav da je $Ax = \lambda x$.

Definicija 1.1. Neka su $\lambda_1, \lambda_2, \dots, \lambda_n$ karakteristični koreni matrice $A \in \mathbb{R}^{n \times n}$. Spektralni radijus $\rho(A)$ matrice A je

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|.$$

Euklidska norma za matricu $A \in \mathbb{R}^{n \times n}$ je

$$\|A\| = \sqrt{\rho(A^T A)}.$$

Za nju važi da je

$$\|A\| = \max_{\|x\|=1} \|Ax\|,$$

pa je $\|Ax\| \leq \|A\| \|x\|$ za svako $x \in \mathbb{R}^n$.

Za nas su od posebnog interesa osobine za simetrične i pozitivno definitne matrice. Matrica A je simetrična ako važi da je $A^T = A$.

Definicija 1.2. Matrica $A \in \mathbb{R}^{n \times n}$ je pozitivno semidefinitna ako važi da je

$$x^T A x \geq 0 \quad \text{za svako } x \in \mathbb{R}^n,$$

a pozitivno definitna ako postoji $\alpha \in \mathbb{R}_+$ za koje važi da je

$$x^T A x \geq \alpha x^T x \quad \text{za svako } x \in \mathbb{R}^n.$$

Teorema 1.2. Simetrična matrica $A \in \mathbb{R}^{n \times n}$ je pozitivno definitna ako i samo ako su svi njeni karakteristični koreni pozitivni.

Teorema 1.3. Ako su $\lambda_1, \lambda_2, \dots, \lambda_n$ karakteristični koreni simetrične matrice $A \in \mathbb{R}^{n \times n}$ poređani po veličini u rastućem poretku, tada za svaki vektor $x \in \mathbb{R}^n$ važi da je

$$\lambda_1 x^T x \leq x^T A x \leq \lambda_n x^T x.$$

Za simetričnu pozitivno definitnu matricu $A \in \mathbb{R}^{n \times n}$ važi da je $\|A\| = \rho(A)$.

Napomenuli bismo da u disertaciji posmatramo probleme u kojima imamo realne funkcije i vektore funkcija s realnim komponentama i da su promenljive iz skupa \mathbb{R}^n , te za nas kompaktan skup predstavlja zatvoren i ograničen skup. Za dva puta diferencijabilnu funkciju $f : \mathbb{R}^n \rightarrow \mathbb{R}$ sa ∇f je označen gradijent,

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix},$$

a sa $\nabla^2 f$ Hesijan, odnosno Jakobijeva matrica funkcije ∇f . Sa ∇f označavamo i Jakobijevu matricu diferencijabilne funkcije $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $m > 1$, pa je tada

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \cdots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \cdots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix},$$

pri čemu je $f(x) = (f_1(x), \dots, f_m(x))^T$.

Navodimo još teoremu srednje vrednosti, ali pre toga da je duž $[a, b]$ koja spaja tačke a i b iz \mathbb{R}^n definisana sa

$$[a, b] = \{(1-t)a + tb : t \in [0, 1]\}.$$

Teorema 1.4. Neka je D otvoren skup u \mathbb{R}^n , $[a, b] \subseteq D$ i $f : D \rightarrow \mathbb{R}$ diferencijabilna funkcija, tada postoji $t \in (0, 1)$ takvo da je

$$f(b) - f(a) = \nabla^T f((1-t)a + tb)(b - a).$$

1.2 Teorija verovatnoće i statistika

U disertaciji koristimo pojmove slučajna promenljiva, odnosno slučajan vektor, matematičko očekivanje, skoro sigurna konvergencija, jaki zakon velikih brojeva. Sad ćemo navesti njihove definicije i poneke osobine.

Prvo moramo da vidimo šta je prostor verovatnoće, a za to je potrebno da krenemo od eksperimenta koji predstavlja osnovni model u teoriji verovatnoće. Mogući ishod nekog eksperimenta je elementarni događaj ω , a Ω je skup svih mogućih elementarnih događaja posmatranog eksperimenta. Slučajan događaj, odnosno događaj, je podskup skupa Ω . Pod realizacijom slučajnog događaja se podrazumeva da se tokom eksperimenta ostvario elementarni događaj koji mu pripada.

Neka je \mathcal{F} familija slučajnih događaja eksperimenta definisanog sa skupom elementarnih događaja Ω .

Definicija 1.3. Podskup \mathcal{F} partitivnog skupa skupa Ω je σ -algebra nad Ω ako su zadovoljeni uslovi:

1. $\Omega \in \mathcal{F}$,
2. ako $A \in \mathcal{F}$, tada i $\Omega \setminus A \in \mathcal{F}$,
3. ako je $\{A_i\}_{i \in \mathbb{N}} \subseteq \mathcal{F}$, tada $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

Skup (Ω, \mathcal{F}) je prostor sa σ -algebrom, a elementi σ -algebre \mathcal{F} su merljivi skupovi. Najmanja σ -algebra koja sadrži skup \mathbb{R}^n je Borelova σ -algebra $\mathcal{B}_{\mathbb{R}^n}$, koja je generisana otvorenim, zatvorenim ili poluotvorenim intervalima.

Definicija 1.4. Neka je \mathcal{F} σ -algebra nad Ω . Funkcija $P: \mathcal{F} \rightarrow [0, 1]$ je funkcija verovatnoće ako važi da je

1. $P(\Omega) = 1$,
2. ako $A_i \in \mathcal{F}$, $i = 1, 2, \dots$ i $A_i \cap A_j = \emptyset$ za svako $i \neq j$, $i, j = 1, 2, \dots$, tada je $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Uređena trojka (Ω, \mathcal{F}, P) je prostor verovatnoće, a $P(A)$ verovatnoća slučajnog događaja A .

Događaj Ω je siguran događaj jer se uvek realizuje. Događaj A s osobinom da je $A \neq \Omega$ i $P(A) = 1$ je skoro siguran (s.s.) događaj.

Dolazimo do pojmova slučajna promenljiva i slučajan vektor, kao i funkcije raspodele koja ih potpuno određuje, ali i numeričkih karakteristika: kvantil, matematičko očekivanje i varijansa.

Definicija 1.5. *n -dimenzionalana slučajna promenljiva (\mathcal{F} -merljivo preslikavanje) X nad prostorom verovatnoće (Ω, \mathcal{F}, P) je preslikavanje $X : \Omega \rightarrow \mathbb{R}^n$ za koje važi da*

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \quad \text{za svako } B \in \mathcal{B}_{\mathbb{R}^n}.$$

Teorema 1.5. *Preslikavanje $X = (X_1, \dots, X_n)^T$, $X : \Omega \rightarrow \mathbb{R}^n$, je slučajan vektor (n -dimenzionalna slučajna promenljiva) ako i samo ako za svako $i = 1, \dots, n$ važi da je X_i jednodimenzionalana slučajna promenljiva.*

Definicija 1.6. *Neka je $X : \Omega \rightarrow \mathbb{R}^n$ slučajan vektor X nad prostorom verovatnoće (Ω, \mathcal{F}, P) . Funkcija*

$$P_X(B) = P\{X \in B\} = P\{\omega \in \Omega : X(\omega) \in B\} \quad \text{za svako } B \in \mathcal{B}_{\mathbb{R}^n}$$

je raspodela verovatnoće slučajnog vektora X .

Definicija 1.7. *Funkcija raspodele slučajne promenljive X nad prostorom verovatnoće (Ω, \mathcal{F}, P) je funkcija $F_X : \mathbb{R} \rightarrow [0, 1]$ definisana sa*

$$F_X(x) = P\{X \leq x\} = P\{\omega \in \Omega : X(\omega) \leq x\}.$$

Definicija 1.8. *Funkcija raspodele slučajnog vektora $X = (X_1, \dots, X_n)^T$ nad prostorom verovatnoće (Ω, \mathcal{F}, P) je funkcija $F_X : \mathbb{R}^n \rightarrow [0, 1]$ definisana sa*

$$\begin{aligned} F_X(x) &= F_X(x_1, \dots, x_n) = P\{\{X_1 \leq x_1\} \cap \dots \cap \{X_n \leq x_n\}\} = \\ &= P\{\omega \in \Omega : \{X_1(\omega) \leq x_1\} \cap \dots \cap \{X_n(\omega) \leq x_n\}\}. \end{aligned}$$

Za $\alpha \in (0, 1)$ definisan je kvantil reda α slučajne promenljive X kao

$$z_\alpha = \inf_{F_X(x) \geq \alpha} x.$$

Definicija 1.9. *Matematičko očekivanje (očekivanje, očekivana vrednost, srednja vrednost) slučajne promenljive X nad prostorom verovatnoće (Ω, \mathcal{F}, P) je*

$$E[X] = \int_{\Omega} X dP.$$

Za slučajnu promenljivu X se kaže da je P-integrabilna ako je $E[X]$ dobro definisano i konačno.

Definicija 1.10. *Varijansa (disperzija ili centralni momenat 2. reda) slučajne promenljive X nad prostorom verovatnoće (Ω, \mathcal{F}, P) je*

$$D[X] = E[(X - E[X])^2].$$

Veličina $\sigma(X) = \sqrt{D[X]}$ je standardno odstupanje slučajne promenljive X .

Neka su X i Y slučajne promenljive i c konstanta, tada važe osobine:

1. $D[X] = E[X^2] - (E[X])^2$
2. $D[X] \geq 0$
3. $E[c] = c, \quad D[c] = 0$
4. $E[cX] = cE[X], \quad D[cX] = c^2D[X]$
5. $E[X + Y] = E[X] + E[Y]$.

Pored toga, ako slučajna promenljiva ima varijansu, tada ima i očekivanu vrednost.

Matematičko očekivanje slučajnog vektora $X = (X_1, \dots, X_n)^T$ je vektor

$$E[X] = (E[X_1], \dots, E[X_n])^T.$$

Analogno, varijansa slučajnog vektora X je

$$D[X] = (D[X_1], \dots, D[X_n])^T.$$

Kada je slučajna promenljiva X diskretnog tipa, što znači da joj je skup slika $\{x_1, x_2, \dots\}$ najviše prebrojiv skup, tada je njeno matematičko očekivanje

$$E[X] = \sum_{i=1}^{\infty} x_i p(x_i),$$

gde je verovatnoća $p(x_i) = P\{X = x_i\}$. Očekivanje $E[X]$ postoji ako je suma s kojom je određeno apsolutno konvergentna, odnosno ako je

$$\sum_{i=1}^{\infty} |x_i| p(x_i) < \infty.$$

Slučajna promenljiva X je apsolutno neprekidnog tipa ako postoji nene-
gativna integrabilna funkcija $\varphi_X(x)$ (gustina raspodele verovatnoće) sa $x \in \mathbb{R}$
takva da je

$$P\{X \in B\} = \int_B \varphi_X(x) dx \quad \text{za svako } B \in \mathcal{B}_{\mathbb{R}}.$$

Njeno matematičko očekivanje je

$$E[X] = \int_{-\infty}^{\infty} x \varphi_X(x) dx$$

i ono postoji ako integral koji ga određuje apsolutno konvergira, odnosno ako
važi da je

$$\int_{-\infty}^{\infty} |x| \varphi_X(x) dx < \infty.$$

Očekivana vrednost slučajne promenljive koja je transformacija neke druge
slučajne promenljive može se izračunati preko gustine raspodele te slučajne
promenljive. Ova osobina važi za Borelove funkcije, tj. za funkciju $f: \mathbb{R} \rightarrow \mathbb{R}$
za koju važi da je $f^{-1}(B) \in \mathcal{B}_{\mathbb{R}}$ za svako $B \in \mathcal{B}_{\mathbb{R}}$.

Teorema 1.6. *Neka je slučajna promenljiva X apsolutno neprekidnog tipa s
gustinom raspodele $\varphi_X(x)$ i neka je funkcija $f: \mathbb{R} \rightarrow \mathbb{R}$ Borelova funkcija.
Tada važi da je*

$$E[f(X)] = \int_{-\infty}^{\infty} f(x) \varphi_X(x) dx.$$

Za nas su posebno interesantne nezavisne slučajne promenljive.

Definicija 1.11. *Slučajne promenljive X i Y definisane nad istim prostorom
verovatnoće (Ω, \mathcal{F}, P) su nezavisne ako su njima generisane σ -algebre neza-
visne, odnosno ako za svako $A, B \in \mathcal{B}_{\mathbb{R}}$ važi da je*

$$P(X^{-1}(A) \cap X^{-1}(B)) = P(X^{-1}(A))P(X^{-1}(B)).$$

Za nezavisne slučajne promenljive X_1, \dots, X_n , $n \in \mathbb{N}$, važi da je njihova
zajednička funkcija raspodele, odnosno funkcija raspodele slučajnog vektora
 $X = (X_1, \dots, X_n)^T$ je

$$F_X(x_1, \dots, x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n).$$

Pored toga,

$$E[X_1 \dots X_n] = E[X_1] \dots E[X_n] \quad \text{i} \quad D[X_1 + \dots + X_n] = D[X_1] + \dots + D[X_n].$$

Teorema 1.7. *Neka su $f, g : \mathbb{R} \rightarrow \mathbb{R}$ Borelove funkcije. Ako su slučajne promenljive X i Y nezavisne, tada su i $f(X)$ i $g(Y)$ nezavisne slučajne promenljive.*

U stohastičkoj optimizaciji su značajni pojmovi skoro sigurna konvergencija i konvergencija u raspodeli, kao i zakoni velikih brojeva, ali pre nego što pređemo na te pojmove, navešćemo neke osobine slučajne promenljive s normalnom raspodelom.

Normalna raspodela $\mathcal{N}(\mu, \sigma^2)$, sa $\mu \in \mathbb{R}$ i $\sigma > 0$, je određena s gustinom raspodele

$$\varphi_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

Vrednost μ je očekivana vrednost, a σ^2 varijansa. Raspodela $\mathcal{N}(0, 1)$ se naziva standardnom normalnom raspodelom. Ako je X slučajna promenljiva, tada se za slučajnu promenljivu

$$X^* = \frac{X - E[X]}{\sigma(X)}$$

kaže da je standardizovana i važi da je $E[X^*] = 0$, $D[X^*] = 1$. Ako slučajna promenljiva X ima $\mathcal{N}(\mu, \sigma^2)$ raspodelu, tada X^* ima $\mathcal{N}(0, 1)$ raspodelu.

Neka su X_1, X_2, \dots slučajne promenljive sa istom raspodelom nad datim prostorom verovatnoće (Ω, \mathcal{F}, P) . Razlikujemo sledeće konvergencije niza $\{X_k\}_{k \in \mathbb{N}}$ ka slučajnoj promenljivoj X .

Definicija 1.12. *Niz slučajnih promenljivih $\{X_k\}$ skoro sigurno (s.s.) konvergira ka slučajnoj promenljivoj X ako je*

$$P \left\{ \lim_{k \rightarrow \infty} |X_k - X| = 0 \right\} = 1.$$

Definicija 1.13. *Niz slučajnih promenljivih $\{X_k\}$ konvergira u verovatnoći ka slučajnoj promenljivoj X ako za svako $\varepsilon > 0$*

$$\lim_{k \rightarrow \infty} P\{|X_k - X| \geq \varepsilon\} = 0.$$

Definicija 1.14. *Neka slučajne promenljive X_1, X_2, \dots imaju svoje funkcije raspodele $F_{X_1}(x), F_{X_2}(x), \dots$, a slučajna promenljiva X funkciju raspodele $F_X(x)$. Niz $\{X_k\}$ konvergira u raspodeli ka X ako*

$$\lim_{k \rightarrow \infty} F_{X_k}(x) = F_X(x)$$

za svaku tačku x u kojoj je funkcija F_X neprekidna.

Definicija 1.15. Niz slučajnih promenljivih $\{X_k\}$, s osobinom da je $E[X_k^2] < \infty$ za svako $k = 1, 2, \dots$, konvergira srednje kvadratno ka slučajnoj promenljivoj X ako

$$\lim_{k \rightarrow \infty} E[(X_k - X)^2] = 0.$$

Veze između navedenih konvergencija su formulisane u narednim teoremama.

Teorema 1.8. Ako niz slučajnih promenljivih $\{X_k\}_{k \in \mathbb{N}}$ skoro sigurno konvergira ka slučajnoj promenljivoj X , tada taj niz konvergira i u verovatnoći.

Teorema 1.9. Ako niz slučajnih promenljivih $\{X_k\}_{k \in \mathbb{N}}$ konvergira u verovatnoći ka slučajnoj promenljivoj X , tada taj niz konvergira i u raspodeli.

Teorema 1.10. Ako niz slučajnih promenljivih $\{X_k\}_{k \in \mathbb{N}}$ konvergira srednje kvadratno ka slučajnoj promenljivoj X , tada taj niz konvergira i u verovatnoći.

Neka su sada slučajne promenljive X_1, X_2, \dots još i nezavisne. Zakoni velikih brojeva se bave pitanjem ponašanja slučajne promenljive S_n/n kada n teži u beskonačno, pri čemu je

$$S_k = X_1 + X_2 + \dots + X_k.$$

Ukoliko se posmatra skoro sigurna konvergencija, radi se o jakim zakonima velikih brojeva, dok se po slabim zakonima velikih brojeva posmatra konvergencija u verovatnoći. Centralne granične teoreme se bave pitanjem kada niz slučajnih promenljivih konvergira u raspodeli ka slučajnoj promenljivoj sa normalnom raspodelom.

Teorema 1.11. Neka su slučajne promenljive X_1, X_2, \dots nezavisne i neka imaju uniformno ograničene varijanse (tj. postoji konstanta C za koju važi da je $D[X_k] \leq C$ za $k \in \mathbb{N}$). Tada

$$\lim_{k \rightarrow \infty} \frac{S_k}{k} = \frac{1}{k} \sum_{k=1}^n E[X_k] \quad \text{u verovatnoći.}$$

Teorema 1.12. Neka su slučajne promenljive X_1, X_2, \dots nezavisne i neka imaju istu raspodelu s konačnom očekivanom vrednošću $E[X_k] = \mu$, $k \in \mathbb{N}$. Tada

$$\lim_{k \rightarrow \infty} \frac{S_k}{k} = \mu \quad \text{s.s.}$$

Dakle, možemo da kažemo da je $\frac{1}{k} \sum_{j=1}^k X_j$ približno jednaka sa matematičkim očekivanjem posmatranih slučajnih promenljivih za dovoljno veliko k , pod navedenim uslovima.

Teorema 1.13. *Neka su slučajne promenljive X_1, X_2, \dots nezavisne i sa istom raspodelom i neka imaju konačnu varijansu $D[X_k] = \sigma^2$, $k \in \mathbb{N}$ (i konačnu očekivanu vrednost $E[X_k] = \mu$, $k \in \mathbb{N}$). Tada za standardizovanu slučajnu promenljivu S_k^* važi da je*

$$\lim_{k \rightarrow \infty} F_{S_k^*}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Dakle, pod navedenim uslovima, niz standardizovanih slučajnih promenljivih $S_k^* = \frac{S_k - E[S_k]}{\sqrt{D[S_k]}}$ konvergira u raspodeli ka slučajnoj promenljivoj sa standardnom normalnom raspodelom $\mathcal{N}(0, 1)$.

Ostalo nam je da se osvrnemo na statističke pojmove: uzorak, uzoračko očekivanje, nepristrasna i konzistentna ocena, interval poverenja.

Populacija, odnosno skup na kojem posmatramo određenu pojavu – obeležje, za nas predstavlja slučajnu promenljivu X sa svojom funkcijom raspodele. Uzorak nam predstavlja niz slučajnih promenljivih X_1, \dots, X_N sa istom funkcijom raspodele kao i slučajna promenljiva X . Taj niz se može posmatrati kao N realizacija slučajne promenljive X . Naime, uzorak je N -dimenzionalni slučajni vektor $(X_1, \dots, X_N)^T$. Statistika je funkcija od uzorka, a samim tim i slučajna promenljiva. Za nas su značajne statistike:

$$\triangleright \text{uzoračko očekivanje} \quad \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i,$$

$$\triangleright \text{uzoračka varijansa} \quad \bar{S}^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2.$$

Prilikom ocene karakteristike θ slučajne promenljive X sa statistikom $\hat{\theta}$ na osnovu uzorka X_1, \dots, X_N značajno je utvrditi kvalitet te ocene. U tome imaju ulogu osobine:

$$\triangleright \text{nepristrasnost (centriranost):} \quad E[\hat{\theta}] = \theta,$$

$$\triangleright \text{konzistentnost (postojanost):} \quad \hat{\theta} \text{ konvergira u verovatnoći ka } \theta \text{ kada } N \rightarrow \infty.$$

Ocena je dobra ukoliko je nepristrasna i konzistentna. Među dvema nepristrasnim ocenama efikasnija je ona koja ima manju varijansu. Kada su X_1, \dots, X_N nezavisne slučajne promenljive sa istom raspodelom, tada su ocene \bar{X} i \bar{S}^2 nepristrasne.

Glava 2

Deterministička optimizacija

Pre nego što se posvetimo optimizaciji problema sa stohastičkim ograničenjima osvrnimo se na determinističku optimizaciju. S obzirom na to da su u osnovi našeg istraživanja problemi s ograničenjima, stavljamo akcenat na njihovo razmatranje. Među postupcima za njihovo rešavanje koncentrisali smo se na kaznene postupke. Suština tih postupaka je da se problem s ograničenjima preformuliše u problem bez ograničenja, te da se rešava niz problema bez ograničenja. Zbog toga je značajno razmotriti i probleme bez ograničenja. Tako počinjemo s posmatranjem uslova optimalnosti problema bez ograničenja, a potom navodimo metode za njihovo rešavanje. Ističemo linijsko pretraživanje pošto taj pristup koristimo u našem istraživanju.

Veliki broj knjiga je posvećen determinističkoj optimizaciji, na primer [8, 9,16,21,39], a kad je u pitanju teorija s primenom, [3,4].

2.1 Optimizacija problema bez ograničenja

Optimizacija, odnosno matematičko programiranje, je matematički alat za analiziranje fizičkih sistema. Proces optimizacije se sastoji od modeliranja problema definisanog fizičkim sistemom, odnosno utvrđivanja promenljivih, funkcije cilja i ograničenja, i rešavanja modela uz pomoć algoritma. Algoritmi u optimizaciji su mahom iterativni, što znači da se njima generiše niz aproksimacija, polazeći od početne aproksimacije i generišući svaku narednu aproksimaciju koristeći prethodnu, s ciljem da se dobije niz tačaka koje konvergiraju ka rešenju problema optimizacije. Kada je model u potpunosti poznat, poseduju se vrednosti funkcija iz modela i rezultat algoritma je isti za iste početne aproksimacije, tada se govori o determinističkoj optimizaciji.

Problemi optimizacije se mogu podeliti na probleme bez ograničenja i na probleme s ograničenjima. Opšta formulacija problema bez ograničenja je

$$\min_{x \in \mathbb{R}^n} f(x) \quad (2.1)$$

gde je funkcija $f : \mathbb{R}^n \rightarrow \mathbb{R}$ funkcija cilja. Problem maksimizacije se ekvivalentno može zapisati preko problema minimizacije, naime

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

U slučaju problema s ograničenjima, promenljiva x pripada nekom podskupu od \mathbb{R}^n .

2.1.1 Uslovi optimalnosti

Pod rešavanjem problema (2.1) podrazumeva se utvrđivanje tačke minimuma x^* u kojoj funkcija f ima najmanju vrednost na skupu \mathbb{R}^n . Dakle,

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

Problem optimizacije je dobro definisan ako ima rešenje, s tim da može da ima više rešenja. Da bi problem minimizacije bez ograničenja bio dobro definisan, funkcija cilja treba da bude ograničena odozdo.

U mnogim slučajevima je veoma teško odrediti minimum funkcije na skupu \mathbb{R}^n zbog posedovanja samo lokalnih informacija o funkciji. Iz tog razloga su definisane sledeće vrste minimuma. Problem (2.1) u tački x^* ima

- ▷ globalni minimum ako je $f(x^*) \leq f(x)$ za svako $x \in \mathbb{R}^n$;
- ▷ (slabi) lokalni minimum ako je $f(x^*) \leq f(x)$ za svako x iz neke okoline tačke x^* ;
- ▷ strogi lokalni minimum ako je $f(x^*) < f(x)$ za svako x , različito od x^* , iz neke okoline tačke x^* .

Za glatke funkcije postoje uslovi koji određuju lokalni minimum. Potreban uslov prvog reda je definisan za neprekidno diferencijabilne funkcije cilja.

Teorema 2.1. ([39]) *Ako je x^* tačka lokalnog minimuma problema (2.1) sa neprekidno diferencijabilnom funkcijom f u nekoj okolini tačke x^* , tada je $\nabla f(x^*) = 0$.*

Tačka u kojoj je gradijent funkcije jednak nuli je stacionarna tačka funkcije. Dakle, potreban uslov da funkcija ima minimum u tački x^* je da je x^* njena stacionarna tačka, odnosno, za tačku koja nije stacionarna tačka funkcije, znamo da ne može biti tačka minimuma.

Potreban uslov drugog reda, kao i dovoljan uslov drugog reda, su dati za dva puta neprekidno diferencijabilne funkcije.

Teorema 2.2. ([39]) *Ako je x^* tačka lokalnog minimuma problema (2.1) sa neprekidno diferencijabilnim gradijentom ∇f u nekoj okolini tačke x^* , tada je $\nabla f(x^*) = 0$ i Hesijan $\nabla^2 f(x^*)$ je pozitivno semidefinitan.*

Teorema 2.3. ([39]) *Ako je Hesijan $\nabla^2 f$ neprekidan u nekoj okolini tačke x^* , $\nabla f(x^*) = 0$ i $\nabla^2 f(x^*)$ je pozitivno definitan, tada je x^* tačka strogo lokalnog minimuma problema (2.1).*

Znači, i za tačku koja nema pozitivno semidefinitan Hesijan znamo da ne može biti tačka minimuma. Međutim, za stacionarnu tačku u kojoj je Hesijan pozitivno definitan znamo da je tačka lokalnog minimuma.

Kada je funkcija cilja konveksna, tada imamo informacije i o globalnom minimumu.

Teorema 2.4. ([9,39]) *Ako je funkcija cilja konveksna, tada je njena tačka lokalnog minimuma i tačka globalnog minimuma.*

Kada je funkcija cilja konveksna i diferencijabilna, tada je svaka stacionarna tačka ujedno i tačka globalnog minimuma.

Ako je funkcija cilja strogo konveksna, tada je njena tačka globalnog minimuma jedinstvena.

2.1.2 Metodi rešavanja

S obzirom na to da su metodi za rešavanje problema optimizacije iterativni, odabere se početna tačka x_0 i generiše se niz iterativnih tačaka $\{x_k\}_{k \in \mathbb{N}}$ tako da se naredna tačka određuje na osnovu prethodne. Cilj je da niz $\{x_k\}$ bude generisan tako da teži ka rešenju x^* problema (2.1). Stoga se izbor naredne iterativne tačke vrši sa ciljem da se dobije manja vrednost funkcije cilja.* Tako, u k -toj iteraciji se bira naredna iterativna tačka x_{k+1} za koju važi da je

*Tačnije, ovaj uslov ne mora da važi za nemonotone algoritme, videti Grippo i dr. [22]. Kod tih algoritama se traži da se tek nakon nekog broja iteracija postigne smanjenje vrednosti funkcije cilja.

$f(x_{k+1}) < f(x_k)$. To se čini odabirom pravca d_k i dužine koraka α_k s kojima će se iz x_k doći u x_{k+1} , što znači da je

$$x_{k+1} = x_k + \alpha_k d_k.$$

Postoje dva osnovna pristupa određivanja pravca i dužine koraka: postupak oblasti poverenja i linijsko pretraživanje.

Po postupku oblasti poverenja prvo se fiksira maksimalna dužina koraka, a potom se odrede pravac pretrage i dužina koraka kojima se dolazi u tačku u kojoj se dovoljno smanji vrednost funkcije cilja. Tako se u k -toj iteraciji fiksira maksimalna dužina koraka u vidu poluprečnika oblasti poverenja $\Delta_k > 0$, te se napravi model m_k koji aproksimira funkciju cilja u okolini x_k . Obično se koristi kvadratni model

$$m_k(p) = f(x_k) + \nabla^T f(x_k)p + \frac{1}{2}p^T H_k p,$$

gde je matrica H_k Hesijan $\nabla^2 f(x_k)$ ili njegova aproksimacija. Pravac pretrage i dužina koraka se određuju istovremeno, (približnim) rešavanjem problema

$$\min_{\|p\| \leq \Delta_k} m_k(p). \quad (2.2)$$

Ukoliko se s tačkom $x_k + p_k$, gde je p_k rešenje potproblema (2.2), dobije dovoljno smanjenje vrednosti funkcije f , ta tačka se uzima za narednu iterativnu tačku, odnosno, tada je $x_{k+1} = x_k + p_k$. U suprotnom se smanjuje poluprečnik oblasti poverenja i rešava novi potproblem. Algoritam i njegove osobine se mogu naći u Nocedal i Wright [39].

Linijsko pretraživanje

Po postupku linijskog pretraživanja u svakoj iteraciji se prvo odredi pravac pretraživanja d_k , a potom dužina koraka α_k tako da se umanju vrednost funkcije cilja. Uzimanjem za pravac pretraživanja opadajući pravac – pravac d_k s osobinom da je

$$d_k^T \nabla f(x_k) < 0,$$

obebeđuje se da se duž pravca d_k može smanjiti vrednost funkcije f . Neki metodi koriste proizvoljan opadajući pravac pretraživanja, dok drugi konkretno definišu izbor pravca pretraživanja. U nastavku ćemo navesti najčešće korišćene pravce, ali ćemo pre toga razmotriti izbor dužine koraka.

Određivanje dužine koraka u linijskom pretraživanju

Dužina koraka s kojom se, duž d_k , dobija najveće umanjeње vrednosti $f(x_k)$ je rešenje problema

$$\min_{\alpha > 0} f(x_k + \alpha d_k). \quad (2.3)$$

Rešavanje potproblema (2.3) je uglavnom skupo, a pored toga je nepotrebno da se u svakoj iteraciji izračuna njegovo tačno rešenje, te se zbog toga, često, α_k određuje proveravanjem vrednosti $f(x_k + \alpha d_k)$ za određen niz vrednosti α . U nekim slučajevima, uglavnom u Njutnovom postupku, se čak fiksira dužina koraka na $\alpha_k = 1$.

Prilikom proveravanja niza vrednosti α postavlja se uslov kojim se određuje prihvatljivost za α . S obzirom na to da je cilj da se umanja vrednost funkcije cilja, uslov koji se postavlja je Armižo uslov

$$f(x_k + \alpha d_k) \leq f(x_k) + \eta \alpha d_k^T \nabla f(x_k), \quad (2.4)$$

gde je parametar $\eta \in (0, 1)$. Praktično se obično uzima da je $\eta = 10^{-4}$.

Armižo uslov je zadovoljen za svako dovoljno malo α , pa niz iteracija $\{x_k\}$ zasnovan samo na njemu može veoma sporo napredovati ka rešenju problema (2.1). Zbog toga se dodaje i uslov krivine

$$d_k^T \nabla f(x_k + \alpha d_k) \geq \eta_2 d_k^T \nabla f(x_k) \quad (2.5)$$

s konstantom $\eta_2 \in (\eta, 1)$. Prilikom linijskih pretraživanja koja nisu jako restriktivna koristi se $\eta_2 = 0.9$, dok se za što preciznija linijska pretraživanja računa sa $\eta_2 = 0.1$. U slučaju Njutnovog i kvazi-Njutnovih metoda uobičajeno je uzeti $\eta_2 = 0.9$.

Armižo uslov i uslov krivine čine Volfove uslove, s ciljem da se smanji vrednost funkcije cilja i da se spreči da dužina koraka bude neprihvatljivo mala. Naredna teorema potvrđuje da je, za neprekidno diferencijabilne i odozdo ograničene funkcije cilja, linijsko pretraživanje zasnovano na Volfovim uslovima dobro definisano, odnosno da se uvek može dobiti dužina koraka koja zadovoljava Volfove uslove.

Teorema 2.5. ([39]) *Neka je funkcija $f : \mathbb{R}^n \rightarrow \mathbb{R}$ neprekidno diferencijabilna i ograničena odozdo duž prave $\{x_k + \alpha d_k : \alpha > 0\}$, gde je d_k opadajući pravac za f u x_k . Tada za parametre η i η_2 s osobinom da je $0 < \eta < \eta_2 < 1$ postoje intervali kojima pripada dužina koraka koja zadovoljava uslove (2.4) i (2.5).*

Algoritam za linijsko pretraživanje zasnovano na Volfovim uslovima za neku maksimalnu dozvoljenu dužinu koraka $\alpha_{max} > 0$ utvrđuje podinterval intervala $(0, \alpha_{max})$ u kojem se nalaze dužine koraka koje zadovoljavaju Volfove uslove, a potom, koristeći interpolaciju, određuje traženu dužinu koraka. Detaljan algoritam može se naći u [16,39].

Ukoliko se Armižo uslov primenjuje s postupkom u kojem niz testiranih vrednosti α opada s unapred definisanim faktorom, dobija se postupak linijskog pretraživanja koji eliminiše prekratke dužine koraka. To je slučaj s takozvanom *backtracking* procedurom. U pitanju je jednostavan postupak sledećeg oblika.

Algoritam 2.1. *Backtracking* linijsko pretraživanje

K0 Izabrati: $\bar{\alpha} > 0, \beta, \eta \in (0, 1)$ i uzeti da je $\alpha = \bar{\alpha}$.

K1 Dokle god je

$$f(x_k + \alpha d_k) > f(x_k) + \eta \alpha d_k^T \nabla f(x_k)$$

stavljati da je $\alpha = \beta \alpha$.

Postaviti da je $\alpha_k = \alpha$.

Backtracking linijsko pretraživanje je takođe dobro definisano jer se zaustavlja za $\alpha = \bar{\alpha}$ ili za dovoljno malo α za koje važi Armižo uslov. Kada se koristi u sklopu Njutnovog ili kvazi-Njutnovog postupka, uzima se početna dužina koraka $\bar{\alpha} = 1$.

Pravci pretraživanja u linijskom pretraživanju

Najčešće korišćeni pravac pretraživanja je

$$d_k = -H_k^{-1} \nabla f(x_k) \quad (2.6)$$

sa simetričnom i regularnom matricom H_k . Kada je H_k pozitivno definitna, a x_k nije stacionarna tačka funkcije f , tada je

$$d_k^T \nabla f(x_k) = -\nabla^T f(x_k) H_k^{-1} \nabla f(x_k) < 0,$$

pa je pravac pretraživanja definisan sa (2.6) opadajući pravac.

Za $H_k = I$ dobija se pravac negativnog gradijenta

$$d_k = -\nabla f(x_k)$$

duž kojeg najbrže opada vrednost funkcije f . Taj pravac se koristi u postupku najbržeg pada koji je najjednostavniji postupak i ne zahteva računanje Hesijana, ali može biti jako spor.

Kada je H_k Hesijan u tački x_k , dobija se Njutnov pravac

$$d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k). \quad (2.7)$$

Kako se duž Njutnovog pravca s dužinom koraka $\alpha_k = 1$ dobija najveće smanjenje vrednosti funkcije f , kada je Hesijan $\nabla^2 f(x_k)$ pozitivno definitan, uzima se da je $\alpha_k = 1$. Zbog toga Njutnov postupak, kod kojeg je niz iteracija $\{x_k\}$ generisan sa $x_{k+1} = x_k + d_k$, gde je d_k određeno sa (2.7), spada u najbrže postupke. S obzirom na to da Hesijan $\nabla^2 f(x_k)$ ne mora biti pozitivno definitan (kada je daleko od rešenja), samim tim pravac (2.7) ne mora da bude opadajući pravac. Tada može da bude čak i singularan. Jedan od načina da se izbegnu ti problemi je da se matrica H_k modifikuje dodavanjem pozitivne dijagonalne matrice na Hesijan $\nabla^2 f(x_k)$, obezbeđujući da postane dovoljno pozitivno definitna.

Kada se želi izbeći računanje Hesijana, a ne želi se upotrebiti metod najbržeg pada, tada su dobar izbor kvazi-Njutnovi postupci. Po kvazi-Njutnovim postupcima se za matricu H_k uzima aproksimacija Hesijana $\nabla^2 f(x_k)$. Pošto je

$$\nabla^2 f(x_k)(x_{k+1} - x_k) \approx \nabla f(x_{k+1}) - \nabla f(x_k),$$

matrica H_k se ažurira u svakoj iteraciji tako da nova matrica H_{k+1} bude simetrična i zadovolji jednačinu sečice

$$H_{k+1}s_k = y_k,$$

gde je

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f(x_{k+1}) - \nabla f(x_k).$$

Jedna od najpoznatijih formula za ažuriranje tog tipa je BFGS formula, koju su predložili Broyden, Fletcher, Goldfarb i Shanno,

$$H_{k+1} = H_k - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}. \quad (2.8)$$

Radi izbegavanja faktorizacije matrice H_k prilikom računanja pravca pretraživanja u (2.6), koristi se ekvivalent formuli (2.8), u kojoj se direktno ažurira matrica $B_k := (H_k)^{-1}$,

$$B_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k} \right) B_k \left(I - \frac{y_k s_k^T}{y_k^T s_k} \right) + \frac{s_k s_k^T}{y_k^T s_k}. \quad (2.9)$$

Za početne matrice H_0 , odnosno B_0 , najčešće se uzima jedinična matrica.

Pravila ažuriranja (2.8) i (2.9) čuvaju pozitivnu definitnost kada je zadovoljen uslov

$$s_k^T y_k > 0. \quad (2.10)$$

Preciznije, ako su H_k i B_k pozitivno definitne matrice i važi (2.10), tada su i matrice H_{k+1} i B_{k+1} pozitivno definitne. Kada je x_{k+1} određeno dužinom koraka α_k koja zadovoljava uslov krivine (2.5), tada važi uslov (2.10). Naime,

$$\begin{aligned} s_k^T y_k &= \alpha_k (d_k^T \nabla f(x_{k+1}) - d_k^T \nabla f(x_k)) > \\ &> \alpha_k (d_k^T \nabla f(x_{k+1}) - \eta_2 d_k^T \nabla f(x_k)) \geq 0, \end{aligned}$$

jer je $\alpha_k > 0$, $\eta_2 \in (0, 1)$ i $d_k^T \nabla f(x_k) < 0$. Dakle, linijsko pretraživanje s dužinom koraka koja zadovoljava Volfove uslove i BFGS pravcem s pozitivno definitnom matricom H_0 (ili B_0) je dobro definisano.

Umesto linijskog pretraživanja zasnovanog na Volfovim uslovima može se upotrebiti *backtracking* linijsko pretraživanje s $\bar{\alpha} = 1$, s tim da se u tom slučaju mora uvesti mera zaštite (*safeguard*) jer uslov (2.10) može biti narušen. Mera zaštite podrazumeva da se za $s_k^T y_k$ koje je negativno ili previše blizu nuli matrica aproksimacije Hesijana ostavi nepromenjena. Tako za dovoljno pozitivno $s_k^T y_k$ se H_{k+1} , odnosno B_{k+1} , ažurira po formuli (2.8), odnosno (2.9), a u suprotnom se stavi da je $H_{k+1} = H_k$, odnosno $B_{k+1} = B_k$.

2.2 Optimizacija problema s ograničenjima

Problem optimizacije s ograničenjima možemo formulisati kao

$$\min_{x \in \Omega} f(x) \quad (2.11)$$

gde je funkcija $f : \mathbb{R}^n \rightarrow \mathbb{R}$ funkcija cilja, kao i u problemu bez ograničenja, a skup Ω pravi podskup skupa \mathbb{R}^n . Za skup Ω kažemo da je dopustiv skup, a za tačke koje mu pripadaju da su dopustive tačke. Najopštija formulacija dopustivog skupa je preko ograničenja tipa jednakosti i ograničenja tipa nejednakosti:

$$\Omega = \{x \in \mathbb{R}^n : h_i(x) = 0, i \in \mathcal{E}, h_i(x) \geq 0, i \in \mathcal{I}\}, \quad (2.12)$$

gde su funkcije $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$, $i \in \mathcal{E} \cup \mathcal{I}$, funkcije ograničenja.

2.2.1 Uslovi optimalnosti

Rešenja problema s ograničenjima su definisana analogno kao kod problema bez ograničenja s tim da rešenje mora biti dopustiva tačka. Tako, rešiti problem (2.11) znači pronaći tačku $x^* \in \Omega$ u kojoj funkcija f ima najmanju vrednost na skupu Ω . Za problem (2.11) x^* je tačka

- ▷ globalnog minimuma ako je $x^* \in \Omega$ i $f(x^*) \leq f(x)$ za svako $x \in \Omega$;
- ▷ lokalnog minimuma ako je $x^* \in \Omega$ i $f(x^*) \leq f(x)$ za svako dopustivo x iz neke okoline tačke x^* ;
- ▷ strogog lokalnog minimuma ako je $x^* \in \Omega$ i $f(x^*) < f(x)$ za svako dopustivo x , različito od x^* , iz neke okoline tačke x^* .

Za problem s ograničenjima kažemo da je nedopustiv ako mu je dopustiv skup prazan i tada on nema rešenje. Problem s ograničenjima nema rešenje ni kada njegova funkcija cilja nije ograničena odozdo na dopustivom skupu.

Kod problema s ograničenjima, pored uslova na gradijent funkcije cilja i uslova da je optimalna tačka dopustiva, i gradijenti funkcija ograničenja su uključeni u uslove optimalnosti. Stoga je definisana Lagranžova funkcija problema (2.11) s dopustivim skupom (2.12):

$$\mathcal{L}(x, \lambda) := f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i h_i(x),$$

gde su $\lambda_i \in \mathbb{R}$ Lagranžovi množitelji.

Pre nego što formulišemo uslove optimalnosti, potrebno je da kažemo šta se podrazumeva pod skupom aktivnih ograničenja i LICQ uslovom. Ograničenje je aktivno u dopustivoj tački x ako je $h_i(x) = 0$, te je skup aktivnih ograničenja u dopustivoj tački x

$$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} : h_i(x) = 0\}.$$

LICQ (*Linear Independence Constraint Qualification*) uslov je zadovoljen u tački x ako su gradijenti funkcija ograničenja koja su aktivna u x linearno nezavisni, odnosno ako su funkcije iz skupa $\{\nabla h_i(x) : i \in \mathcal{A}(x)\}$ linearno nezavisne. Zadovoljen LICQ uslov u rešenju obezbeđuje da se, u okolini rešenja, linearnom aproksimacijom dopustivog skupa dobija dobra aproksimacija. Odnosno da je skup linearizovanih dopustivih pravaca

$$\begin{aligned} \mathcal{F}(x) := \{d \in \mathbb{R}^n : & d^T \nabla h_i(x^*) = 0 \quad \text{za svako } i \in \mathcal{E} \\ & d^T \nabla h_i(x^*) \geq 0 \quad \text{za svako } i \in \mathcal{I} \cap \mathcal{A}(x)\} \end{aligned}$$

u dopustivoj tački x istovetan sa skupom svih tangenti na dopustiv skup u x .

Postoje razni potrebni uslovi optimalnosti prvog reda. Navodimo KKT (Karush-Kuhn-Tucker) uslove definisane u sledećoj teoremi, a potom i uslov vezan za pravce iz skupa linearizovanih dopustivih pravaca.

Teorema 2.6. ([39]) *Neka je x^* tačka lokalnog minimuma problema (2.11) sa dopustivim skupom (2.12) i neprekidno diferencijabilnim funkcijama f i h_i , $i \in \mathcal{E} \cup \mathcal{I}$, i neka je zadovoljen LICQ uslov u x^* . Tada postoji jedinstven Lagranžov množitelj λ^* , s komponentama $\lambda_i^* \in \mathbb{R}$, $i \in \mathcal{E} \cup \mathcal{I}$, za koji važe KKT uslovi:*

- (i) $\nabla_x \mathcal{L}(x^*, \lambda^*) = 0$
- (ii) $h_i(x^*) = 0$ za svako $i \in \mathcal{E}$
- (iii) $h_i(x^*) \geq 0$ za svako $i \in \mathcal{I}$
- (iv) $\lambda_i^* \geq 0$ za svako $i \in \mathcal{I}$
- (v) $\lambda_i^* h_i(x^*) = 0$ za svako $i \in \mathcal{E} \cup \mathcal{I}$.

Primetimo da za probleme optimizacije s ograničenjima tipa jednakosti KKT uslov čine uslovi (i) i (ii).

Potreban uslov za lokalni minimum, u kojem važi LICQ uslov, je da nijedan pravac u skupu linearizovanih dopustivih pravaca ne bude opadajući.

Teorema 2.7. ([39]) *Neka je x^* tačka lokalnog minimuma problema (2.11) sa dopustivim skupom (2.12) i neprekidno diferencijabilnim funkcijama f i h_i , $i \in \mathcal{E} \cup \mathcal{I}$, i neka je zadovoljen LICQ uslov u x^* . Tada važi da je*

$$d^T \nabla f(x^*) \geq 0 \quad \text{za svako } d \in \mathcal{F}(x^*).$$

Ako za x^* postoji Lagranžov množitelj λ^* takav da za njega važe KKT uslovi, tada za pravce d iz skupa $\mathcal{F}(x^*)$ s osobinom da je

$$d^T \nabla h_i(x^*) = 0 \quad \text{za svako } i \in \mathcal{A}(x^*) \text{ za koje je } \lambda_i^* > 0 \quad (2.13)$$

važi da je

$$\begin{aligned} d^T \nabla h_i(x^*) &= 0 && \text{za svako } i \in \mathcal{E} \\ d^T \nabla h_i(x^*) &= 0 && \text{za svako } i \in \mathcal{I} \cap \mathcal{A}(x^*) \text{ sa } \lambda_i^* > 0 \\ d^T \nabla h_i(x^*) &\geq 0 && \text{za svako } i \in \mathcal{I} \cap \mathcal{A}(x^*) \text{ sa } \lambda_i^* = 0 \end{aligned}$$

što implicira da je $d^T \nabla f(x^*) = 0$, pa je pitanje da li se duž pravca d može još smanjiti vrednost funkcije cilja f . Odgovor na to pitanje daju uslovi optimalnosti drugog reda. Potreban i dovoljan uslov drugog reda su redom formulisani u naredne dve teoreme. Sa $\mathcal{C}(x^*, \lambda^*)$ je označen skup pravaca $d \in \mathcal{F}(x^*)$ za koje važi osobina (2.13).

Teorema 2.8. ([39]) *Neka je x^* tačka lokalnog minimuma problema (2.11) sa dopustivim skupom (2.12) i dva puta neprekidno diferencijabilnim funkcijama f i h_i , $i \in \mathcal{E} \cup \mathcal{I}$, neka je zadovoljen LICQ uslov i neka je λ^* Lagranžov množitelj za koji su zadovoljeni KKT uslovi. Tada je*

$$d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) d \geq 0 \quad \text{za svako } d \in \mathcal{C}(x^*, \lambda^*). \quad (2.14)$$

Dakle, negativna krivina Hesijana Lagranžove funkcije duž nekog pravca iz $\mathcal{C}(x^*, \lambda^*)$ daje informaciju da x^* nije tačka lokalnog minimuma.

Teorema 2.9. ([39]) *Ako za dopustivu tačku $x^* \in \mathbb{R}^n$ postoji Lagranžov množitelj λ^* za koji su zadovoljeni KKT uslovi i ako je*

$$d^T \nabla_{xx}^2 \mathcal{L}(x^*, \lambda^*) d > 0 \quad \text{za svako } d \in \mathcal{C}(x^*, \lambda^*) \setminus \{0\}, \quad (2.15)$$

tada problem (2.11) s dopustivim skupom (2.12) ima strogi lokalni minimum u tački x^ .*

Znači, pozitivna krivina Hesijana Lagranžove funkcije duž svih pravca iz $\mathcal{C}(x^*, \lambda^*) \setminus \{0\}$ pruža informaciju da je dostignut lokalni minimum.

Svaki lokalni minimum je i globalni minimum kad je u pitanju problem konveksnog programiranja – problem sa konveksnom funkcijom cilja i konveksnim dopustivim skupom, što za problem (2.11) sa (2.12) znači da su funkcije ograničenja tipa jednakosti linearne, a funkcije ograničenja tipa nejednakosti su konkavne.

Teorema 2.10. ([9]) *Lokalni minimum konveksne funkcije cilja na konveksnom dopustivom skupu je i globalni minimum. Ako je funkcija cilja još i strogo konveksna, tada je tačka globalnog minimuma jedinstvena.*

2.2.2 Metodi rešavanja

Iterativni postupci za rešavanje problema s ograničenjima se mogu grupisati u postupke koji vrše pretragu samo kroz dopustiv skup – dopustivi metodi i postupke koji dozvoljavaju da iterativne tačke budu nedopustive. Za primere

dopustivih metoda navodimo postupke zasnovane na projekciji na dopustiv skup. U tim postupcima se određuje kandidat za narednu iterativnu tačku, te se on projektuje na dopustiv skup. S obzirom na to da je projekcija jedinstveno određena ako se preslikavanje vrši na zatvoren i konveksan skup, da bi ti postupci bili dobro definisani, potrebno je da je dopustiv skup zatvoren i konveksan. Šta više, postupci koji koriste projekcije su praktični jedino kada je projektovanje na dopustiv skup relativno jednostavno, kao što je slučaj s ograničenjima koja predstavljaju samo donje i gornje granice na promenljivu, odnosno kada je dopustiv skup $\{x \in \mathbb{R}^n : l \leq x \leq u\}$ za neke $l, u \in \mathbb{R}^n$.

Najjednostavniji dopustiv metod zasnovan na projekciji na dopustiv skup je metod projekcije gradijenta u kojem se pretražuje po pravcu negativnog gradijenta. Pretpostavlja se da je funkcija cilja f neprekidno diferencijabilna, a da je dopustiv skup Ω neprazan, konveksan i zatvoren. Za novu iteraciju se uzima

$$x_{k+1} = x_k + \alpha_k d_k, \quad d_k = \Pi_{\Omega}(x_k - \gamma_k \nabla f(x_k)) - x_k,$$

gde je $\Pi_{\Omega}(\cdot)$ projekcija na dopustiv skup Ω definisana sa

$$\Pi_{\Omega}(x) := \arg \min_{y \in \Omega} \|x - y\|,$$

a dužina koraka $\alpha_k \in (0, 1]$ i γ_k je pozitivan parametar. U Bertsekas [9] je pokazano da ako se za γ_k uzme proizvoljan pozitivan parametar, na primer $\gamma_k = \gamma, k = 1, 2, \dots$, a dužina koraka α_k se određuje *backtracking* linijskim pretraživanjem, tada je tačka nagomilavanja niza $\{x_k\}$ stacionarna tačka. Drugi načini određivanja α_k i γ_k koji takođe obezbeđuju konvergenciju ka stacionarnoj tački, kao i više detalja o dopustivim metodima, mogu se naći u Bertsekas [9].

Kazneni postupci

Prilikom rešavanja problema s ograničenjima u kojima se dozvoljava nedopustivost iterativne tačke, zbog narušenosti uslova dopustivosti, javlja se pitanje prednosti između smanjivanja nedopustivosti i umanjivanja vrednosti funkcije cilja po iteracijama. Jedan vid balansiranja se može videti u filter metodama, a drugi u postupcima koji koriste *merit* funkciju. U oba pristupa se na neki način, na primer nekom funkcijom θ , meri nedopustivost.

Suština filter metoda je da se razdvojeno tretira problem minimizacije funkcije cilja f i problem minimizacije mere nedopustivosti θ . Grubo rečeno, dozvoljava se da se u narednoj iteraciji poveća nedopustivost ili vrednost funkcije cilja u odnosu na vrednosti u svim prethodnim iteracijama, s tim da se ne

dozvoljava istovremeno povećavanje obe vrednosti. Tako se za iterativnu tačku x_{k+1} bira tačka za koju ne postoji iterativna tačka x_j sa $j \in \{0, 1, \dots, k\}$ za koju važi da je $f(x_j) \leq f(x_{k+1})$ i $\theta(x_j) \leq \theta(x_{k+1})$. Ideja o filter metodama potiče od Fletcher i Leyffer [17], pa je dalje istraživana u [57–59].

U postupcima sa *merit* funkcijom se funkcija cilja f i mera nedopustivosti θ povezuju u *merit* funkciju, te se ona minimizira. Kazneni postupci spadaju u ove metode. Kod njih je *merit* funkcija kaznena funkcija koja je zbir funkcije cilja i mere nedopustivosti pomnožene kaznenim parametrom. Dakle,

$$\phi(x; \mu) := f(x) + \mu \theta(x),$$

pri čemu je μ pozitivan kazneni parametar kojim se kažnjava nedopustivost. Time se umesto problema s ograničenjem (2.11), rešava niz problema bez ograničenja. Problemi u tom nizu imaju rastuće vrednosti kaznenog parametra. Puštanjem da niz kaznenih parametara teži u beskonačno, postiže se dopustivost. Opšti oblik algoritama kaznenih postupaka se može formulisati na sledeći način.

Algoritam 2.2.

- K0** Izabrati: početni kazneni parametar $\mu_0 > 0$, početnu tačku x_0^s i niz tolerancija $\{\tau_k\}$ koji teži ka nuli.
- K1** Staviti da je $k = 0$.
- K2** Izračunati približno rešenje x_k problema $\min_{x \in \mathbb{R}^n} \phi(x; \mu_k)$ koristeći početnu tačku x_k^s i kriterijum zaustavljanja $\|\nabla_x \phi(x; \mu_k)\| \leq \tau_k$.
- K3** Ako x_k zadovoljava kriterijum za konvergenciju problema (2.11), zaustaviti algoritam.
- K4** Odabrati kazneni parametar $\mu_{k+1} > \mu_k$ i staviti da je $x_{k+1}^s = x_k$.
- K5** Postaviti da je $k = k + 1$ i preći na korak K2.

Postoje egzaktni kazneni postupci kod kojih se do rešenja problema s ograničenjima može doći rešavanjem jednog problema bez ograničenja. Razlog za to je što za njih važi da postoji kazneni parametar $\mu^* > 0$ takav da je rešenje problema s ograničenjima tačka lokalnog minimuma kaznene funkcije $\phi(x; \mu)$ za proizvoljan kazneni parametar $\mu > \mu^*$.

Mana kaznenih postupaka je da kaznena funkcija može da bude neograničena odozdo iako problem s ograničenjem ima jedinstveno rešenje. S obzirom

na to da je mera nedopustivosti nenegativna, $\theta(x) \geq 0$, tada se ta mana ne može javiti ako je funkcija cilja f ograničena odozdo.

Najjednostavniji kazneni postupak za rešavanje problema s ograničenjima tipa jednakosti (komponente funkcije h su h_i , $i \in \mathcal{E}$) je kvadratni kazneni postupak s kaznenom funkcijom

$$\phi_K(x; \mu) := f(x) + \frac{\mu}{2} \|h(x)\|^2.$$

Mera nedopustivosti ove funkcije je neprekidno diferencijabilna, te je za neprekidno diferencijabilne funkcije f i h i kaznena funkcija ϕ_K neprekidno diferencijabilna i zbog toga se problemi u koraku K2 algoritma 2.2 mogu rešavati tehnikama za glatke probleme bez ograničenja. Za niz $\{\tau_k\}$ koji teži u 0 i niz $\{\mu_k\}$ koji teži u beskonačno, uz neprekidno diferencijabilne funkcije f i h i LICQ uslov u rešenju, dobija se da je tačka nagomilavanja niza iteracija generisanog algoritmom 2.2 KKT tačka problema s ograničenjima tipa jednakosti.

Diferencijabilnost kaznene funkcije za kazneni postupak za rešavanje problema (2.11) ne važi jer je u tom slučaju kaznena funkcija

$$\phi_{KK}(x; \mu) := f(x) + \frac{\mu}{2} \|h(x)\|^2 + \frac{\mu}{2} \sum_{i \in \mathcal{I}} (\max\{0, -h_i(x)\})^2. \quad (2.16)$$

Primer egzaktnog kaznenog postupka za rešavanje problema (2.11) je postupak s kaznenom funkcijom

$$\phi_1(x; \mu) := f(x) + \mu \sum_{i \in \mathcal{E}} |h_i(x)| + \mu \sum_{i \in \mathcal{I}} \max\{0, -h_i(x)\}.$$

Ni ova kaznena funkcija nije diferencijabilna, ali ima veliku prednost u odnosu na funkciju (2.16) zbog egzaktnosti kaznenog postupka definisanog njime.

Razne verzije postupaka koje koriste *merit* funkciju mogu se naći u [16,21, 39].

Glava 3

Stohastička optimizacija

Značaj stohastičke optimizacije se može prepoznati i u naglom razvoju istraživanja na tu temu u toku poslednjih nekoliko decenija. Za početak bismo spomenuli knjige [10,15,27,53,54] koje sadrže mnoštvo primera, modela problema stohastičke optimizacije, kao i prikaz njihovih matematičkih osobina i postupaka za njihovo rešavanje. Uvid u zastupljenost stohastičkog programiranja daju [15,60] sa kolekcijama primena na realne probleme iz raznih oblasti, kao i programskim kodovima napisanim za određene modele.

U okviru ovog poglavlja navodimo šta je stohastička optimizacija, pa podrobnije razmatramo probleme stohastičkog programiranja u kojima su funkcije u obliku matematičkog očekivanja, kao i metode za njihovo rešavanje. Posebno ističemo SAA i VSS metode, s pregledom rezultata o njima, jer one čine osnovu istraživanja prikazanog u narednim poglavljima. Kako je predmet našeg istraživanja problem stohastičkog programiranja s ograničenjima, ovo poglavlje završavamo kratkim osvrtom na tehnike rešavanja problema s ograničenjima u stohastičkoj optimizaciji.

3.1 Problem stohastičkog programiranja

Pojam stohastička optimizacija pokriva stohastičke probleme optimizacije, ali i stohastičke postupke. Tako, na primer, govorimo o stohastičkoj optimizaciji i kada se za rešavanje determinističkog problema optimizacije upotrebi iterativni postupak u kojem se iterativne tačke ili pravci pretraživanja biraju na slučajan način (na osnovu definisane raspodele verovatnoće), kao što je slučaj u direktnim postupcima stohastičkog pretraživanja prikazanim u Spall [54].

U slučajevima kada se rešava problem determinističke optimizacije tako da se raspolaže samo vrednostima dobijenim na osnovu eksperimentalnih merenja

ili stohastičke simulacije, tada nedovoljna preciznost, odnosno nepotpunost, podataka sadrži u sebi neizvesnost. Zbog toga se, obično, do rešenja tih problema dolazi stohastičkim postupcima u kojima je ta neizvesnost modelirana pomoću šuma kao slučajne promenljive dodate na tačnu vrednost. Tako se u iterativnom postupku računa s vrednostima sa šumom (dostupne vrednosti), te se opet radi o stohastičkoj optimizaciji. Najpoznatiji postupak tog tipa je algoritam stohastičke aproksimacije. Govorićemo o njemu u narednom odeljku.

U mnogim slučajevima javlja se potreba za modeliranjem neizvesnosti koja potiče od nedostatka informacija zbog pojave koja zavisi od ishoda u budućnosti (na primer u slučaju formiranja ponude u skladu s potražnjom koja je nepoznata u tom momentu) ili zbog nepoznatih ili nepotpunih informacija (kao kod proizvodnje koja zavisi od klimatskih uslova). Tada se neizvesnost može modelirati uvođenjem slučajne promenljive kao parametra. Modeli koji u sebi sadrže slučajnost predstavljaju stohastičke probleme. Kada ti problemi čine, ili se mogu svesti na probleme optimizacije, tada dolazimo do stohastičkog problema optimizacije koji se može prikazati u obliku

$$\min_{x \in X} F(x, \xi), \quad (3.1)$$

za neku funkciju cilja $F : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$ i oblast definisanosti X (bilo determinističku, bilo stohastičku – kada i funkcije ograničenja zavise od slučajne promenljive), pri čemu je slučajna promenljiva $\xi : \Omega \rightarrow \mathbb{R}^p$ definisana nad prostorom verovatnoće (Ω, \mathcal{F}, P) .

Rešavanje stohastičkog problema (3.1) je obično veoma komplikovano jer se optimalna vrednost traži u odnosu na sve moguće realizacije slučajne promenljive (scenarije), a pri tome za različite realizacije ista vrednost promenljive jednom može biti dopustiva a drugi put nedopustiva, pa je pitanje da li se ona treba posmatrati kao dopustiva ili nedopustiva tačka, odnosno može li ona uopšte biti optimalna tačka. Zbog toga se vrši reformulacija problema (3.1). Jedan od načina da se to učini je da se slučajnost eliminiše tako što se umesto slučajne promenljive posmatra njena očekivana vrednost. Tako se, za oblast definisanosti X koja ne zavisi od slučajne promenljive,* dobija problem

$$\min_{x \in X} F(x, \bar{\xi}),$$

gde je $\bar{\xi} = E[\xi]$. Dobijeni problem je deterministički, te se može rešiti determinističkim postupcima. Međutim, rešenje dobijenog pojednostavljenog problema je uglavnom daleko od rešenja problema (3.1), tako da ovaj pristup nije preporučljiv, [10,15].

*Kada je oblast definisanosti X stohastička, tada se i u funkcijama koje je određuju slučajna promenljiva zamenjuje njenim očekivanjem.

Druga reformulacija stohastičkog problema (3.1) je da se umesto njega posmatra problem

$$E[\min_{x \in X} F(x, \xi)],$$

što je takozvani *wait-and-see* pristup. U ovom slučaju se rešava određen broj determinističkih problema dobijenih uzimanjem konkretnih realizacija slučajne promenljive ξ , te se određuje očekivana vrednost dobijenih rešenja. U većini slučajeva, ovaj pristup je i dalje komplikovan za praktičnu upotrebu, te ga nećemo dalje razmatrati.

Treći način reformulisanja problema je optimizacija očekivanih performansi stohastičkog problema (*here-and-now* pristup), odnosno da se posmatra matematičko očekivanje stohastičke funkcije umesto nje same. Tako se rešava problem

$$\min_{x \in X} E[F(x, \xi)] \quad (3.2)$$

kada oblast definisanosti ne zavisi od slučajne promenljive, tj. za $X \subset \mathbb{R}^n$. Kada je i oblast definisanosti stohastička, tada se uzima i očekivana vrednost stohastičkih funkcija koje određuju X . Dakle, za

$$X = \{x \in \mathbb{R}^n : F_i(x, \xi) = 0, i = 1, \dots, m, F_i(x, \xi) \leq 0, i = m + 1, \dots, s\}$$

posmatra se problem

$$\min_{x \in Y} E[F(x, \xi)], \quad (3.3)$$

pri čemu je

$$Y = \{x \in \mathbb{R}^n : E[F_i(x, \xi)] = 0, i = 1, \dots, m, E[F_i(x, \xi)] \leq 0, i = m + 1, \dots, s\}.$$

Ova formulacija, bez ograničenja ili sa ograničenjima, se može videti u mnogim radovima, na primer u [20,25,43,46,56].

Kada je slučajna promenljiva ξ potpuno poznata, odnosno poznata je njena raspodela, tada se s očekivanom vrednošću uklanja slučajnost, pa problemi (3.2) i (3.3) postaju deterministički i njihovom rešavanju se može pristupiti determinističkim postupcima. Međutim, retko se može odrediti analitički izraz matematičkog očekivanja, a u slučaju kada se poseduju samo delimične informacije o slučajnoj promenljivoj (na primer kada se optimizira na osnovu merenih podataka ili simulacije) ni ne može se odrediti. U situacijama kada se ne može odrediti matematičko očekivanje ili je to previše komplikovano, deterministički problem se može rešiti stohastičkim postupkom ili se može izvršiti aproksimacija matematičkog očekivanja i rešavati problem optimizacije (3.2)

ili (3.3) koristeći tu aproksimaciju. U ovom svetlu, problemi (3.2) i (3.3) se nazivaju problemima stohastičkog programiranja.

Kada je poznata funkcija gustine slučajne promenljive, tada se aproksimacija matematičkog očekivanja može izvršiti aproksimacijom podintegralne funkcije u očekivanju. Uglavnom se tada podintegralna funkcija aproksimira tako da se dobije višestruki integral koji se može zapisati kao zbir jednostrukih integrala jer se jednostruki integrali mogu izračunati i numeričkim postupcima, na primer aproksimirati kvadraturnim formulama. Pogodnost aproksimacije umnogome zavisi od samog problema, što se može videti u Birge i Louveaux [10], te nećemo dalje razmatrati ovaj pristup.

Matematičko očekivanje se može još aproksimirati uzoračkim očekivanjem. U tom slučaju se uzorak od N slučajnih promenljivih tretira kao skup konkretnih realizacija ξ_1, \dots, ξ_N slučajne promenljive koje predstavljaju scenarije koji se javljaju s verovatnoćom $1/N$. Time se integral aproksimira sumom, odnosno za gustinu raspodele φ_ξ slučajne promenljive ξ ,

$$E[F(x, \xi)] = \int_{\Omega} F(x, t) \varphi_\xi(t) dt$$

se aproksimira sa

$$\sum_{i=1}^N F(x, \xi_i) \frac{1}{N}.$$

Za dobru aproksimaciju je potreban veliki broj N , na šta ukazuje sledeća činjenica.

Ako slučajne promenljive u uzorku ξ_1, \dots, ξ_N imaju istu raspodelu kao i ξ , tada i slučajne promenljive $F(x, \xi_1), \dots, F(x, \xi_N)$, za neko fiksirano $x \in X$, imaju istu raspodelu kao i $F(x, \xi)$, što znači da imaju i istu očekivanu vrednost, odnosno da je

$$E[F(x, \xi)] = E[F(x, \xi_1)] = \dots = E[F(x, \xi_N)].$$

Stoga, na osnovu jakog zakona velikih brojeva, ako su ξ_1, \dots, ξ_N još i nezavisne i očekivanje $E[F(x, \xi)]$ je konačno, tada

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N F(x, \xi_i) = E[F(x, \xi)] \quad \text{s.s.}$$

3.2 Metodi rešavanja problema stohastičkog programiranja

Rezultat *no free lunch theorems* iz Wolpert i Macready [64] ukazuje na to da svi algoritmi (kako deterministički, tako i stohastički) imaju podjednake performanse posmatrajući skup svih mogućih problema. To u stvari znači da se algoritam koji ima dobre performanse na jednoj grupi problema, sigurno loše ponaša na drugoj grupi problema. Pored toga, čak i kad je u pitanju grupa problema s određenim osobinama ne može se reći hoće li performanse nekog algoritma biti dobre. Međutim, ukoliko je algoritam zasnovan na određenim osobinama problema, tada se za njega može znati da ima dobre performanse na problemima tog tipa.

Kako ne postoji algoritam za koji možemo reći da ima najbolje performanse, u ovom delu razmatramo najzastupljenije stohastičke postupke za rešavanje problema (3.2) i pravimo kratak pregled rezultata vezanih za njih. U pitanju su postupak stohastičke aproksimacije – SA (*Stochastic Approximation*) metod, SAA (*Sample Average Approximation*) metod i postupak s promenljivom veličinom uzorka – VSS (*Variable Sample Size*) metod.

Dakle, posmatramo problem

$$\min_{x \in X} f(x), \quad (3.4)$$

gde je $f(x) := E[F(x, \xi)]$, funkcija $F : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$, $X \subset \mathbb{R}^n$ i slučajna promenljiva $\xi : \Omega \rightarrow \mathbb{R}^p$ je definisana nad prostorom verovatnoće (Ω, \mathcal{F}, P) .

3.2.1 SA metod

Ideja SA metoda se pojavila u radu Robbins i Monro [45] kao postupak za rešavanje sistema jednačina za koji se poseduju jedino vrednosti sa šumom. U pitanju je uopštavanje postupka najbržeg pada u smislu da se niz aproksimacija rešenja generiše koristeći vrednosti sa šumom i unapred definisane dužine koraka. Konvergencija metoda, konvergencija u verovatnoći, je postignuta pogodnim izborom niza dužina koraka. Kasnija istraživanja su pokazala da se, pod odgovarajućim pretpostavkama, može postići i skoro sigurna konvergencija.

S obzirom na to da se rešavanje problema (3.4) za $X = \mathbb{R}^n$ svodi na određivanje stacionarne tačke, odnosno na rešavanje sistema jednačina $\nabla f(x) = 0$, SA metod se može primeniti za rešavanje tog problema. Formulacija postupka

za taj slučaj je sledeća. Za unapred definisan niz dužina koraka $\{a_k\}_{k \in \mathbb{N}}$ i realizacije slučajne promenljive $\xi: \xi_1, \xi_2, \dots$, generiše se niz iteracija

$$x_{k+1} = x_k - a_k \nabla_x F(x_k, \xi_k), \quad (3.5)$$

uz pretpostavku da je funkcija $F(\cdot, \xi)$ diferencijabilna za skoro svako $\xi \in \Omega$. Izraz $\nabla_x F(x, \xi)$ se naziva stohastičkim gradijentom, te se SA algoritam sa (3.5) naziva postupkom stohastičkog gradijenta. Jedan od uslova kojim se postiže skoro sigurna konvergencija postupka je formulisan u sledećoj teoremi.

Teorema 3.1. ([54]) *Neka je x^* jedinstveno rešenje problema (3.4) sa $X = \mathbb{R}^n$ i neka važe uslovi:*

$$(i) \ a_k > 0, k \in \mathbb{N}, \quad \lim_{k \rightarrow \infty} a_k = 0, \quad \sum_k a_k = \infty \quad i \quad \sum_k a_k^2 < \infty$$

(ii) *postoji simetrična i pozitivno definitna matrica B sa osobinom da za svako $\eta \in (0, 1)$ važi da je*

$$\inf_{\eta < \|x - x^*\| < 1/\eta} (x - x^*)^T B \nabla f(x) > 0$$

(iii) $E[\nabla_x F(x, \xi_k) - \nabla f(x)] = 0$ za svako x i k

(iv) *postoji $c > 0$ takvo da za svako x i k važi da*

$$\|\nabla f(x)\|^2 + E[\|\nabla_x F(x, \xi_k) - \nabla f(x)\|^2] \leq c(1 + \|x\|^2).$$

Tada za niz $\{x_k\}_{k \in \mathbb{N}}$ generisan sa (3.5) važi da je

$$\lim_{k \rightarrow \infty} x_k = x^* \quad s.s.$$

Primer niza $\{a_k\}$ koji zadovoljava uslov (i) iz prethodne teoreme je

$$a_k = \frac{a}{(k+1)^\alpha}, \quad a > 0, \alpha \in (0.5, 1]. \quad (3.6)$$

Niz dužina koraka ima veliki uticaj na stabilnost i performanse postupka, te su rađena mnoga istraživanja na tu temu. Pregled navedenih istraživanja se može videti u Spall [54]. Pokazalo se da je stepen konvergencije iteracija x_k ,

dobijenih sa (3.6), ka rešenju x^* stohastičkog reda $O_v(1/\sqrt{k^\alpha})$.[†] To znači da se može postići konvergencija reda $O_v(1/\sqrt{k})$.

Uslov (iii) iz teoreme 3.1 predstavlja nepristrasnost ocene $\nabla_x F(x, \xi_k)$ i zadovoljen je pod standardnim pretpostavkama za stohastičku optimizaciju, na primer pod pretpostavkama naredne teoreme.

Teorema 3.2. ([53]) *Neka je za datu tačku \tilde{x} funkcija očekivanja $f(\tilde{x})$ dobro definisana i konačna, a funkcija $F(\cdot, \xi)$ diferencijabilna u \tilde{x} za skoro svako $\xi \in \Omega$. Neka još postoji pozitivna slučajna promenljiva $C(\xi)$ za koju važi da je $E[C(\xi)] < \infty$ i za svako y i z iz okoline \tilde{x} i skoro svako $\xi \in \Omega$ važi da*

$$|F(y, \xi) - F(z, \xi)| \leq C(\xi) \|y - z\|.$$

Tada je i $f(\cdot)$ diferencijabilna u \tilde{x} i važi da je

$$\nabla f(\tilde{x}) = \nabla E[F(\tilde{x}, \xi)] = E[\nabla_x F(\tilde{x}, \xi)].$$

Ako važi uslov (iii), tada uslov (iv) predstavlja ograničenost drugog momenta ocene kvadratnom funkcijom od x :

$$E[\|\nabla_x F(x, \xi_k)\|^2] \leq c(1 + \|x\|^2),$$

jer je

$$\begin{aligned} E[\|\nabla_x F(x, \xi_k) - \nabla f(x)\|^2] &= \\ &= E[\|\nabla_x F(x, \xi_k)\|^2 - 2\nabla^T f(x) \nabla_x F(x, \xi_k) + \|\nabla f(x)\|^2] \\ &= E[\|\nabla_x F(x, \xi_k)\|^2] - 2\nabla^T f(x) E[\nabla_x F(x, \xi_k)] + \|\nabla f(x)\|^2 \\ &= E[\|\nabla_x F(x, \xi_k)\|^2] - \|\nabla f(x)\|^2. \end{aligned}$$

Primena SA metode na problem optimizacije sa ograničenjima može se videti u radovima [19,38,50]. U tom slučaju je potrebno primeniti projekciju na dopustiv skup. Neka je funkcija $\Pi_X : \mathbb{R}^n \rightarrow X$ projekcija na skup X , pa je

$$\Pi_X(x) := \arg \min_{y \in X} \|x - y\|.$$

Po SA postupku za rešavanje problema (3.4) definiše se niz dužina koraka $\{a_k\}_{k \in \mathbb{N}}$, te se generiše niz iteracija

$$x_{k+1} = \Pi_X(x_k - a_k \nabla_x F(x_k, \xi_k)), \quad (3.7)$$

[†]Ako je $\{X_k\}_{k \in \mathbb{N}}$ niz slučajnih promenljivih, a $\{c_k\}_{k \in \mathbb{N}}$ niz konstanti, tada $X_k = O_v(c_k)$ označava da je X_k/c_k stohastički ograničeno, odnosno da za svako $\varepsilon > 0$ postoji $C > 0$ takvo da je $P\{|X_k/c_k| > C\} \leq \varepsilon$ za svako $k \in \mathbb{N}$.

pri čemu su ξ_1, ξ_2, \dots realizacije slučajne promenljive ξ .

Za iteracije x_k generisane SA algoritmom (3.7) primenjenim sa $a_k := a/k$, $k \in \mathbb{N}$, na problem (3.4) može se dobiti da je očekivana greška iteracije x_k reda $O(1/\sqrt{k})$, a očekivana greška vrednosti $f(x_k)$ reda $O(1/k)$. Naime, po Nemirovski i dr. [38], sledeće pretpostavke:

- (i) X je neprazan, kompaktan i konveksan podskup od \mathbb{R}^n
- (ii) funkcija očekivanja $f(x)$ je dobro definisana i konačna na X
- (iii) funkcija $F(\cdot, \xi)$ je konveksna i diferencijabilna na X (što implicira da je i $f(x)$ konveksna i diferencijabilna na X i da operator očekivanja i gradijent mogu da zamene mesta)
- (iv) ξ_1, ξ_2, \dots su generisane tako da su nezavisne i imaju istu raspodelu kao slučajna promenljiva ξ
- (v) $f(x)$ je još i strogo konveksno na X , odnosno ako postoji $c > 0$ takvo da za svako $x, y \in X$ važi da

$$(x - y)^T (\nabla f(x) - \nabla f(y)) \geq c \|x - y\|^2$$

(što znači da je rešenje x^* problema (3.4) jedinstveno)

- (vi) postoji $M > 0$ takvo da za svako $x \in X$ važi da je

$$E[\|\nabla_x F(x, \xi)\|^2] \leq M^2,$$

obebeđuju da se sa $a_k := a/k$, $k \in \mathbb{N}$, gde je $a > 1/(2c)$, dobija da je

$$E[\|x_k - x^*\|^2] = O\left(\frac{1}{\sqrt{k}}\right) \quad \text{i} \quad E[f(x_k) - f(x^*)] = O\left(\frac{1}{k}\right).$$

Nemirovski i drugi, u radu [38], uporedili su SA metod sa SAA postupkom, na konveksnim problemima stohastičkog programiranja. Koristili su različite verzije nizova dužina koraka za SA metod. Teoretski i numerički rezultati su dali sličan kvalitet ocena, s tim da je poređenje vremena izračunavanja u numeričkim primerima prikazalo prednost SA postupka (konkretno postupka *robust mirror descent SA*).

Napomenuli bismo da je navedena prednost SA algoritma u odnosu na SAA metod dobijena na problemima sa strogo konveksnim funkcijama cilja i konveksnim dopustivim skupom. S obzirom na to da naše istraživanje obuhvata i nekonveksne probleme, a još za probleme s nekonveksnim dopustivim skupom iterativno pravilo (3.7) nije dobro definisano, za nas je praktičnije koristiti SAA metod.

3.2.2 SAA metod

SAA postupak je veoma rasprostranjen, [11,26,28,34,35,48,51,61]. Može se naći i pod imenima *sample-path optimization*, Plambeck i dr. [42], i *stochastic counterpart method*, Rubinstein i Shapiro [49].

Osnovna ideja SAA metoda je da se očekivana vrednost $f(x) = E[F(x, \xi)]$ aproksimira uzoračkom očekivanom vrednošću

$$f_N(x) := \frac{1}{N} \sum_{i=1}^N F(x, \xi_i),$$

gde su ξ_1, \dots, ξ_N realizacije slučajne promenljive ξ , te da se reši tako dobijeni problem. Dakle, umesto problema (3.4) rešava se SAA problem

$$\min_{x \in X} f_N(x). \quad (3.8)$$

To je deterministički problem jer je definisan za konkretan uzorak, te se može rešiti pogodnim determinističkim postupkom.

Detalji o SAA oceni i njenim statističkim osobinama kada su ξ_1, \dots, ξ_N nezavisne i imaju istu raspodelu se mogu naći u [52,53]. U tom slučaju $f_N(x)$ je nepristrasna i konzistentna ocena za $f(x)$. Naime, videli smo u prethodnom odeljku da uz uslov da je očekivanje $E[F(x, \xi)]$ konačno za fiksirano $x \in X$ važi da je

$$\lim_{N \rightarrow \infty} f_N(x) = f(x) \quad \text{s.s.},$$

što predstavlja konzistentnost za $x \in X$. Nepristrasnost važi jer je

$$E[f_N(x)] = \frac{1}{N} \sum_{i=1}^N E[F(x, \xi_i)] = \frac{1}{N} \sum_{i=1}^N E[F(x, \xi)] = f(x).$$

Pored toga se dobija da je varijansa SAA ocene

$$D[f_N(x)] = \frac{1}{N^2} \sum_{i=1}^N D[F(x, \xi_i)] = \frac{1}{N^2} \sum_{i=1}^N D[F(x, \xi)] = \frac{1}{N} D[F(x, \xi)].$$

Pod dodatnim uslovima važi i uniformna konvergencija SAA ocene ka funkciji f , što je formulisano u narednoj teoremi, ali pre toga navodimo pod kojim uslovima nad funkcijom dominira integrabilna funkcija. Ako postoji nenegativna funkcija $G(\xi)$ takva da je $E[G(\xi)] < \infty$ i za svako $x \in X$ važi da je $|F(x, \xi)| \leq G(\xi)$ s.s., tada nad funkcijom $F(x, \xi)$, $x \in X$, dominira integrabilna funkcija $G(\xi)$.

Teorema 3.3. ([53]) *Neka je X neprazan i kompaktan podskup od \mathbb{R}^n , neka je za svako $x \in X$ funkcija $F(\cdot, \xi)$ neprekidna u x za skoro svako $\xi \in \Omega$ i neka nad funkcijom $F(x, \xi)$, $x \in X$, dominira integrabilna funkcija. Ako su slučajne promenljive u uzorku ξ_1, \dots, ξ_N nezavisne i imaju istu raspodelu, tada je funkcija $f(x)$ konačna i neprekidna na X i*

$$\lim_{N \rightarrow \infty} f_N(x) = f(x) \quad \text{s.s. uniformno na } X,$$

odnosno

$$\lim_{N \rightarrow \infty} \sup_{x \in X} |f_N(x) - f(x)| = 0 \quad \text{s.s.}$$

Konzistentnost SAA ocene obezbeđuje da greška koja se pravi primenom ocene teži ka nuli kada veličina uzorka teži ka beskonačnosti. Za meru greške koja se pravi konkretnim uzorkom može se koristiti dužina intervala poverenja za $f(x)$, s fiksnim $x \in X$, na nivou $1 - \alpha$:

$$\left[f_N(x) - z_{\alpha/2} \frac{\hat{\sigma}_N(x)}{\sqrt{N}}, f_N(x) + z_{\alpha/2} \frac{\hat{\sigma}_N(x)}{\sqrt{N}} \right] \quad (3.9)$$

gde je $z_{\alpha/2}$ kvantil reda $\alpha/2$ slučajne promenljive sa standardnom normalnom raspodelom,[‡] a $\hat{\sigma}_N^2(x)$ uzoračka varijansa definisana sa

$$\hat{\sigma}_N^2(x) := \frac{1}{N-1} \sum_{i=1}^N (F(x, \xi_i) - f_N(x))^2.$$

Opravdanost navedene mere leži u tome da (3.9) znači da je

$$P \left\{ \left| \frac{f_N(x) - f(x)}{\frac{\hat{\sigma}_N(x)}{\sqrt{N}}} \right| \leq z_{\alpha/2} \right\} = 1 - \alpha$$

i da na osnovu centralne granične teoreme, uz dodatni uslov da je varijansa $\sigma^2(x) := D[F(x, \xi)]$ konačna, važi da niz standardizovanih slučajnih promenljivih

$$\frac{f_N(x) - E[f_N(x)]}{\sqrt{D[f_N(x)]}}, \quad \text{odnosno} \quad \frac{f_N(x) - f(x)}{\frac{\sigma(x)}{\sqrt{N}}},$$

konvergira u raspodeli ka slučajnoj promenljivoj sa standardnom normalnom raspodelom. To znači da $f_N(x)$ ima približno normalnu raspodelu s očekivanjem $f(x)$ i varijansom $\sigma^2(x)/N$, za dovoljno veliko N , kao i da je greška ocene $f_N(x)$ stohastičkog reda $O_p(1/\sqrt{N})$.

[‡]Obično se uzima 90% ili 95% interval poverenja, te je $z_{0.05} \approx 1.64$ ili $z_{0.025} \approx 1.96$.

Uslovi pod kojima su i SAA ocene rešenja problema (3.4) konzistentne su formulisani u narednoj teoremi. Označimo sa f^* i f_N^* optimalne vrednosti originalnog problema (3.4) i SAA problema (3.8), respektivno, a sa S^* i S_N^* skupove svih rešenja.

Teorema 3.4. ([53]) *Neka su skupovi S^* i S_N^* neprazni i neka postoji neprazan i kompaktan skup $C \subset \mathbb{R}^n$ za koji važi:*

- (i) $S^* \subseteq C$ i $S_N^* \subseteq C$ s.s. za dovoljno veliko N
- (ii) funkcija $f(x)$ je konačna i neprekidna na C
- (iii) $\lim_{N \rightarrow \infty} f_N(x) = f(x)$ s.s. uniformno na C .

Tada

$$\lim_{N \rightarrow \infty} f_N^* = f^* \quad i \quad \lim_{N \rightarrow \infty} \sup_{x \in S_N^*} d(x, S^*) = 0^{\S} \quad s.s.$$

Ako je dopustiv skup X stohastički, uopšteno

$$X = \{x \in \mathbb{R}^n : E[F_i(x, \xi)] = 0, i = 1, \dots, m, \\ E[F_i(x, \xi)] \leq 0, i = m + 1, \dots, s\}, \quad (3.10)$$

gde $F_i : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}$, $i = 1, \dots, s$, tada problemu (3.4) sa (3.10) odgovara SAA problem

$$\min_{x \in X_N} f_N(x), \quad (3.11)$$

gde je

$$X_N = \{x \in \mathbb{R}^n : \frac{1}{N} \sum_{i=1}^N F_i(x, \xi) = 0, i = 1, \dots, m, \\ \frac{1}{N} \sum_{i=1}^N F_i(x, \xi) \leq 0, i = m + 1, \dots, s\}.$$

U tom slučaju su potrebne dodatne pretpostavke na dopustiv skup SAA problema da bi se dobila konzistentnost SAA ocena rešenja.

[§]Oznaka $d(x, A)$ predstavlja rastojanje tačke x od skupa A i definisano je sa

$$d(x, A) := \inf_{y \in A} \|x - y\|.$$

Teorema 3.5. ([53]) *Neka važe pretpostavke teoreme 3.4 i neka osim toga važi:*

(iv) *ako $x_N \in X_N$ i $\lim_{N \rightarrow \infty} x_N = x$ s.s., tada $x \in X$*

(v) *ako za neko $x \in S$ postoji niz $\{x_N\}$ takav da $x_N \in X_N$ i $\lim_{N \rightarrow \infty} x_N = x$ s.s.*

Tada za optimalne vrednosti originalnog problema (3.4) sa (3.10) i SAA problema (3.11), redom f^ i f_N^* , i skupove rešenja, redom S^* i S_N^* , važi da*

$$\lim_{N \rightarrow \infty} f_N^* = f^* \quad \text{i} \quad \lim_{N \rightarrow \infty} \sup_{x \in S_N^*} d(x, S^*) = 0 \quad \text{s.s.}$$

Kada je dopustiv skup deterministički, tada postojanje rešenja originalnog i SAA problema može da sledi iz nekih njegovih osobina. Takav primer je konačan dopustiv skup. Međutim, kada je u pitanju stohastički dopustiv skup, tada je značajno pretpostaviti postojanje rešenja SAA problema. Naime, SAA problem može biti nedopustiv za neki konkretan uzorak i u slučaju da dopustiv skup originalnog problema nije prazan. Na primer, za

$$X = \{x \in \mathbb{R} : E[x^2 + \xi] \leq 0\}$$

sa $E[\xi] = 0$ je $X = \{0\}$, a skup

$$X_N = \{x \in \mathbb{R} : \frac{1}{N} \sum_{i=1}^N (x^2 + \xi_i) \leq 0\}.$$

Za uzorke sa pozitivnim uzoračkim očekivanjem skup X_N je prazan. Stoga je pretpostavka postojanja rešenja uobičajena u stohastičkoj optimizaciji, što potvrđuju i radovi [7,34,49].

Pod uslovima koji obezbeđuju da očekivanje $f(x)$ i varijansa $\sigma^2(x)$, za nezavisne ξ_1, \dots, ξ_N sa istom raspodelom kao slučajna promenljiva ξ , imaju konačne vrednosti na kompaktnom dopustivom skupu, važi da je greška ocene optimalne vrednosti SAA problema najviše reda $O_v(1/\sqrt{N})$. U slučaju jedinstvenog rešenja originalnog problema greška ocene je baš reda $O_v(1/\sqrt{N})$.

Teorema 3.6. ([53]) *Neka važe pretpostavke:*

(i) *skup X je kompaktan*

(ii) *ξ_1, \dots, ξ_N su nezavisne i imaju istu raspodelu kao ξ*

(iii) *očekivanje $E[(F(\tilde{x}, \xi))^2]$ je konačno za neko $\tilde{x} \in X$*

(iv) postoji merljiva funkcija $C : \Omega \rightarrow \mathbb{R}_+$ takva da je $E[(C(\xi))^2]$ konačno i za svako $x, y \in X$ i skoro svako $\xi \in \Omega$ važi da

$$|F(x, \xi) - F(y, \xi)| \leq C(\xi) \|x - y\|.$$

Tada za optimalnu vrednost f_N^* SAA problema (3.8) važi da

$$\lim_{N \rightarrow \infty} \sqrt{N}(f_N^* - f^*) = \inf_{x \in S^*} Y(x) \quad \text{u raspodeli,}$$

pri čemu su $Y(x)$ slučajne promenljive s normalnom raspodelom $\mathcal{N}(0, \sigma^2(x))$. Ako je još $S^* = \{x^*\}$, tada $\sqrt{N}(f_N^* - f^*)$ konvergira u raspodeli ka slučajnoj promenljivoj s normalnom raspodelom $\mathcal{N}(0, \sigma^2(x^*))$.

Videli smo da se prilikom ocenjivanja $f(x)$ sa $f_N(x)$ kvalitet te ocene, odnosno greška $f_N(x) - f(x)$, može oceniti dužinom intervala poverenja (3.9). Kada je u pitanju kvalitet ocene rešenja x^* originalnog problema sa rešenjem x_N^* SAA problema, s determinističkim dopustivim skupom, može se posmatrati ocena greške $f(x_N^*) - f^*$. Greška (*optimality gap*)

$$g(\tilde{x}) := f(\tilde{x}) - f^*,$$

gde je \tilde{x} odabrani kandidat za rešenje, može se oceniti na više načina, pogledati [28,35,36,52,53]. Kada se koriste Monte Karlo ocene, što znači da su realizacije u uzorku nezavisne i da imaju istu raspodelu, jedan od načina da se oceni $g(\tilde{x})$ je sledeći. Uzme se M nezavisnih uzoraka veličine N :

$$\xi_{1,1}, \xi_{2,1}, \dots, \xi_{N,1}, \quad \dots, \quad \xi_{1,M}, \xi_{2,M}, \dots, \xi_{N,M},$$

pri čemu su $\xi_{1,j}, \dots, \xi_{N,j}$ nezavisni i imaju istu raspodelu kao i ξ , i za svaki od njih se odredi rešenje SAA problema, odnosno optimalna vrednost $f_{N,j}^*$. Ako je $\tilde{x} \in X$, tada je $g(\tilde{x}) \geq 0$, pa ostaje da se oceni gornja granica. Greška $g(\tilde{x})$ se može oceniti sa

$$\overline{f_{N,M}}(\tilde{x}) - \overline{f_{N,M}^*} := \frac{1}{M} \sum_{j=1}^M f_{N,j}(\tilde{x}) - \frac{1}{M} \sum_{j=1}^M f_{N,j}^*, \quad (3.12)$$

gde je

$$f_{N,j}(\tilde{x}) = \frac{1}{N} \sum_{i=1}^N F(\tilde{x}, \xi_{i,j}).$$

Varijansa ocene (3.12) je

$$D \left[\frac{1}{M} \sum_{j=1}^M (f_{N,j}(\tilde{x}) - f_{N,j}^*) \right] = \frac{1}{M^2} \sum_{j=1}^M D[f_{N,j}(\tilde{x}) - f_{N,j}^*],$$

a $D[f_{N,j}(\tilde{x}) - f_{N,j}^*]$, $j = 1, \dots, M$, se može oceniti sa

$$\begin{aligned} \hat{\sigma}_M^2 &= \frac{1}{M-1} \sum_{j=1}^M \left((f_{N,j}(\tilde{x}) - f_{N,j}^*) - \frac{1}{M} \sum_{i=1}^M (f_{N,i}(\tilde{x}) - f_{N,i}^*) \right)^2 \\ &= \frac{1}{M-1} \sum_{j=1}^M \left((f_{N,j}(\tilde{x}) - f_{N,j}^*) - (\overline{f_{N,M}}(\tilde{x}) - \overline{f_{N,M}^*}) \right)^2, \end{aligned}$$

te se za gornju granicu intervala poverenja za $g(\tilde{x})$ na nivou $1 - \alpha$ može uzeti

$$\overline{f_{N,M}}(\tilde{x}) - \overline{f_{N,M}^*} + z_\alpha \frac{\hat{\sigma}_M}{\sqrt{M}}, \quad (3.13)$$

pri čemu je z_α kvantil reda α slučajne promenljive sa standardnom normalnom raspodelom.

Na osnovu konzistentnosti SAA ocena važi da je za dovoljno veliki uzorak rešenje SAA problema dobra aproksimacija rešenja originalnog problema, a na osnovu ocene (3.13) se vidi da se dobra aproksimacija dobija i za veliko M . Stoga, kada računaska složenost rešavanja SAA problema brzo raste s porastom veličine uzorka, praktičnije je smanjiti veličinu uzorka i rešiti neki broj SAA problema s različitim uzorcima te veličine (međusobno nezavisnim). Od izračunatih rešenja SAA problema: $x_{N,1}^*, \dots, x_{N,M}^*$ za rešenje se može uzeti $x_{N,j}^*$ za koje je $\overline{f_{N,M}}(x_{N,j}^*)$ najbolja ocena. Razni postupci za odabir najboljeg rešenja mogu se naći u Nelson i dr. [37]. Rešavanje problema stohastičkog programiranja rešavanjem niza SAA problema se može videti u radovima [28,35]. U Kleywegt i dr. [28] su rešenja SAA problema dobijena koristeći Monte Karlo ocene, dok su u Linderoth i dr. [35] rešenja SAA problema dobijena koristeći uzorke bez uslova nezavisnosti.

U Dai i dr. [11] je pokazano da se greška $f(x_N^*) - f^*$ eksponencijalno smanjuje porastom veličine uzorka. Tačnije, pod odgovarajućim uslovima, za fiksirano $\varepsilon > 0$ verovatnoće

$$\begin{aligned} P\{f(x_N^*) - f^* \leq \varepsilon\}, \\ P\{\|x_N^* - x^*\| \leq \varepsilon\} \end{aligned} \quad (3.14)$$

eksponencijalnom brzinom teže ka jedinici kada $N \rightarrow \infty$. Pokazano je još da eksponencijalna konvergencija verovatnoće (3.14) važi i u slučaju kada ne važi uslov nezavisnosti u uzorku.

Razmatranjem SAA metoda s uzorkom bez uslova nezavisnosti i iste raspodele su se bavili i autori radova [26,65]. Pokazano je da se, pod određenim uslovima, i u ovom slučaju može dobiti konvergencija rešenja SAA problema ka rešenju originalnog problema.

3.2.3 Metodi s promenljivom veličinom uzorka

U prethodno delu smo videli da je za dobru SAA aproksimaciju matematičkog očekivanja potreban veliki uzorak. To u mnogim problemima povlači za sobom veliki broj izračunavanja vrednosti funkcije. U problemima u kojima je evaluacija funkcije skupa, odnosno predstavlja dominantan trošak, direktno rešavanje SAA problema determinističkim postupcima je skupo, te se stoga koriste metodi s promenljivom veličinom uzorka, takozvani VSS metodi, u kojima se u procesu optimizacije umesto jednog velikog uzorka koriste i uzorci manje veličine. Globalna ideja VSS metoda je da se računa s malim uzorcima dok je iterativna tačka daleko od rešenja, a da se koriste veliki uzorci kada se iterativna tačka približi rešenju. Time što se dopušta računanje i s manjim uzorcima dolazi se do smanjenja troškova.

Variranje veličine uzorka se javlja u više vidova. Jedan vid je rešavanje niza SAA problema s različitim veličinama uzoraka. Tu spadaju postupak dijagonalizacije, Polak i Royset [43], i retrospektivna aproksimacija (*retrospective approximation*), Pasupathy [40]. Suština oba postupka je da se rešava niz SAA problema i da se svaki problem u nizu reši s određenim brojem iteracija, odnosno odgovarajućom tačnošću, i da se naredni SAA problem rešava s uzorkom čija je veličina veća u odnosu na prethodni. Za početnu tačku novog problema uzima se aproksimacija iz poslednje iteracije prethodnog problema. U postupku dijagonalizacije u radu Polak i Royset [43] broj iteracija rešavanja svakog SAA problema, kao i veličina uzorka s kojim se rešava, su određeni pomoćnim problemom optimizacije tako da obezbeđuju linearnu konvergenciju SAA postupaka. U postupku retrospektivne aproksimacije su unapred određeni niz veličina uzoraka i niz tolerancija za greške s kojima se rešavaju SAA problemi, uz uslov da niz veličina uzorka teži ka beskonačnosti, a da niz tolerancija teži ka nuli. U radu Pasupathy [40] su navedeni primeri tih nizova s kojima se postiže određen red konvergencije.

Drugi vid variranja veličine uzorka je kada se veličina menja u svakoj iteraciji. Među postupcima ovog tipa prvo su se pojavili postupci u kojima je

niz veličina uzoraka neopadajući, kao u [12,25]. Deng i Ferris su, u radu [12], prikazali algoritam za rešavanje problema stohastičkog programiranja bez ograničenja koji se zasniva na postupku oblasti poverenja i u svakoj iteraciji određuje veličinu uzorka na osnovu Bajesove formule za testiranje dovoljne redukcije modela. Pokazali su skoro sigurnu konvergenciju algoritma, pod određenim uslovima, izostavivši uslov da niz veličina uzoraka teži ka beskonačnosti. Zbog toga se skoro sigurna konvergencija algoritma može postići i s ograničenim nizom veličina uzoraka, što su i pokazali na primeru. U radu Homem-de-Mello [25] je pokazano da se do optimalne tačke problema stohastičkog programiranja s determinističkim ograničenjima skoro sigurno može doći VSS postupcima u kojima se u svakoj iteraciji koristi novi uzorak (s realizacijama koje imaju istu raspodelu, međusobno su nezavisne i nezavisne su i od prethodnih uzoraka), pod uslovom da veličina uzorka raste odgovarajućom brzinom. Prikazan je primer u kojem se veličina uzorka utvrđivala u svakoj iteraciji primenom t-testa radi određivanja postoji li statistički značajna razlika među vrednostima SAA aproksimacija funkcije cilja u tekućoj i narednoj iterativnoj tački. Da bi se obezbedila brzina rasta veličine uzorka kojom se po teorijskim rezultatima dobija konvergencija, u svakoj unapred definisanoj iteraciji se povećava veličina uzorka nezavisno od rezultata t-testa.

Kasnije su se pojavili postupci u kojima se u procesu optimizacije može i smanjiti veličina uzorka, [6,29–31]. U tim radovima se u svakoj iteraciji određuje veličina uzorka i to na osnovu tačnosti aproksimacije rešenja problema i preciznosti aproksimacije funkcije cilja. U radovima [6,29,30] je posmatran problem stohastičkog programiranja bez ograničenja, odnosno SAA problem s unapred definisanom veličinom uzorka, te su predloženi VSS postupci definisani tako da se koristi uzorak s unapred definisanom maksimalnom veličinom. Pokazana je konvergencija predloženih algoritama ka rešenju SAA problema s maksimalnim uzorkom, a na numeričkim primerima je ilustrovana prednost predstavljenih VSS postupaka u odnosu na rešavanje SAA problema s maksimalnim uzorkom. U Bastin i dr. [6] algoritam je zasnovan na postupku oblasti poverenja, dok su ideje u radovima Krejić i Krklec [29] i Krejić i Krklec Jerinkić [30] redom zasnovane na monotonom i nemonotonom linijskom pretraživanju. Krejić i Krklec Jerinkić su u radu [31] posmatrali problem stohastičkog programiranja s ograničenjima u kojem je dopustiv skup deterministički. Predstavili su VSS postupak, zasnovan na nemonotonom linijskom pretraživanju i metodu projekcije gradijenta, u kojem niz veličina uzoraka teži ka besonačnosti, te se dobija skoro sigurna konvergencija ka rešenju originalnog problema, pod odgovarajućim pretpostavkama.

Na osnovu navedenih postupaka možemo da vidimo da se i sa stanovišta

načina ažuriranja veličine uzorka razlikuju VSS postupci. Na primer u postupcima u radovima [40,43] je korišćeno unapred definisano pravilo za promenu veličine uzorka – heurističko ažuriranje. S druge strane, u radovima [6,12,29–31] se veličina uzorka utvrđuje dinamički, u svakoj iteraciji, na osnovu informacija koje pruža algoritam. Stoga se za taj vid ažuriranja veličine uzorka kaže da je adaptivan.

Numerički primeri potvrđuju uštedu u broju izračunavanja vrednosti funkcija ili u vremenu računanja VSS postupaka iz radova [6,7,12,29–31] u odnosu na SAA postupak. U radovima [29–31], s adaptivnim ažuriranjem veličine uzorka, je izvršeno i poređenje sa postupcima koji koriste heurističku šemu za ažuriranje. Rezultati ukazuju na to da se adaptivnim ažuriranjem može uštedeti u broju izračunavanja funkcija.

Konzistentnost SAA ocena je značajna za obezbeđivanje konvergencije postupka optimizacije u kojem se koristi SAA aproksimacija. Videli smo da, na osnovu jakog zakona velikih brojeva, za uzorak sa nezavisnim realizacijama koje imaju istu raspodelu i dobro definisanu i konačnu funkciju f skoro sigurno važi konzistentnost SAA ocene f_N . U radu Homem-de-Mello [25] je pokazano da se uslov da realizacije u uzorku imaju istu raspodelu može zameniti uslovom da niz veličina uzoraka $\{N_k\}$ raste uz osobinu da važi

$$\sum_{k=1}^{\infty} \alpha^{N_k} < \infty \quad \text{za svako } \alpha \in (0, 1)$$

ili

$$\sum_{k=1}^{\infty} \frac{1}{N_k} < \infty.$$

Kod VSS metoda s adaptivnim ažuriranjem značajnu ulogu ima ocena konzistentnosti, odnosno gornja granica greške $|f_N(x) - f(x)|$. Za kumulativne uzorke (što znači da svaki novi uzorak sadrži prethodni uzorak) koji imaju nezavisne realizacije sa istom raspodelom, u radu Homem-de-Mello [25], je izvedena granica

$$|f_N(x) - f(x)| \leq C \sqrt{\frac{\ln(\ln N)}{N}}, \quad (3.15)$$

za dovoljno veliko N , gde je $C = C(x)$ odgovarajući pozitivan parametar vezan za varijansu od $F(x, \xi)$. Kada se u svakoj iteraciji koristi novi uzorak veličine N_k , s realizacijama koje imaju istu raspodelu, koje su međusobno nezavisne i nezavisne su i od prethodnih uzoraka, tada se pod uslovom da veličina uzorka raste s odgovarajućom brzinom, tačnije za

$$N_k \geq ak^b \quad \text{sa } a > 0, b > 2,$$

dobija granica

$$|f_{N_k}(x) - f(x)| \leq C \sqrt{\frac{\ln N_k}{N_k}}$$

za dovoljno veliko k , pri čemu je $C = C(x)$ odgovarajući pozitivan parametar određen varijansom od $F(x, \xi)$. Međutim, Polak i Royset u [43] predlažu da se umesto ocene (3.15) koristi ocena (3.9), odnosno

$$|f_{N_k}(x) - f(x)| \leq z_{\alpha/2} \frac{\hat{\sigma}_{N_k}(x)}{\sqrt{N_k}}.$$

3.3 Problemi s ograničenjima u stohastičkom programiranju

Pošto u narednim poglavljima razmatramo rešavanje problema stohastičkog programiranja s ograničenjima koristeći ideje iz SAA, odnosno VSS metoda, ovaj odeljak smo posvetili osvrtu na pregled tehnika rešavanja problema s ograničenjima u stohastičkoj optimizaciji prvenstveno kada se primenjuju ti postupci.

Na primer u Homem-de-Mello [25] je dopustiv skup deterministički i koristi se slučajna pretraga po dopustivom skupu. U radovima [19,20,31,38,50] je dopustiv skup takođe deterministički, ali se u njima koristi projekcija da bi iterativne tačke bile u dopustivom skupu.

Stohastički dopustiv skup, preciznije problem stohastičkog programiranja s ograničenjima tipa nejednakosti kod kojih su funkcije u obliku matematičkog očekivanja, može se naći u radovima [33,34,47]. U radu Li i dr. [34] KKT uslovi posmatranog problema su preformulisani u polugladak sistem jednačina, te je formulisan algoritam za njegovo rešavanje zasnovan na ideji SAA metoda i Njutnovog postupka. Navedeni su uslovi pod kojima konvergira postupak. Lan i Zhou u radu [33] su se ograničili na posmatranje konveksnih problema. Prikazali su dva algoritama formulisana na osnovu ideje SA metoda, s tim da na slučajan način biraju da li će za određivanje pravca pretraživanja koristiti gradijent funkcije cilja ili gradijent funkcije ograničenja. Navedeni su i uslovi pod kojima konvergiraju algoritmi. Royset u radu [47], s jedne strane, aproksimira originalni problem glačanjem i koristeći SAA aproksimacije, a s druge strane pravi linearizaciju funkcije cilja i ograničenja i uključuje je u kvadratni problem koji koristi kao funkciju optimalnosti. Originalni problem rešava nizom njenih aproksimacija, a preciznost tih aproksimacija meri ocenom funkcije optimalnosti. Kada dođe do dovoljno precizne aproksimacije problema,

poveća veličinu uzorka. Dao je uslove pod kojima se dobija konvergencija ka stacionarnoj tački originalnog problema.

Radovi [43,62] ukazuju na primenu kaznenih postupaka i u stohastičkoj optimizaciji. U oba rada je upotrebljen egzaktni kazneni postupak za rešavanje problema s funkcijom cilja u obliku matematičkog očekivanja i determinističkim ograničenjima. Stohastički problem koji su Polak i Royset, [43], razmatrali je problem s ograničenjima tipa nejednakosti. Za problem transformisan (uz pomoć egzaktno kaznene funkcije) u problem bez ograničenja, sa dovoljno velikim kaznenim parametrom, prikazali su algoritam za postupak dijagonalizacije, koji smo spominjali u prethodnom odeljku. Pokazana skoro sigurna konvergencija algoritma se odnosi na konvergenciju ka rešenju transformisanog problema. U radu Wang, Ma i Yuan [62] su posmatrani stohastički problemi s determinističkim ograničenjima tipa jednakosti i prikazani su algoritmi kojima se dobija ε -stohastički kritična tačka[¶] posmatranog problema. U navedenim algoritmima se u svakoj iteraciji rešavaju dva potproblema. Prvim potproblemom je definisano pravilo određivanja kaznenog parametra, a drugi je problem minimizacije neglatke i nekonveksne kaznene funkcije za koji je dat algoritam za rešavanje primenom stohastičke aproksimacije.

[¶]Grubo rečeno, tačka je ε -stohastički kritična ako važi da je očekivana vrednost mere KKT uslova u njoj manja od $\sqrt{\varepsilon}$.

Glava 4

Kazneni postupak s promenljivom veličinom uzorka za rešavanje SAA problema

Počevši od ovog poglavlja posmatramo probleme stohastičkog programiranja s ograničenjima tipa jednakosti u obliku matematičkog očekivanja (problem SP). U prethodnom poglavlju smo videli da se prilikom rešavanja problema stohastičkog programiranja često pribegava SAA postupcima, kao i da je za dobru aproksimaciju rešenja potreban veliki uzorak što, zbog iterativnog rešavanja SAA problema, dovodi do velikog broja izračunavanja vrednosti funkcije. Ovaj nedostatak postupka se može korigovati metodama s promenljivom veličinom uzorka (VSS metode) čija je ideja da se tokom iterativnog procesa varira veličina uzorka, odnosno da se radi s malim uzorcima kad god smo daleko od rešenja, a da se veliki uzorci koriste kad se približimo rešenju. Efikasnost VSS postupaka za rešavanje problema minimizacije s funkcijom cilja u obliku matematičkog očekivanja se može videti u radovima [12,29,30] kada su u pitanju problemi bez ograničenja, odnosno u radovima [7,25] za slučaj problema s determinističkim ograničenjima.

U ovom poglavlju posmatramo SAA reformulaciju problema SP. To znači da imamo SAA problem sa unapred zadatim uzorkom, te je on deterministički problem. Predstavićemo novu stohastičku metodu za rešavanje te SAA reformulacije. Postupak spada u VSS metode i kombinuje variranje veličine uzorka s kaznenim postupkom i linijskim pretraživanjem. Koristimo kumulativne uzorke i adaptivnu šemu za određivanje veličine uzorka u iteraciji. Algoritam ćemo predstaviti i detaljno opisati u odeljku 4.1. U odeljku 4.2 pokazaćemo da pod standardnim pretpostavkama novi metod konvergira ka KKT tački SAA

reformulacije. Numerički rezultati biće predstavljeni u poglavlju 6.

4.1 Algoritam

Posmatramo problem optimizacije sa ograničenjima tipa jednakosti

$$\min f(x) \quad \text{tako da je} \quad h(x) = 0 \quad (4.1)$$

gde funkcija cilja $f: \mathbb{R}^n \rightarrow \mathbb{R}$, a funkcija ograničenja je u obliku matematičkog očekivanja, odnosno

$$h(x) = E[H(x, \xi)]$$

s funkcijom $H: \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^m$, slučajnim vektorom $\xi: \Omega \rightarrow \mathbb{R}^p$ i prostorom verovatnoće (Ω, \mathcal{F}, P) .*

Neka je unapred zadat pun uzorak $\xi_1, \xi_2, \dots, \xi_{N_{max}}$. Aproksimiranjem matematičkog očekivanja uzoračkim očekivanjem dolazimo do SAA reformulacije originalnog problema (4.1)

$$\min f(x) \quad \text{tako da je} \quad h_{N_{max}}(x) = 0 \quad (4.2)$$

pri čemu je

$$h_N(x) = \frac{1}{N} \sum_{i=1}^N H(x, \xi_i)$$

SAA funkcija za proizvoljno $N \in \mathbb{N}$ takvo da je $N \leq N_{max}$.

Kao što smo videli u prethodnom poglavlju, do rešenja problema SP može se doći SAA metodom – rešavanjem niza SAA problema (4.2), svaki minimizirajući koristeći uzorak fiksirane veličine. Ako za taj niz problema važi da niz njihovih veličina uzoraka teži ka beskonačnosti, tada, pod standardnim pretpostavkama za teoriju stohastičkog programiranja, niz njihovih rešenja skoro sigurno konvergira ka rešenju originalnog problema (4.1). Dakle, ako se poboljša postupak rešavanja SAA problema (za nas u smislu smanjivanja broja izračunavanja funkcija potrebnog da se dođe do rešenja), tada se efikasnije može doći do rešenja originalnog problema. Zbog toga u ovom poglavlju nadalje posmatramo SAA problem (4.2).

Cilj nam je da rešimo problem (4.2) kombinovanjem efikasnog ažuriranja veličine uzorka N i ažuriranja kaznenog parametra μ kaznene funkcije

$$\phi_N(x; \mu) = f(x) + \mu \theta_N(x), \quad (4.3)$$

*Analiza bi se mogla proširiti i na slučaj kada je i funkcija cilja u obliku matematičkog očekivanja, ali je naš cilj da vidimo šta se dešava kada se VSS postupci primenjuju na ograničenja.

gde je

$$\theta_N(x) = \|h_N(x)\|^2$$

i $N \leq N_{\max}$. Tako u k -toj iteraciji imamo funkciju $\phi_{N_k}(x; \mu_k)$ s uzorkom veličine N_k i kaznenim parametrom μ_k . To znači da je u svakoj iteraciji potrebno odrediti: novu iterativnu tačku, novu veličinu uzorka i novi kazneni parametar. Kad je u pitanju uzorak veličine N_k , podrazumevamo da se on sastoji od prvih N_k elemenata punog uzorka $\xi_1, \xi_2, \dots, \xi_{N_{\max}}$.

Novu iterativnu tačku biramo s namerom da se približimo stacionarnoj tački kaznene funkcije sa veličinom uzorka N_k i kaznenim parametrom μ_k . Konkretnije, x_{k+1} određujemo pomoću linijskog pretraživanja, uzimajući opadajući pravac pretraživanja d_k i određivanjem dužine koraka α_k primenom *backtracking* procedure s Armizo uslovom postavljenim na kaznenu funkciju $\phi_{N_k}(x_k; \mu_k)$, odnosno uslovom

$$\phi_{N_k}(x_k + \alpha_k d_k; \mu_k) \leq \phi_{N_k}(x_k; \mu_k) + \eta \alpha_k g_k^T d_k, \quad (4.4)$$

gde je η parametar iz intervala $(0, 1)$.

Veličini uzorka N_k dopuštamo da varira iz iteracije u iteraciju. Za njeno određivanje koristimo adaptivnu šemu, odnosno izbor veličine vršimo na osnovu informacija o kvalitetu aproksimacija dobijenih u tekućoj iteraciji. Pratićemo informacije o tačnosti aproksimacije rešenja problema, kao i o preciznosti aproksimacije funkcije ograničenja. Tako, ažuriranje veličine uzorka zasniavamo na dve mere greške: mera optimalnosti dm_k i mera preciznosti $\varepsilon_{\delta}^{N_k}(x_k)$. Kako u svakoj iteraciji možemo da imamo različitu kaznenu funkciju, a ne funkciju $\phi_{N_{\max}}(x; \mu_k)$, mera optimalnosti predstavlja udaljenost od stacionarne tačke funkcije $\phi_{N_k}(x; \mu_k)$, dok je druga aproksimacija udaljenosti od punog uzorka, tj. predstavlja ocenu greške aproksimacije $h_{N_k} \approx h_{N_{\max}}$. Mera dm_k mora biti nenegativna, a za $\varepsilon_{\delta}^{N_k}(x_k)$ pretpostavljamo da je veća od neke pozitivne konstante.

Niz kaznenih parametara $\{\mu_k\}_{k \in \mathbb{N}}$ mora biti neopadajući da bi osigurao da ćemo doći do dopustive tačke. Stoga kazneni parametar povećavamo kada dođemo blizu stacionarne tačke funkcije $\phi_{N_k}(x; \mu_k)$. Preciznije, za novi kazneni parametar μ_{k+1} uzimamo broj veći od μ_k kada mera opadanja kaznene funkcije (u stvari dm_k) postane dovoljno mala, ali samo ako novi uzorak nije isti kao i prethodni u slučaju da nije u pitanju maksimalni uzorak. Poslednji uslov nam obezbeđuje da dostignemo pun uzorak, odnosno da se algoritam ne zaustavi na uzorku koji nije maksimalni.

Imamo još jedan uslov u algoritmu koji ima značajnu ulogu u dostizanju punog uzorka. U pitanju je donja granica veličine uzorka N_k^{\min} . Niz $\{N_k^{\min}\}_{k \in \mathbb{N}}$

se takođe adaptivno određuje, ali tako da bude neopadajući. Njegova uloga je da spreči permanentne promene veličine uzorka N_k .

Nadalje pretpostavimo da imamo na raspolaganju vrednost gradijenta

$$g_k := \nabla_x \phi_{N_k}(x_k; \mu_k).$$

Meru optimalnosti definišimo sa

$$dm_k = dm_k(\alpha_k) := -\alpha_k g_k^T d_k,$$

gde je α_k dužina koraka, a d_k je pravac pretraživanja. Kako pretpostavljamo da je d_k opadajući pravac za kaznenu funkciju $\phi_{N_k}(x_k; \mu_k)$, odnosno da važi da je $g_k^T d_k < 0$, nenegativnost mere optimalnosti je obezbeđena. S obzirom na to da se veličina $\alpha_k g_k^T d_k$ javlja u Armižo uslovu (4.4), ovako definisana mera optimalnosti se ne mora dodatno računati.

Mera optimalnosti bi se još mogla definisati, kao što je navedeno u radovima [13,30], sa

$$dm_k = \alpha_k^2 \beta_k$$

gde je α_k dužina koraka, a $\{\beta_k\}$ ograničen pozitivan niz za koji važi da teži nuli samo kada i gradijent kaznene funkcije teži nuli. Najjednostavniji takav niz se dobija za konstantno i pozitivno β_k , a moglo bi se koristiti i, na primer, $\beta_k = \|g_k\|$.

Što se tiče mere preciznosti, može se koristiti mera dužine intervala poverenja za $h_{N_k}(x_k)$, kao u [5,29–32,35,43]. Na primer

$$\varepsilon_{\delta}^{N_k}(x_k) := \hat{\sigma}_{N_k}(x_k) \frac{1.96}{\sqrt{N_k}}, \quad (4.5)$$

gde je $\hat{\sigma}_{N_k}^2(x_k)$ uzoračka varijansa definisana sa

$$\hat{\sigma}_{N_k}^2(x_k) := \frac{1}{N_k - 1} \sum_{i=1}^{N_k} \|H(x_k, \xi_i) - h_{N_k}(x_k)\|^2.$$

Broj 1.96 potiče od relevantnog kvantila 95% intervala poverenja standardne normalne raspodele.

Sad možemo da navedemo osnovni algoritam. Pošto posmatramo SAA problem (4.2), pretpostavljamo da je pun uzorak $\xi_1, \xi_2, \dots, \xi_{N_{max}}$ generisan na samom početku postupka optimizacije.

Algoritam 4.1.

K0 Izabrati: $N_{min} \in \mathbb{N}$, $x_0 \in \mathbb{R}^n$, $\beta, \eta, v_1 \in (0, 1)$, $\mu_0 > 0$, $\rho > 1$.

K1 Staviti da je $k = 0$, $N_k = N_{min}$, $N_0^{min} = N_{min}$, $x_k = x_0$, $\mu_k = \mu_0$, $t = 1$.

K2 Odabrati opadajući pravac d_k .

K3 Određivanje dužine koraka α_k :

Naći najmanji nenegativan prirodan broj j takav da za $\alpha_k = \beta^j$ važi Armižo uslov (4.4).

Uzeti da je $x_{k+1} = x_k + \alpha_k d_k$ i $dm_k = -\alpha_k g_k^T d_k$.

K4 Ako važi $dm_k \leq \frac{\alpha_k}{\mu_k^2}$, tada je $z_t = x_k$ i $t = t + 1$.

K5 Utvrditi veličinu uzorka N_{k+1} uz pomoć algoritma 4.2.

K6 Odrediti donju granicu veličine uzorka N_{k+1}^{min} koristeći algoritam 4.3.

K7 Određivanje kaznenog parametra μ_{k+1} :

Ako je

$$N_k = N_{k+1} < N_{max} \quad \text{ili} \quad dm_k > \frac{\alpha_k}{\mu_k^2},$$

tada je $\mu_{k+1} = \mu_k$, u suprotnom je $\mu_{k+1} = \rho \mu_k$.

K8 Postaviti da je $k = k + 1$ i preći na korak K2.

Primitimo da u algoritmu 4.1 koristimo uopšteni pravac pretraživanja (korak K2), s tim da za njega pretpostavljamo da je opadajući pravac za kaznenu funkciju $\phi_{N_k}(x_k; \mu_k)$.

Algoritmi 4.2 i 4.3 su zasnovani na algoritmima iz Krejić i Krklec Jerinkić [30]. Uzet je konkretan težinski parametar prilikom modifikacije veličine uzorka, a kod određivanja donje granice veličine uzorka posmatra se mera nedopustivosti $\theta_{N_{k+1}}$ umesto funkcije cilja. Pre nego što navedemo te algoritme, detaljno ćemo opisati njihovu ideju.

S obzirom na to da je glavna ideja dinamičke promene veličine uzorka da se uštedi na broju izračunavanja vrednosti funkcija tokom postupka optimizacije, dok je iteracija daleko od rešenja računa se s aproksimacijama manje preciznosti (upotrebom malih uzoraka), a kada iteracija dođe blizu rešenja povećava se preciznost aproksimacija (prelaskom na velike uzorke). Male vrednosti dm_k sugerišu da se nalazimo u blizini rešenja stacionarne tačke kaznene funkcije $\phi_{N_k}(x; \mu_k)$, odnosno trenutne aproksimacije posmatranog SAA

problema, te u tom slučaju povećavamo veličinu uzorka. S druge strane, velike vrednosti dm_k pokazuju da smo daleko od rešenja, pa tada smanjujemo veličinu uzorka. Naravno, smanjivanje i povećavanje veličine uzorka ostaje u granicama između donje granice veličine uzorka i maksimalne veličine uzorka. Šta su male, a šta velike vrednosti dm_k utvrđujemo na osnovu mere $\varepsilon_\delta^{N_k}(x_k)$. Dakle, uslov $dm_k < d\varepsilon_\delta^{N_k}(x_k)$ implicira povećanje uzorka, a $dm_k > d\varepsilon_\delta^{N_k}(x_k)$ njegovo smanjivanje, pri čemu je d težinski parametar. Kolika će biti promena u veličini uzorka zavisi od odnosa dm_k i $\varepsilon_\delta^{N_{k+1}}(x_k)$. Cilj je da nova veličina uzorka, N_{k+1} , zadovoljava uslov

$$dm_k \approx d\varepsilon_\delta^{N_{k+1}}(x_k) \quad \text{i} \quad N_k^{min} \leq N_{k+1} \leq N_{max}.$$

U literaturi su korišćene različite vrednosti za težinski parametar d . U radu Krejić i Krklec [29] je $d = 1$, dok je u Krejić i Krklec Jerinkić [30] $d \in (0, 1]$. Autori rada [6] su kombinovali $d = 1$ i nešto slično razmatranju odnosa tekuće veličine uzorka i nove veličine. U cilju izbegavanja nepotrebno velikih uzoraka, pogotovo što naše aproksimacije problema zavise i od kaznenog parametra, pa se njegovom promenom lako može promeniti status blizine rešenju, mi smo se odlučili za težinski parametar $d = N_k/N_{k+1}$. Time smo dobili manje promene u veličini uzorka kada je uzorak mali, odnosno kada imamo slabu preciznost i pitanje je koliko smo u stvari blizu rešenja. Kod velikih uzoraka, kada su nam informacije o blizini rešenja tačnije, dozvoljavamo veće promene u veličini uzorka. Tako naš podalgoritam za određivanje veličine uzorka ima sledeći oblik.

Algoritam 4.2.

K0 Ulazni parametri: $dm_k, \varepsilon_\delta^{N_k}(x_k), x_k, N_k, N_k^{min}, v_1 \in (0, 1)$.

K1 Određivanje veličine uzorka N_{k+1} :

- 1) Ako je $dm_k = \varepsilon_\delta^{N_k}(x_k)$, tada je $N_{k+1} = N_k$.
- 2) Ako je $dm_k > \varepsilon_\delta^{N_k}(x_k)$, tada uzeti da je $N = N_k$.
Dokle god je

$$dm_k > \frac{N_k}{N} \varepsilon_\delta^N(x_k) \quad \text{i} \quad N > N_k^{min},$$

stavljati da je $N = N - 1$.

Postaviti da je $N_{k+1} = N$.

- 3) Ako je $dm_k < \varepsilon_\delta^{N_k}(x_k)$, onda:

- i) Ako je $dm_k \geq v_1 \varepsilon_\delta^{N_k}(x_k)$, tada uzeti da je $N = N_k$.
Dokle god je

$$dm_k < \frac{N_k}{N} \varepsilon_\delta^N(x_k) \quad \text{i} \quad N < N_{max},$$

stavljati da je $N = N + 1$.

Postaviti da je $N_{k+1} = N$.

- ii) Ako je $dm_k < v_1 \varepsilon_\delta^{N_k}(x_k)$, tada je $N_{k+1} = N_{max}$.

Primetimo da po konstrukciji algoritma 4.2 uslov $N_k^{min} \leq N_k$, koji po algoritmu 4.1 važi za N_0 , obezbeđuje da je

$$N_k^{min} \leq N_{k+1} \leq N_{max}.$$

Ideja algoritma 4.3 je da se donja granica veličine uzorka povećava na osnovu informacija o meri opadanja mere nedopustivosti θ_N u zavisnosti od iteracija, posmatrajući isti nivo preciznosti (određen sa N). Tako se N_k^{min} povećava kad god se povećavanjem uzorka ($N_k < N_{k+1}$) dođe do uzorka veličine N_{k+1} i iteracije x_{k+1} sa kojima se ne postigne dovoljno opadanje funkcije $\theta_{N_{k+1}}$. Naravno, ovo poređenje je izvodljivo ukoliko smo već koristili uzorak određen sa N_{k+1} . Stoga označimo sa $l(k)$ iteraciju u kojoj smo poslednji put počeli da koristimo uzorak veličine N_{k+1} . Na primer:

k	0	1	2	3	4	5	6	7	8	9
N_k	3	3	10	10	7	8	10	10	6	10
$l(k)$	-	-	-	-	-	2	-	-	6	-

Smatramo da nije došlo do dovoljnog opadanja funkcije $\theta_{N_{k+1}}$ kad mera nedopustivosti između iteracija $x_{l(k)}$ i x_{k+1} u proseku opadne manje od dela mere preciznosti $\varepsilon_\delta^{N_{k+1}}(x_{k+1})$, odnosno ako važi da je

$$\frac{\theta_{N_{k+1}}(x_{l(k)}) - \theta_{N_{k+1}}(x_{k+1})}{k + 1 - l(k)} < \frac{N_{k+1}}{N_{max}} \varepsilon_\delta^{N_{k+1}}(x_{k+1}). \quad (4.6)$$

Navodimo algoritam za određivanje donje granice veličine uzorka.

Algoritam 4.3.

K0 Ulazni parametri: N_k, N_{k+1}, N_k^{min} .

K1 Određivanje donje granice veličine uzorka N_{k+1}^{min} :

- 1) Ako je $N_{k+1} \leq N_k$, tada je $N_{k+1}^{min} = N_k^{min}$.
- 2) Ako je $N_{k+1} > N_k$, onda:
 - i) Ako je N_{k+1} veličina uzorka koja još nije bila upotrebljena, tada je $N_{k+1}^{min} = N_k^{min}$.
 - ii) Ako je N_{k+1} veličina uzorka koja je već korišćena i ne važi nejednakost (4.6), tada je $N_{k+1}^{min} = N_k^{min}$.
 - iii) Ako je N_{k+1} veličina uzorka koja je već korišćena i važi nejednakost (4.6), tada je $N_{k+1}^{min} = N_{k+1}$.

Primetimo da po algoritmima 4.2 i 4.3 važi da $N_k^{min} \leq N_k$ implicira da je

$$N_k^{min} \leq N_{k+1}^{min} \leq N_{k+1}.$$

Dakle, pošto je $N_0^{min} = N_0$, za svako k važi da je $N_k^{min} \leq N_k$, te je

$$N_k^{min} \leq N_{k+1}^{min} \leq N_{k+1} \leq N_{max}.$$

4.2 Analiza konvergencije

Kao što smo videli u prethodnoj glavi, SAA problem (4.2) može biti nedopustiv za neki konkretan uzorak i u slučaju da dopustiv skup originalnog problema (4.1) nije prazan, tako da je neophodno pretpostaviti egzistenciju rešenja. Pored toga, radi postizanja konvergencije iteracija generisanih algoritmom 4.1 potrebne su nam i dodatne pretpostavke. U pitanju su pretpostavke koje su standardne u stohastičkoj optimizaciji, što se može videti u Shapiro i dr. [53].

Pretpostavka 4.1. *Funkcija cilja f je ograničena odozdo na dopustivom skupu SAA problema (4.2), funkcije $f(\cdot)$ i $H(\cdot, \xi_i)$ su neprekidno diferencijabilne na \mathbb{R}^n za svako $i = 1, 2, \dots, N_{max}$ i niz iteracija $\{x_k\}_{k \in \mathbb{N}_0}$ generisan algoritmom 4.1 ima bar jednu tačku nagomilavanja.*

Prethodna pretpostavka osigurava neprekidnost i diferencijabilnost funkcije h_N za svako $N \in \mathbb{N}$, $N \leq N_{max}$ i time se dobija da je i kaznena funkcija $\phi_N(x; \mu)$, definisana sa (4.3), neprekidno diferencijabilna. Nadalje, ograničenost funkcije f odozdo, zbog nenegativnosti mere nedopustivosti θ_N i kaznenog parametra μ , daje da je i kaznena funkcija $\phi_N(x; \mu)$ ograničena odozdo na istom skupu. Ovim se dobija da i problemi minimizacije kaznene funkcije imaju rešenje pod pretpostavkom da je dopustiv skup problema (4.2) neprazan.

Prvo ćemo pokazati da algoritam 4.1 obezbeđuje da mera optimalnosti teži nuli kada se kazneni parametar i veličina uzorka fiksiraju, koristeći istu tehniku kao u prvom delu dokaza leme 4.1 u Krejić i Krklec [29].

Lema 4.1. *Neka važi pretpostavka 4.1 i neka postoje $\bar{n}, \bar{N} \in \mathbb{N}$ i $\bar{\mu} \in \mathbb{R}$ takvi da za kazneni parametar μ_k i veličinu uzorka N_k generisane algoritmom 4.1 važi da je $\mu_k = \bar{\mu}$ i $N_k = \bar{N}$ za svako $k \geq \bar{n}$. Tada*

$$\lim_{k \rightarrow \infty} dm_k = 0.$$

Dokaz. Kako je kazneni parametar fiksiran,

$$\phi_{N_k}(x_{k+1}; \mu_{k+1}) = \phi_{N_k}(x_{k+1}; \mu_k) \quad \text{za svako } k \geq \bar{n},$$

pa Armižo uslov u koraku K3 algoritma 4.1 implicira da je

$$\phi_{N_k}(x_{k+1}; \mu_{k+1}) \leq \phi_{N_k}(x_k; \mu_k) - \eta dm_k,$$

odnosno

$$\phi_{\bar{N}}(x_{k+1}; \bar{\mu}) \leq \phi_{\bar{N}}(x_k; \bar{\mu}) - \eta dm_k \quad \text{za svako } k \geq \bar{n}.$$

Definišući

$$\phi(x) := \phi_{\bar{N}}(x; \bar{\mu}),$$

dobijamo da je

$$\phi(x_{k+1}) \leq \phi(x_k) - \eta dm_k \quad \text{za svako } k \geq \bar{n}. \quad (4.7)$$

Uzevši da je x^* proizvoljna tačka nagomilavanja niza $\{x_k\}_{k \in \mathbb{N}_0}$ (postoji na osnovu pretpostavke 4.1), imamo da je

$$x^* = \lim_{j \rightarrow \infty} x_{k_j}$$

za neki podniz $\{x_{k_j}\}_{j \in \mathbb{N}_0} \subseteq \{x_k\}_{k \in \mathbb{N}_0}$. Bez gubitka opštosti možemo da pretpostavimo da je $k_j \geq \bar{n}$ za svako j . Primenivši (4.7) za svako $k \in [k_{j-1}, k_j]$ dobijamo da

$$\begin{aligned} \phi(x_{k_j}) &\leq \phi(x_{k_{j-1}}) - \eta dm_{k_{j-1}} \leq \\ &\leq \phi(x_{k_{j-2}}) - \eta dm_{k_{j-2}} - \eta dm_{k_{j-1}} \leq \\ &\vdots \\ &\leq \phi(x_{k_{j-1}}) - \eta \sum_{i=k_{j-1}}^{k_j-1} dm_i. \end{aligned}$$

Odavde sledi da je

$$\phi(x_{k_j}) \leq \phi(x_{k_0}) - \eta \sum_{i=k_0}^{k_j-1} dm_i \quad \text{za svako } j. \quad (4.8)$$

Na osnovu neprekidnosti iz pretpostavke 4.1 i konvergentnosti niza $\{x_{k_j}\}$, postoji konstanta M za koju važi da je $\phi(x_{k_j}) \geq M$ za svako j . Time nejednakost (4.8) daje

$$M \leq \phi(x_{k_0}) - \eta \sum_{i=k_0}^{k_j-1} dm_i \quad \text{za svako } j$$

ili ekvivalentno

$$\sum_{i=k_0}^{k_j-1} dm_i \leq \frac{1}{\eta} (\phi(x_{k_0}) - M) \quad \text{za svako } j.$$

Puštajući da j teži beskonačnosti i uzevši u obzir da je $dm_i \geq 0$ za svako i , dobijamo da red $\sum_{i=k_0}^{\infty} dm_i$ konvergira, što znači da mu opšti član teži nuli, čime je lema dokazana. \square

Dalje dokazujemo da nakon konačnog broja iteracija veličina uzorka dostiže maksimalnu veličinu i dalje ostaje nepromenjena. Tok dokaza je sličan dokazu leme 4.1 u Krejić i Krklec [29], ali pre njega navodimo i sledeću pretpostavku.

Pretpostavka 4.2. Za niz iteracija $\{x_k\}_{k \in \mathbb{N}_0}$ generisanih algoritmom 4.1 postoje $\kappa > 0$ i $n_1 \in \mathbb{N}$ za koje važi da je $\varepsilon_{\delta}^{N_k}(x_k) \geq \kappa$ za svako $k \geq n_1$.

Lema 4.2. Pod pretpostavkama 4.1 i 4.2 postoji $q \in \mathbb{N}$ takvo da za veličine uzorka N_k generisane algoritmom 4.1 važi da je $N_k = N_{max}$ za svako $k \geq q$.

Dokaz. Niz $\{N_k\}_{k \in \mathbb{N}_0}$ generisan algoritmom 4.1 je beskonačan jer algoritam nema izlazni kriterijum. Zbog toga će nakon konačnog broja iteracija N_k postati fiksirano ili će se niz $\{N_k\}$ permanentno menjati. Prvo pokazujemo da se niz veličina uzorka ne može ustaliti na broju manjem od N_{max} .

Pretpostavimo da postoji $\tilde{n} > n_1$ takvo da za svako $k \geq \tilde{n}$ važi da je

$$N_k = \tilde{N} < N_{max}.$$

Tada na osnovu pravila ažuriranja kaznenog parametra (korak K7 algoritma 4.1) imamo da je $\mu_{k+1} = \mu_k$ za svako $k \geq \tilde{n}$. Dakle, veličina uzorka i kazneni

parametar su fiksirani za svako $k \geq \tilde{n}$, tj. zadovoljeni su uslovi leme 4.1, pa važi

$$\lim_{k \rightarrow \infty} dm_k = 0.$$

To znači da za dovoljno veliko k imamo da je

$$dm_k < v_1 \kappa,$$

gde je v_1 parametar dat u algoritmu 4.2, a κ parametar iz pretpostavke 4.2. Pošto je

$$\varepsilon_{\delta}^{N_k}(x_k) \geq \kappa,$$

važi da

$$dm_k < v_1 \varepsilon_{\delta}^{N_k}(x_k),$$

te korak K1 3) ii) algoritma 4.2 daje da se veličina uzorka povećava na N_{max} . Međutim, to je u kontradikciji s polaznom pretpostavkom da je $N_k = \tilde{N} < N_{max}$.

Sad pretpostavimo da se niz $\{N_k\}$ permanentno menja. Posmatrajmo niz donjih granica veličina uzoraka $\{N_k^{min}\}_{k \in \mathbb{N}_0}$. Algoritam 4.2 obezbeđuje da

$$N_k^{min} \leq N_{k+1} \leq N_{max},$$

a algoritam 4.3 da je

$$N_{k+1}^{min} = N_{k+1} \quad \text{ili} \quad N_{k+1}^{min} = N_k^{min},$$

čime se dobija da

$$N_k^{min} \leq N_{k+1}^{min},$$

odnosno da je niz donjih granica $\{N_k^{min}\}$ neopadajući. To znači da je $N_k^{min} < N_{max}$ za svako k jer bi u suprotnom bilo da je $N_k \geq N_k^{min} = N_{max}$ za svako k počevši od neke iteracije. Zbog toga se N_k^{min} povećava samo konačno mnogo puta, te postoji $\bar{n} > n_1$ za koje je

$$N_{k+1}^{min} = N_k^{min} \quad \text{za svako } k \geq \bar{n}.$$

Posmatrajmo sad najveću veličinu uzorka koja se beskonačno mnogo puta koristi i označimo je sa \bar{N} . Neka su k_1, k_2, \dots iteracije (veće ili jednake sa \bar{n}) u kojima je veličina uzorka povećana na \bar{N} . Takvih iteracija ima beskonačno mnogo jer bi u suprotnom veličina uzorka postala fiksirana na \bar{N} . Dakle, važi

$$N_{k_i} < N_{k_i+1} = \bar{N} \quad \text{i} \quad N_{k_i+1}^{min} = N_{k_i}^{min} \quad \text{za } i = 1, 2, \dots$$

To znači da je u iteraciji k_1 u algoritmu 4.3 došlo do realizacije koraka K1 2) pod i) ili ii). Međutim, u iteracijama k_2, k_3, \dots je došlo do realizacije koraka K1 2) ii), što znači da je u iteraciji k_i došlo do dovoljnog opadanja funkcije $\theta_{N_{k_i+1}}$ (tj. funkcije $\theta_{\bar{N}}$) od iteracije $l(k_i) = k_{i-1} + 1$, odnosno za $i = 2, 3, \dots$ važi nejednakost

$$\theta_{\bar{N}}(x_{k_{i-1}+1}) - \theta_{\bar{N}}(x_{k_i+1}) \geq \frac{\bar{N}}{N_{max}} (k_i - k_{i-1}) \varepsilon_{\delta}^{\bar{N}}(x_{k_i+1}).$$

Na osnovu pretpostavke 4.2 i činjenice da je $k_i - k_{i-1} \geq 1$ dobijamo da

$$\theta_{\bar{N}}(x_{k_{i-1}+1}) - \theta_{\bar{N}}(x_{k_i+1}) \geq \frac{\bar{N}}{N_{max}} \kappa > 0 \quad \text{za svako } i = 2, 3, \dots$$

Posledica prethodne nejednakosti je

$$\lim_{i \rightarrow \infty} \theta_{\bar{N}}(x_{k_i+1}) = -\infty.$$

Time smo došli do kontradikcije pošto je, po definiciji, funkcija $\theta_{\bar{N}}$ ograničena odozdo, tačnije $\theta_{\bar{N}}(x) \geq 0$.

Dakle, zbog obe kontradikcije zaključujemo da veličina uzorka ne može da se fiksira na broju manjem od N_{max} , niti je moguće da se niz veličina uzoraka permanentno menja, pa ostaje da se veličina uzorka fiksira na N_{max} . \square

Videli smo da se za dovoljno velik broj iteracija optimizacija nastavlja s punim uzorkom. Naredna teorema daje informaciju o tome šta se dešava sa kaznenim parametrom, ali nam je potrebna i sledeća pretpostavka.

Pretpostavka 4.3. *Pravac pretraživanja d_k je opadajući, ograničen i za proizvoljan podniz $K \subseteq \mathbb{N}$ zadovoljava implikaciju*

$$\lim_{k \in K} g_k^T d_k = 0 \implies \lim_{k \in K} g_k = 0.$$

Primetimo da pravac najbržeg pada zadovoljava implikaciju koju tražimo. Naime, za pravac negativnog gradijenta $d_k = -g_k$ važi:

$$0 = \lim_{k \in K} g_k^T d_k = \lim_{k \in K} (-\|g_k\|^2) \implies \lim_{k \in K} \|g_k\| = 0.$$

Teorema 4.1. *Ako važe pretpostavke 4.1–4.3, tada za kaznene parametre μ_k generisane algoritmom 4.1 važi da*

$$\lim_{k \rightarrow \infty} \mu_k = \infty.$$

Dokaz. Na osnovu leme 4.2 postoji $q \in \mathbb{N}$ takvo da je $N_k = N_{max}$ za svako $k \geq q$. Stoga, na osnovu koraka K7 algoritma 4.1, postoje dve mogućnosti za $k \geq q$:

1. $\mu_{k+1} = \mu_k$ ako je $dm_k > \alpha_k / \mu_k^2$.
2. $\mu_{k+1} = \rho \mu_k$ ako je $dm_k \leq \alpha_k / \mu_k^2$.

Kako je $\rho > 1$, po drugom slučaju se μ_k povećava. Ukoliko se beskonačno mnogo puta dogodi taj slučaj, tada važi tvrđenje teoreme. U cilju da dođemo do kontradikcije, pretpostavimo suprotno: da se drugi slučaj konačno mnogo puta dogodi, odnosno pretpostavimo da postoji iteracija $q_1 > q$ za koju važi da

$$dm_k > \frac{\alpha_k}{\mu_k^2} \quad \text{za svako } k > q_1.$$

Tada je $\mu_k = \mu_{q_1}$ za svako $k > q_1$, te je prethodna nejednakost ekvivalentna sa

$$-g_k^T d_k > \frac{1}{\mu_{q_1}^2}. \quad (4.9)$$

To znači da je niz $\{-g_k^T d_k\}_{k > q_1}$ ograničen odozdo.

S druge strane, pošto su veličina uzorka i kazneni parametar fiksirani za svako $k > q_1$, lema 4.1 implicira da je

$$0 = \lim_{k \rightarrow \infty} dm_k = \lim_{k \rightarrow \infty} (-\alpha_k g_k^T d_k).$$

Kao posledicu, uzimajući u obzir prethodno dobijenu ograničenost odozdo, važi

$$\lim_{k \rightarrow \infty} \alpha_k = 0. \quad (4.10)$$

Odavde sledi da postoji iteracija $q_2 > q_1$ takva da je dužina koraka α_k manja od 1 za svako $k > q_2$. Zbog toga za svako $k > q_2$ postoji α'_k za koje važi da je $\alpha_k = \beta \alpha'_k$ i α'_k ne zadovoljava Armižo uslov (4.4), tj. važi

$$\phi_{N_{max}}(x_k + \alpha'_k d_k; \mu_{q_1}) > \phi_{N_{max}}(x_k; \mu_{q_1}) + \eta \alpha'_k g_k^T d_k.$$

Radi kraćeg zapisa izostavljamo N_{max} i μ_{q_1} iz preostalog dela dokaza, te definišemo

$$\phi(x_k) := \phi_{N_{max}}(x_k; \mu_{q_1}).$$

Prethodna nejednakost je ekvivalentna sa

$$\frac{\phi(x_k + \alpha'_k d_k) - \phi(x_k)}{\alpha'_k} > \eta d_k^T \nabla \phi(x_k).$$

Na osnovu teoreme o srednjoj vrednosti postoji $t_k \in (0, 1)$ koje zadovoljava

$$\frac{\alpha'_k d_k^T \nabla \phi(x_k + t_k \alpha'_k d_k)}{\alpha'_k} > \eta d_k^T \nabla \phi(x_k),$$

odnosno

$$d_k^T \nabla \phi(x_k + t_k \alpha'_k d_k) > \eta d_k^T \nabla \phi(x_k). \quad (4.11)$$

Neka je x^* proizvoljna tačka nagomilavanja niza iteracija $\{x_k\}_{k \in \mathbb{N}}$ (postoji na osnovu pretpostavke 4.1), te za neko $K_1 \subseteq \{k \in \mathbb{N} : k > q_2\}$ važi

$$\lim_{k \in K_1} x_k = x^*.$$

Po pretpostavci 4.3, niz pravaca pretraživanja $\{d_k\}_{k \in \mathbb{N}}$ je ograničen, pa postoje d^* i podskup $K \subseteq K_1$ takvi da je

$$\lim_{k \in K} d_k = d^*.$$

Nadalje, na osnovu $\alpha_k = \beta \alpha'_k$ i (4.10) važi

$$\lim_{k \rightarrow \infty} \alpha'_k = 0.$$

Odavde, uz ograničenost niza $\{d_k\}$ i uslov da $t_k \in (0, 1)$ za svako $k > q_2$, sledi

$$\lim_{k \in K} (x_k + t_k \alpha'_k d_k) = x^*.$$

Pošto je, zbog pretpostavke 4.1, funkcija $\nabla \phi$ neprekidna, puštajući u nejednakosti (4.11) za $k \in K$ da $k \rightarrow \infty$, dobijamo

$$(d^*)^T \nabla \phi(x^*) \geq \eta (d^*)^T \nabla \phi(x^*)$$

ili ekvivalentno

$$(d^*)^T \nabla \phi(x^*) (1 - \eta) \geq 0.$$

Parametar $\eta \in (0, 1)$, pa prethodna nejednakost implicira da je

$$(d^*)^T \nabla \phi(x^*) \geq 0. \quad (4.12)$$

Pored toga, na osnovu pretpostavke 4.3, pravac pretraživanja je opadajući, te je $d_k^T \nabla \phi(x_k) < 0$ za svako $k \in K$. Puštanjem limesa po K , dobijamo

$$(d^*)^T \nabla \phi(x^*) \leq 0,$$

što sa (4.12) daje da je

$$(d^*)^T \nabla \phi(x^*) = 0.$$

Međutim, to je u kontradikciji s nejednakošću (4.9), pa smo došli do željenog zaključka. \square

Porast niza kaznenih parametara u beskonačno omogućava konvergenciju prikazanog algoritma, što pokazujemo u narednoj teoremi. Za početak, primetimo da smo u prethodnom dokazu dobili da je uslov $dm_k \leq \alpha_k / \mu_k^2$ zadovoljen beskonačno mnogo puta. To znači da se u algoritmu 4.1 korak K4 izvršava beskonačno mnogo puta i time se kreira beskonačan niz iteracija $\{z_t\}$. Pod posmatranim pretpostavkama, svaka tačka nagomilavanja tog niza je stacionarna tačka problema minimizacije mere nedopustivosti posmatranog SAA problema. Dodavanjem pretpostavke da su gradijenti funkcije ograničenja linearno nezavisni u toj tački, dobija se da je ta tačka nagomilavanja KKT tačka SAA problema.

Teorema 4.2. *Neka su zadovoljene pretpostavke 4.1–4.3. Tada je svaka tačka nagomilavanja x^* niza $\{z_t\}_{t \in \mathbb{N}}$ stacionarna tačka za funkciju $\theta_{N_{max}}$. Ako je još i LICQ uslov zadovoljen, tada je x^* KKT tačka SAA problema (4.2).*

Dokaz. Neka je x^* proizvoljna tačka nagomilavanja niza $\{z_t\}$. Pošto je $\{z_t\}$ podniz niza iteracija $\{x_k\}$, postoji skup indeksa $K \subseteq \mathbb{N}$ takav da je

$$\lim_{k \in K} x_k = x^*$$

i

$$dm_k \leq \frac{\alpha_k}{\mu_k^2} \quad \text{za svako } k \in K. \quad (4.13)$$

Pored toga, na osnovu leme 4.2 postoji prirodan broj q za koji je

$$N_k = N_{max} \quad \text{za svako } k \geq q,$$

a teorema 4.1 daje da

$$\lim_{k \rightarrow \infty} \mu_k = \infty.$$

Bez gubitka opštosti možemo da pretpostavimo da je $k \geq q$ za svako k iz skupa K .

Kako na osnovu nejednakost (4.13) za svako $k \in K$ važi

$$-g_k^T d_k \leq \frac{1}{\mu_k^2},$$

divergencija niza kaznenih parametara $\{\mu_k\}$ ka beskonačnosti i pozitivnost izraza $-g_k^T d_k$ (po pretpostavci da je d_k opadajući pravac) impliciraju

$$\lim_{k \in K} g_k^T d_k = 0.$$

Ovim iz pretpostavke 4.3 sledi

$$\lim_{k \in K} g_k = 0. \quad (4.14)$$

S obzirom na to da je

$$g_k = \nabla_x \phi_{N_{max}}(x_k; \mu_k) = \nabla f(x_k) + \mu_k \nabla \theta_{N_{max}}(x_k),$$

dobija se

$$\|\mu_k \nabla \theta_{N_{max}}(x_k)\| \leq \|g_k\| + \|\nabla f(x_k)\|,$$

odnosno

$$\|\nabla \theta_{N_{max}}(x_k)\| \leq \frac{1}{\mu_k} (\|g_k\| + \|\nabla f(x_k)\|).$$

Sada nam neprekidnost funkcija $\nabla \theta_{N_{max}}$, ∇f i g_k (kao posledica pretpostavke 4.1) obezbeđuje da se puštanjem limesa po K dobija

$$\lim_{k \in K} \|\nabla \theta_{N_{max}}(x_k)\| = 0,$$

što implicira

$$\nabla \theta_{N_{max}}(x^*) = 0. \quad (4.15)$$

Dakle, x^* je stacionarna tačka funkcije $\theta_{N_{max}}$.

Ukoliko je još i LICQ uslov zadovoljen, tada Jakobijeva matrica funkcije ograničenja $\nabla h_{N_{max}}(x^*)$ ima pun rang. Pošto je

$$\nabla \theta_{N_{max}}(x^*) = 2\nabla^T h_{N_{max}}(x^*) h_{N_{max}}(x^*),$$

(4.15) dovodi do zaključka da $h_{N_{max}}(x^*)$ mora da bude nula, što u stvari znači da je x^* dopustiva tačka SAA problema.

Ostalo nam je da pokažemo da je gradijent Lagranžove funkcije

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T h_{N_{max}}(x)$$

jednak nuli u x^* , odnosno da je

$$\nabla f(x^*) + \nabla^T h_{N_{max}}(x^*) \lambda = 0$$

za neko $\lambda \in \mathbb{R}^m$. U tu svrhu definišimo λ_k sa

$$\lambda_k := 2\mu_k h_{N_{max}}(x_k).$$

Tada je

$$g_k = \nabla f(x_k) + \nabla^T h_{N_{max}}(x_k) \lambda_k, \quad (4.16)$$

tj.

$$\nabla^T h_{N_{max}}(x_k) \lambda_k = g_k - \nabla f(x_k). \quad (4.17)$$

Sada nam neprekidnost funkcije $\nabla h_{N_{max}}$ implicira da za dovoljno veliko k matrice $\nabla h_{N_{max}}(x_k)$ imaju pun rang, te je $\nabla h_{N_{max}}(x_k) \nabla^T h_{N_{max}}(x_k)$ regularna matrica za dovoljno veliko k . Tako, nakon množenja (4.17) sa $\nabla h_{N_{max}}(x_k)$ sa leve strane i sređivanja, dobijamo da za dovoljno veliko k važi

$$\lambda_k = (\nabla h_{N_{max}}(x_k) \nabla^T h_{N_{max}}(x_k))^{-1} \nabla h_{N_{max}}(x_k) (g_k - \nabla f(x_k)).$$

Puštajući da $k \in K$ teži ka beskonačnosti, koristeći (4.14), sledi

$$\lim_{k \in K} \lambda_k = - (\nabla h_{N_{max}}(x^*) \nabla h_{N_{max}}(x^*)^T)^{-1} \nabla h_{N_{max}}(x^*) \nabla f(x^*).$$

Označimo izraz na desnoj strani sa λ^* . Iskoristivši zapis (4.16), činjenica (4.14), nam daje

$$0 = \lim_{k \in K} g_k = \nabla f(x^*) + \nabla^T h_{N_{max}}(x^*) \lambda^*.$$

Dakle, dobili smo da je x^* KKT tačka problema (4.2) s Lagranžovim množiteljem λ^* . \square

Implementacija algoritma 4.1 na test problemima potvrđuje da tim algoritmom dolazi do uštede u broju izračunavanja funkcije poredeći s rešavanjem SAA problema s fiksiranim punim uzorkom, kao i uporedivom heurističkom šemom za određivanje veličine uzorka. Detalji su navedeni u poglavlju 6.

Glava 5

Kazneni postupak s promenljivom veličinom uzorka za rešavanje problema s ograničenjima u formi matematičkog očekivanja

U prethodnom poglavlju smo rešavanje problema stohastičkog programiranja s ograničenjima tipa jednakosti u obliku matematičkog očekivanja (problem SP) sveli na rešavanje SAA reformulacije problema, odnosno na rešavanje aproksimativnog problema, a sada rešavamo originalan problem. To znači da nemamo više unapred određen uzorak, a samim tim determinističku konvergenciju, već stohastičku. Ostajemo pri kombinovanju kvadratnog kaznenog postupka i metode s promenljivom veličinom uzorka, koristeći linijsko pretraživanje. U svakoj iteraciji računamo sa SAA ocenom funkcije ograničenja, s tim da nemamo maksimalan uzorak, odnosno radimo s neograničenim uzorkom. I u ovom pristupu koristimo kumulativne uzorke i adaptivno ažuriranje veličine uzorka. Osim toga, adaptivno ažuriramo i kazneni parametar.

Deo 5.1 posvećujemo formulisanju algoritma i analiziranju njegovih detalja. U odeljku 5.2 ćemo pokazati da se kreiranim algoritmom generišu niz veličina uzoraka i niz kaznenih parametara koji teže ka beskonačnosti i time obezbeđuje da dobijemo skoro sigurnu konvergenciju niza iteracija ka KKT tački problema SP, pod standardnim pretpostavkama za stohastičku optimizaciju. Numerička implementacija algoritma će biti predstavljena u glavi 6.

5.1 Algoritam

Neka je funkcija $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ograničena odozdo, neka je stohastička funkcija $H : \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^m$ i neka je

$$h(x) := E[H(x, \xi)],$$

pri čemu je $\xi : \Omega \rightarrow \mathbb{R}^p$ slučajan vektor definisan na prostoru verovatnoće (Ω, \mathcal{F}, P) . Cilj nam je da dođemo do rešenja problema stohastičkog programiranja

$$\min f(x) \quad \text{tako da je} \quad h(x) = 0. \quad (5.1)$$

U tu svrhu kreiramo iterativni postupak u kojem u svakoj iteraciji za ocenu matematičkog očekivanja koristimo SAA funkciju

$$h_N(x) = \frac{1}{N} \sum_{i=1}^N H(x, \xi_i), \quad (5.2)$$

s uzorkom od $N \in \mathbb{N}$ realizacija slučajnog vektora ξ : $\xi_1, \xi_2, \dots, \xi_N$. Veličina uzorka nije ograničena i u svakoj iteraciji se određuje veličina narednog uzorka, s tim da su uzorci iste veličine jednaki i veći uzorci se dobijaju proširivanjem manjih. To znači da u k -toj iteraciji radimo s aproksimacijom problema (5.1)

$$\min f(x) \quad \text{tako da je} \quad h_{N_k}(x) = 0 \quad (5.3)$$

tj. SAA problemom s uzorkom veličine N_k . Tačnije, funkciju cilja i ograničenja ovog problema spajamo u kaznenu funkciju

$$\phi_{N_k}(x; \mu) = f(x) + \mu \theta_{N_k}(x), \quad (5.4)$$

gde je μ kazneni parametar, a

$$\theta_{N_k}(x) := \|h_{N_k}(x)\|^2$$

mera nedopustivosti. Ovim se, pored određivanja naredne iterativne tačke i nove veličine uzorka, javlja potreba i za određivanjem kaznenog parametra za narednu iteraciju. Kao i u klasičnom kvadratnom kaznenom postupku, neophodno je da kazneni parametar teži u beskonačno. To obezbeđujemo adaptivnim ažuriranjem sličnim kao i u prethodnom poglavlju, odnosno u radu Krejić i dr. [32]. Radi postizanja dobre aproksimacije funkcije ograničenja i veličina uzorka treba da raste ka beskonačnosti. Za niz kaznenih parametara je potrebno da bude monotono neopadajući, međutim to nije slučaj s nizom veličina uzorka. Dozvoljavamo da se veličina uzorka smanji, ali kontrolišemo opadanje

uz pomoć niza donjih granica veličine uzoraka. Taj niz je isto monotono neopadajući. Veličinu uzorka i njegovu donju granicu takođe adaptivno ažuriramo.

Da bismo precizirali pravila ažuriranja, pretpostavimo da u svakoj iteraciji k imamo na raspolaganju vrednost gradijenta

$$g_k := \nabla_x \phi_{N_k}(x_k; \mu_k).$$

Narednu iterativnu tačku x_{k+1} određujemo primenom linijskog pretraživanja na kaznenu funkciju $\phi_{N_k}(x_k; \mu_k)$. Za pravac pretraživanja d_k uzimamo opadajući pravac, odnosno tražimo da uslov

$$d_k^T g_k < 0$$

bude zadovoljen. Time postižemo da se može primeniti monotono linijsko pretraživanje.* Dužinu koraka α_k biramo pomoću Armižo *backtracking* procedure, te α_k zadovoljava uslov

$$\phi_{N_k}(x_k + \alpha_k d_k; \mu_k) \leq \phi_{N_k}(x_k; \mu_k) + \eta \alpha_k g_k^T d_k, \quad (5.5)$$

gde je $\eta \in (0, 1)$. S obzirom na to da pretpostavljamo da nam je funkcija cilja ograničena odozdo (pretpostavka 5.1) i da je kazneni deo kaznene funkcije nenegativan, a kaznena funkcija je ograničena odozdo, time je Armižo linijsko pretraživanje dobro definisano.

Za uobičajen izbor pravca pretraživanja: $d_k = -g_k$ ili $d_k = -B_k g_k$ s uniformno pozitivno definitnom matricom B_k , veličina $\alpha_k g_k^T d_k$ predstavlja meru opadanja kaznene funkcije. Zato se

$$dm_k := -\alpha_k g_k^T d_k$$

može upotrebiti kao ocena rastojanja tačke x_k od stacionarne tačke kaznene funkcije $\phi_{N_k}(\cdot; \mu_k)$, odnosno kao mera optimalnosti u k -toj iteraciji. Primitimo da uslov da je pravac pretraživanja opadajući obezbeđuje da je dm_k nenegativno.

Jedna od uloga mere dm_k je utvrđivanje kada da se poveća kazneni parametar. Tako, kada je

$$dm_k \leq \frac{\alpha_k}{\mu_k^2}, \quad (5.6)$$

*Moglo bi se primeniti i nemonotono linijsko pretraživanje kao što je to učinjeno u Krejić i Krklec Jerinkić [30].

smatramo da smo se dovoljno približili stacionarnoj tački kaznene funkcije s kaznenim parametrom μ_k , pa pojačavamo uslov nedopustivosti povećavanjem kaznenog parametra (uzimamo $\mu_{k+1} > \mu_k$). U analizi konvergencije algoritma od posebnog su značaja iterativne tačke x_k za koje je zadovoljen uslov (5.6), te ih izdvajamo u niz $\{z_t\}$.

Druga značajna uloga mere optimalnosti dm_k je u određivanju veličine uzorka, ali ćemo pre razmatranja te uloge formulisati glavni algoritam.

Algoritam 5.1.

K0 Ulazni parametri: $N_0 \in \mathbb{N}$, $x_0 \in \mathbb{R}^n$, $\beta, \eta \in (0, 1)$, $\mu_0 > 0$, $\rho > 1$.

K1 Staviti da je $k = 0$, $N_k = N_0$, $x_k = x_0$, $\mu_k = \mu_0$, $t = 1$, $N_0^{min} = N_0$.

K2 Ako je

$$g_k = 0 \quad \text{i} \quad h_{N_k}(x_k) = 0, \quad (5.7)$$

tada je $x_{k+1} = x_k$, $\mu_{k+1} = \mu_k$,

$$N_{k+1} = N_k + 1 \quad \text{i} \quad N_{k+1}^{min} = \max\{N_{k+1}, N_k^{min} + 1\}$$

i preći na korak K7.

K3 Odrediti pravac pretraživanja d_k koji zadovoljava uslov $d_k^T g_k < 0$.

K4 Određivanje dužine koraka α_k :

Naći najmanji nenegativan prirodan broj j takav da za $\alpha_k = \beta^j$ važi Armijo uslov (5.5).

Postaviti da je $x_{k+1} = x_k + \alpha_k d_k$ i $dm_k = -\alpha_k g_k^T d_k$.

K5 Odrediti veličinu uzorka N_{k+1} i donju granicu veličine uzorka N_{k+1}^{min} primenom algoritma 5.2.

K6 Određivanje kaznenog parametra μ_{k+1} :

Ako je $dm_k \leq \frac{\alpha_k}{\mu_k^2}$, tada je

$$\mu_{k+1} = \rho \mu_k, \quad z_t = x_k \quad \text{i} \quad t = t + 1.$$

U suprotnom postaviti da je $\mu_{k+1} = \mu_k$.

K7 Postaviti da je $k = k + 1$ i poći na korak K2.

Primetimo da se veličina uzorka određuje u koraku K5, ali se može modifikovati u koraku K2. Objašnjenje za postojanje tog koraka leži u KKT uslovima SAA problema. Naime,

$$g_k = \nabla f(x_k) + \mu_k \nabla \theta_{N_k}(x_k) = \nabla f(x_k) + 2\mu_k \nabla^T h_{N_k}(x_k) h_{N_k}(x_k),$$

pa definišući

$$\lambda_k := 2\mu_k h_{N_k}(x_k)$$

dobijamo

$$g_k = \nabla f(x_k) + \nabla^T h_{N_k}(x_k) \lambda_k.$$

Pri tome, KKT uslovi za SAA problem (5.3), s Lagranžovom funkcijom

$$\mathcal{L}_{N_k}(x, \lambda) = f(x) + \lambda^T h_{N_k}(x),$$

su

$$\begin{aligned} \nabla_x \mathcal{L}_{N_k}(x, \lambda) = \nabla f(x) + \nabla^T h_{N_k}(x) \lambda &= 0 \\ h_{N_k}(x) &= 0, \end{aligned}$$

za neko $\lambda \in \mathbb{R}^m$. Dakle, uslov (5.7) implicira da je x_k KKT tačka SAA problema posmatranog u k -toj iteraciji s Lagranžovim množiteljem λ_k . Pošto je to najbolja moguća aproksimacija rešenja problema (5.1) s uzorkom veličine N_k , to znači da preostaje da se poveća preciznost aproksimacije ograničenja, tj. da se poveća veličina uzorka, kao i da se ne koristi više uzorak te ili manje veličine.

Određivanje veličine uzorka u koraku K5 predstavlja sprovođenje algoritma 5.2. Tu se vraćamo na drugu ulogu mere optimalnosti dm_k . Rekli smo da ona predstavlja aproksimaciju rastojanja x_k od stacionarne tačke funkcije $\phi_{N_k}(\cdot; \mu_k)$. Tako, s jedne strane imamo grešku dobijenu zbog aproksimacije rešenja SAA problema, merenu s dm_k , a s druge strane grešku nastalu zbog primene SAA ocene funkcije ograničenja (grešku aproksimacije $h_{N_k}(x) \approx h(x)$). Tu grešku merimo s $e(x_k; N_k)$ pri čemu tražimo da funkcija $e(\cdot; N)$ za fiksirano $N \in \mathbb{N}$ bude ograničena odozdo pozitivnim brojem. Kao i u prethodnom poglavlju, ažuriranje veličine uzorka vršimo tako da ove dve mere budu približno jednake. Znači, kada je $dm_k > e(x_k; N_k)$ smatramo da možemo da smanjimo preciznost aproksimacije funkcije ograničenja smanjivši veličinu uzorka, naravno, ukoliko to donja granica veličine uzorka dozvoljava. Preciznije rečeno, uzimamo najveću veličinu uzorka N_{k+1} iz intervala $[N_k^{min}, N_k]$ za koju važi da je

$$dm_k \leq \frac{N_k}{N_{k+1}} e(x_k; N_{k+1}). \quad (5.8)$$

Ako ne postoji takvo N_{k+1} , tada uzimamo najmanju dozvoljenu vrednost, tj. $N_{k+1} = N_k^{min}$. Obrnuti poredak mera, $dm_k < e(x_k; N_k)$, nagoveštava da se treba povećati veličina uzorka da bi se poboljšala SAA ocena, tj. preciznost. U prethodnom poglavlju smo uzeli najmanje N_{k+1} veće od N_k koje zadovoljava uslov

$$dm_k \geq \frac{N_k}{N_{k+1}} e(x_k; N_{k+1}), \quad (5.9)$$

ali nam je uzorak bio ograničen, pa je nova veličina uzorka mogla biti najviše veličina punog uzorka. Sada nas, zbog neograničenosti uzorka, uslov (5.9) može dovesti do veoma velikog uzorka i to nepotrebno. Na primer kada za nepreciznu SAA ocenu ($e(x_k; N_k)$ je veliko) dođemo veoma blizu stacionarne tačke kaznene funkcije definisane s veličinom uzorka N_k (kada je dm_k veoma malo), tada je mera nedopustivosti s veličinom uzorka N_k najverovatnije mala, ali se ona može znatno promeniti i malim povećavanjem uzorka. Ukoliko se povećanjem uzorka poveća mera nedopustivosti, tada nam je to pokazatelj pogoršane mere optimalnosti kaznene funkcije s uvećanim uzorkom, a samim tim nema potrebe za daljim proširivanjem uzorka. Ovo razmatranje nas je dovelo do modifikacije uslova koji određuje koliko će se povećati uzorak. Tako uzimamo najmanje N_{k+1} veće od N_k koje zadovoljava uslov (5.9) ili uslov

$$\|h_{N_{k+1}}(x_k)\| \geq \frac{N_k}{N_{k+1}(N_{k+1} - N_k)} e(x_k; N_{k+1}). \quad (5.10)$$

Količnik $N_k / (N_{k+1}(N_{k+1} - N_k))$ iz prethodnog uslova, kao i količnik N_k / N_{k+1} u uslovima (5.8) i (5.9), ima ulogu da dozvoljava manje promene u veličini uzorka kada su u pitanju manji uzorci jer tada preciznost nije tako dobra. Kod velikih uzoraka on dozvoljava veće promene pošto je tada preciznost dobra, pa mera optimalnosti postaje precizniji pokazatelj blizine rešenja originalnom problemu.

Primetimo da novi uslov ne zahteva dodatne proračune kada se za meru preciznosti $e(x_k; N_k)$ koristi veličina $h_{N_k}(x_k)$, kao što je slučaj s merom dužine intervala poverenja za $h_{N_k}(x_k)$.

Analiza konvergencije će nam pokazati da uslov (5.10) nema ulogu u postizanju konvergencije. Međutim, ima veliki uticaj na broj izračunavanja vrednosti funkcija ograničenja što potvrđuju test primeri prikazani u poglavlju 6. Na tim test primerima je ilustrovan i uticaj količnika u uslovu (5.10), odnosno prikazano je ponašanje niza veličina uzoraka u zavisnosti od toga da li je upotrebljen količnik N_k / N_{k+1} , N_k / N_{k+1}^2 ili $N_k / (N_{k+1}(N_{k+1} - N_k))$.

Nakon što se odredi nova veličina uzorka u algoritmu 5.2, utvrđuje se nova donja granica veličine uzorka. Ključna osobina pravila ažuriranja niza N_k^{min} je

da niz bude neopadajući, s tim da se donja granica poveća kad god promena veličine uzorka ne doprinese zadovoljavajućem opadanju mere nedopustivosti. Opadanje smatramo zadovoljavajućim ako je prosečan pad mere nedopustivosti $\theta_{N_{k+1}}$ od poslednjeg prelaska na uzorak veličine N_{k+1} najmanje jednak delu mere preciznosti tog uzorka u novoj iterativnoj tački. Udeo mere preciznosti je definisan pozitivnom rastućom funkcijom γ čija je uloga da obezbedi veće restrikcije za veće uzorke radi postizanja višeg nivoa preciznosti.

Algoritam 5.2.

K0 Ulazni parametri: $dm_k, e(x_k; N_k), x_k, N_k, N_k^{min}$.

K1 Određivanje veličine uzorka N_{k+1} :

- 1) Ako je $dm_k = e(x_k; N_k)$, tada je $N_{k+1} = N_k$.
- 2) Ako je $dm_k > e(x_k; N_k)$, tada uzeti da je $N = N_k$.
Dokle god je

$$dm_k > \frac{N_k}{N} e(x_k; N) \quad \text{i} \quad N > N_k^{min},$$

stavljati da je $N = N - 1$.

Uzeti da je $N_{k+1} = N$.

- 3) Ako je $dm_k < e(x_k; N_k)$, tada je $N = N_k + 1$.
Dokle god je

$$dm_k < \frac{N_k}{N} e(x_k; N) \quad \text{i} \quad \|h_N(x_k)\| < \frac{N_k}{N(N - N_k)} e(x_k; N)$$

stavljati da je $N = N + 1$.

Uzeti da je $N_{k+1} = N$.

K2 Određivanje donje granice veličine uzorka N_{k+1}^{min} :

Ako je $N_{k+1} \neq N_k$ i postoji $i \in \{1, \dots, k-1\}$ takvo da je $N_i = N_{k+1}$, naći $l(k)$ za koje važi da

$$l(k) = \max\{i \in \{1, \dots, k-1\} : N_i = N_{k+1}, N_{i-1} \neq N_{k+1}\}.$$

Ako je

$$\frac{\theta_{N_{k+1}}(x_{l(k)}) - \theta_{N_{k+1}}(x_{k+1})}{k+1-l(k)} < \gamma(N_{k+1}) e(x_{k+1}; N_{k+1}), \quad (5.11)$$

tada uzeti da je

$$N_{k+1}^{min} = \max\{N_{k+1}, N_k^{min} + 1\}. \quad (5.12)$$

U suprotnom uzeti da je $N_{k+1}^{min} = N_k^{min}$.

Analizirajmo algoritam 5.2. Ako je $N_k^{min} \leq N_k$, realizacija koraka K1 1) ili 3) (s korakom K2) daje

$$N_k^{min} \leq N_k \leq N_{k+1}^{min} \leq N_{k+1}. \quad (5.13)$$

Kada se realizuje korak K1 2) tako da je $N_{k+1} > N_k^{min}$, uslovi (5.13) ostaju obezbeđeni. Međutim, ako se N_k smanji na $N_{k+1} = N_k^{min}$, uslov $N_{k+1}^{min} \leq N_{k+1}$ biva narušen kada se koristi uzorak koji je već ranije upotrebljen i nema dovoljne redukcije mere nedopustivosti (kada je nejednakost (5.11) zadovoljena). Tada se povećava donja granica veličine uzorka, te je

$$N_{k+1}^{min} = N_k^{min} + 1 > N_{k+1}.$$

Ipak, u tom slučaju važi

$$N_{k+1} = N_{k+1}^{min} - 1 \quad \text{i} \quad N_{k+1}^{min} \geq N_k^{min}.$$

Pošto se za $N_k = N_k^{min} - 1$ dobija da korak K1 3) daje

$$N_{k+1} > N_k = N_k^{min} - 1 \quad \text{i} \quad N_{k+1}^{min} \geq N_k^{min},$$

a koraci K1 1) i 2) daju

$$N_{k+1} = N_k = N_k^{min} - 1 = N_{k+1}^{min} - 1,$$

zaključujemo da algoritam 5.2 obezbeđuje da se u k -toj iteraciji sa veličinom uzorka $N_k \geq N_k^{min} - 1$ dobija

$$N_{k+1} \geq N_{k+1}^{min} - 1 \quad \text{i} \quad N_{k+1}^{min} \geq N_k^{min}. \quad (5.14)$$

Primetimo da je algoritam 5.1 definisan tako da je veličina uzorka određena korakom K2 ili K5. Kako smo već razmotrili korak K5, odnosno algoritam 5.2, ostaje nam da vidimo šta se dešava s korakom K2. Ako je $N_k \geq N_k^{min} - 1$ i važi (5.7), tada je

$$N_{k+1} = N_k + 1 \geq N_k^{min} \quad \text{i} \quad N_{k+1}^{min} = \max\{N_{k+1}, N_k^{min} + 1\} > N_k^{min}.$$

Dakle, algoritmi 5.1 i 5.2 s početnim uslovom $N_0^{min} = N_0$ obezbeđuju da za svako k važe nejednakosti (5.14). Primetimo još da se ulaskom u korak K2 algoritma 5.1 donja granica veličine uzorka sigurno povećava.

5.2 Analiza konvergencije

U ovom delu će biti predstavljene osobine navedenog algoritma i dokazana njegova konvergencija. U tu svrhu potrebna nam je pretpostavka koja obezbeđuje neprekidnu diferencijabilnost aproksimacije h_N za svaki prirodan broj N .

Pretpostavka 5.1. *Funkcija cilja f je ograničena odozdo na \mathbb{R}^n , dok su funkcije $f(\cdot)$ i $H(\cdot, \xi)$ neprekidno diferencijabilne na \mathbb{R}^n za svako ξ .*

Pretpostavka 5.1 obezbeđuje i neprekidnu diferencijabilnost kaznene funkcije (5.4). Pošto je mera nedopustivosti nenegativna, a kazneni parametar je pozitivan, za kaznenu funkciju imamo osiguranu i ograničenost odozdo. Time je postignuto da je dobro definisan problem minimizacije kaznene funkcije, kao i da je i linijsko pretraživanje (određivanje dužine koraka pretraživanja – korak K4 algoritma 5.1) dobro definisano.

Radi postizanja da niz veličina uzoraka teži u beskonačno, potrebna nam je još i sledeća pretpostavka.

Pretpostavka 5.2. *Za funkciju $e : \mathbb{R}^n \times \mathbb{N} \rightarrow \mathbb{R}_+$ i proizvoljan fiksiran prirodan broj N postoji e_N takav da*

$$e(x; N) \geq e_N > 0 \quad \text{za svako } x \in \mathbb{R}^n.$$

Ovo je blag uslov pošto postoji veliki broj funkcija koje ga zadovoljavaju. Najjednostavniji primer bi mogao biti $e(x; N) = 1/N$. Pretpostavku 5.2 zadovoljava i ocena mere dužine intervala poverenja aproksimacije h_N uvećana za $1/N$, tj.

$$e(x; N) = \hat{\sigma}_N^2(x) \frac{z_{\alpha/2}}{\sqrt{N}} + \frac{1}{N},$$

gde je $\hat{\sigma}_N^2(x)$ uzoračka varijansa, a $z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2)$ sa uobičajenim vrednostima $\alpha = 0.9$ i $\alpha = 0.95$, pri čemu je $\Phi(\cdot)$ kumulativna funkcija raspodele standardne normalne raspodele.

Prvo dokazujemo da niz veličina uzoraka ne može da postane stagnirajući.

Lema 5.1. *Neka važe pretpostavke 5.1 i 5.2. Tada ne postoji $q \in \mathbb{N}$ takvo da, za algoritmom 5.1 generisane veličine uzorka N_k , važi da je $N_{k+1} = N_k$ za svako $k \geq q$.*

Dokaz. Pretpostavimo suprotno, da postoje prirodni brojevi \bar{q} i \bar{N} za koje važi

$$N_{k+1} = N_k = \bar{N} \quad \text{za svako } k \geq \bar{q}. \quad (5.15)$$

Pokazujemo da taj uslov implicira

$$\liminf_{k \rightarrow \infty} dm_k = 0. \quad (5.16)$$

Primetimo da (5.15) isključuje realizaciju koraka K2 algoritma 5.1 i posmatrajmo kazneni parametar. Postoje dve mogućnosti:

S1. μ_k se konačno mnogo puta menja,

S2. μ_k se beskonačno mnogo puta menja.

S obzirom na to da je po konstrukciji algoritma 5.1 niz kaznenih parametara neopadajući, slučaj S1 podrazumeva da postoji $\bar{\mu}$ za koje je $\mu_k = \bar{\mu}$ za svako dovoljno veliko k . Bez gubitka opštosti važi

$$\mu_k = \bar{\mu} \quad \text{za svako } k \geq \bar{q}.$$

Kako su veličina uzorka i kazneni parametar fiksirani za svako $k \geq \bar{q}$, radi kraćeg zapisa uvodimo oznaku

$$\bar{\phi}(x_k) := \phi_{\bar{N}}(x_k; \bar{\mu}).$$

Korak K4 algoritma 5.1 daje da za svako $k \geq \bar{q}$ važi

$$\bar{\phi}(x_{k+1}) \leq \bar{\phi}(x_k) - \eta dm_k,$$

što implicira da je

$$\bar{\phi}(x_{k+1}) \leq \bar{\phi}(x_{\bar{q}}) - \eta \sum_{i=\bar{q}}^k dm_i.$$

Posledica pretpostavke 5.1 je postojanje realnog broja M za koji važi da je $\bar{\phi}(x_k) \geq M$ za svako k , te dobijamo da

$$M \leq \bar{\phi}(x_{\bar{q}}) - \eta \sum_{i=\bar{q}}^k dm_i,$$

odnosno

$$\sum_{i=\bar{q}}^k dm_i \leq \frac{1}{\eta} (\bar{\phi}(x_{\bar{q}}) - M).$$

Puštajući da $k \rightarrow \infty$ sledi

$$0 \leq \sum_{i=\bar{q}}^{\infty} dm_i \leq \frac{1}{\eta} (\bar{\phi}(x_{\bar{q}}) - M) < \infty,$$

pa je

$$\lim_{i \rightarrow \infty} dm_i = 0.$$

Što se tiče slučaja S2, to što svaka promena kaznenog parametra μ_k predstavlja njegov porast znači da imamo beskonačno mnogo iteracija u kojima je

$$\mu_{k+1} = \rho \mu_k \quad \text{sa } \rho > 1. \quad (5.17)$$

Iz toga sledi da je niz $\{\mu_k\}$ neograničen, te

$$\lim_{k \rightarrow \infty} \mu_k = \infty. \quad (5.18)$$

Neka su k_1, k_2, \dots iteracije u kojima je došlo do porasta kaznenog parametra, tj. za $k = k_i$, $i = 1, 2, \dots$, važi (5.17), što na osnovu koraka K6 algoritma 5.1 znači da je zadovoljen uslov

$$dm_{k_i} \leq \frac{\alpha_{k_i}}{\mu_{k_i}^2}.$$

Pošto dužina koraka pripada intervalu $(0, 1]$, dobijamo

$$dm_{k_i} \leq \frac{1}{\mu_{k_i}^2}.$$

Odavde (5.18) i nenegativnost dm_{k_i} obezbeđuju da važi

$$\lim_{i \rightarrow \infty} dm_{k_i} = 0.$$

Pokazali smo da i S1 i S2 impliciraju (5.16), odnosno da u oba slučaja postoji iteracija $s \geq \bar{q}$ za koju važi $dm_s < e_{\bar{N}}$ gde je $e_{\bar{N}}$ konstanta iz pretpostavke 5.2 (jer je \bar{N} fiksirano). Dakle, važi

$$dm_s < e_{\bar{N}} \leq e(x_s; \bar{N}),$$

te se realizuje korak K1 3) algoritma 5.2, što implicira $N_{s+1} > N_s$. Međutim, to je u kontradikciji sa (5.15), pa polazna pretpostavka nije tačna i time je dokazano tvrđenje. \square

Primetimo da je prethodni rezultat deterministički, iako su nizovi koji se generišu glavnim algoritmom stohastički. Razlog tome je što taj rezultat ne zavisi od realizacije uzorka – po tvrđenju je prvo uzeta proizvoljna realizacija uzorka, a tek nakon toga je razmatrano kakve osobine su zadovoljene, pa za svaku (ne skoro svaku) realizaciju važi pokazani rezultat.

Isti je slučaj i sa sledećim rezultatom, kada pokazujemo da niz veličina uzoraka teži ka beskonačnosti. Pre nego što to učinimo, potrebno je da formalno definišemo funkciju γ .

Pretpostavka 5.3. Funkcija $\gamma: \mathbb{N} \rightarrow (0, 1)$ je rastuća i zadovoljava uslov

$$\lim_{N \rightarrow \infty} \gamma(N) = 1.$$

Na primer $\gamma(N) := \exp(-1/N)$ zadovoljava uslove gornje pretpostavke.

Teorema 5.1. Ako su pretpostavke 5.1–5.3 zadovoljene, tada niz veličina uzoraka $\{N_k\}_{k \in \mathbb{N}_0}$ generisan algoritmom 5.1 ima osobinu

$$\lim_{k \rightarrow \infty} N_k = \infty.$$

Dokaz. Do tvrđenja teoreme dolazimo posmatrajući niz donjih granica veličina uzoraka $\{N_k^{min}\}_{k \in \mathbb{N}_0}$. Dokaz je konstruisan na sledeći način. Prvo posmatramo slučaj kada je niz $\{N_k^{min}\}$ neograničen, koji direktno implicira tvrđenje teoreme. Zatim posmatramo slučaj kada je $\{N_k^{min}\}$ ograničen. Tada rezultat pokazujemo kontrapozicijom.

Algoritam 5.1 je definisan tako da se generiše beskonačan niz veličina uzoraka $\{N_k\}$, i to kroz korake K2 i K5, ali i beskonačan niz donjih granica veličina uzoraka $\{N_k^{min}\}$. Videli smo da elementi tih nizova zadovoljavaju nejednakost

$$N_k \geq N_k^{min} - 1 \quad \text{za svako } k.$$

To znači da uslov

$$\lim_{k \rightarrow \infty} N_k^{min} = \infty \quad (5.19)$$

obežbeđuje tvrđenje teoreme. Kako je po konstrukciji algoritama 5.1 i 5.2 niz $\{N_k^{min}\}$ neopadajući (druga nejednakost u (5.14)), ako je još i neograničen, važi uslov (5.19).

Pogledajmo sada šta se dešava kada je niz $\{N_k^{min}\}$ ograničen. Tada postoje q i N_{max}^{min} iz \mathbb{N} za koje je

$$N_k^{min} = N_{max}^{min} \quad \text{za svako } k \geq q. \quad (5.20)$$

S obzirom na to da se u koraku K2 algoritma 5.1 povećava donja granica veličine uzoraka ($N_{k+1}^{min} > N_k^{min}$), veličina uzorka N_k se može samo konačno mnogo puta promeniti kroz korak K2 algoritma 5.1. To znači da postoji $\bar{q} \geq q$ za koje uslov (5.7) nije zadovoljen ni za jedno $k \geq \bar{q}$, pa se tada veličina uzorka određuje u koraku K5 algoritma 5.1, odnosno na osnovu algoritma 5.2. Pretpostavimo sada suprotno, da ne važi tvrđenje teoreme, odnosno da postoji ograničen beskonačan podniz niza $\{N_k\}$. Tada postoje $\bar{N} \in \mathbb{N}$ i beskonačan podniz $\{\bar{k}_1, \bar{k}_2, \dots\} \subset \{k \in \mathbb{N} : k \geq \bar{q}\}$ takvi da

$$N_{\bar{k}_i+1} = \bar{N} \quad \text{za svako } i = 1, 2, \dots$$

Bez gubitka opštosti možemo uzeti da je

$$N_j \neq \bar{N} \quad \text{za svako } j \in \{k \in \mathbb{N} : k \geq \bar{q}, k \neq \bar{k}_i, i = 1, 2, \dots\}.$$

Na osnovu leme 5.1 niz $\{N_k\}$ ne može da stagnira, pa mora da postoji beskonačan podniz $\{k_1, k_2, \dots\} \subset \{\bar{k}_1, \bar{k}_2, \dots\}$ sa osobinom

$$\bar{N} = N_{k_i+1} \neq N_{k_i} \quad \text{za } i = 1, 2, \dots \quad (5.21)$$

Možemo pretpostaviti, bez gubitka opštosti, da podniz $\{k_1, k_2, \dots\}$ sadrži sve indekse podniza $\{\bar{k}_1, \bar{k}_2, \dots\}$ za koje važi osobina (5.21). Kako važi (5.20), korak K2 algoritma 5.2 implicira

$$\frac{\theta_{\bar{N}}(x_{l(k_i)}) - \theta_{\bar{N}}(x_{k_i+1})}{k_i + 1 - l(k_i)} \geq \gamma(\bar{N}) e(x_{k_i+1}; \bar{N})$$

za svako $i = 1, 2, \dots$. Radi jednostavnijeg zapisivanja, uvedimo oznaku $k_i^+ := k_i + 1$. Primetimo da je tada $l(k_i) = k_{i-1}^+$. Tako, prethodna nejednakost s pretpostavkama 5.2 i 5.3 i činjenicom da je $k_i + 1 - l(k_i) \geq 1$ implicira nejednakost

$$\theta_{\bar{N}}(x_{k_i^+}) - \theta_{\bar{N}}(x_{k_i^+}) \geq \gamma(\bar{N}) e_{\bar{N}} > 0,$$

za svako $i = 2, 3, \dots$. Međutim, to je u kontradikciji s ograničenošću θ_N odozdo (zbog njene nenegativnosti). Dakle, ne može da postoji beskonačan podniz niza veličina uzoraka koji je ograničen, te je teorema dokazana. \square

Preostalo nam je da navedemo još dve pretpostavke koje su nam potrebne radi obezbeđivanja konvergencije. Prvom pretpostavkom je definisan pravac pretraživanja.

Pretpostavka 5.4. *Pravac pretraživanja je u obliku $d_k = -B_k g_k$ gde je B_k simetrična, uniformno ograničena i uniformno pozitivno definitna matrica po iteracijama.*

Primetimo da je prethodna pretpostavka zadovoljena za pravac negativnog pada jer je u tom slučaju B_k jedinična matrica za svako k . Da bi za Kvazi-Njutnove pravce (na primer BFGS pravac) važila pretpostavka 5.4 potrebno je primeniti meru zaštite (*safeguard*) prilikom njihove upotrebe.

Pozitivna definitnost matrice B_k implicira da za svako k važi relacija

$$\lambda_{\min}^k \|g_k\|^2 \leq g_k^T B_k g_k \leq \lambda_{\max}^k \|g_k\|^2,$$

gde je λ_{min}^k najmanji karakteristični koren matrice B_k , a λ_{max}^k najveći. Dalje, uniformna ograničenost i uniformna pozitivna definitnost obezbeđuju postojanje pozitivne donje granice λ_{min} niza $\{\lambda_{min}^k\}_{k \in \mathbb{N}}$ i gornje granice $\lambda_{max} > 0$ niza $\{\lambda_{max}^k\}_{k \in \mathbb{N}}$. Stoga za svako k važi

$$0 \leq \lambda_{min} \|g_k\|^2 \leq g_k^T B_k g_k \leq \lambda_{max} \|g_k\|^2,$$

odnosno

$$0 \leq \lambda_{min} \|g_k\|^2 \leq -g_k^T d_k. \quad (5.22)$$

Iz prethodne nejednakosti sledi implikacija

$$\lim_{k \rightarrow \infty} g_k^T d_k = 0 \implies \lim_{k \rightarrow \infty} \lambda_{min} \|g_k\|^2 = 0 \implies \lim_{k \rightarrow \infty} \|g_k\| = 0.$$

Pored toga, na osnovu osobina norme matrice sledi

$$\|d_k\| = \|B_k g_k\| \leq \|B_k\| \|g_k\| = \lambda_{max}^k \|g_k\|,$$

pa je

$$\|d_k\| \leq \lambda_{max} \|g_k\| \quad \text{za svako } k. \quad (5.23)$$

Dakle, pretpostavka 5.4 obezbeđuje da je pravac pretraživanja d_k ograničen kad god je g_k ograničen, kao i da konvergencija niza $\{g_k^T d_k\}$ ka nuli povlači za sobom da i niz $\{g_k\}$ konvergira ka nuli.

U poglavlju 3 smo videli da su konzistentnost i nepristrasnost značajne osobine SAA ocena. Ulogu u obezbeđivanju tih stohastičkih osobina ocene (5.2) ima naredna pretpostavka.

Pretpostavka 5.5. ξ_1, ξ_2, \dots su nezavisni i imaju istu raspodelu, a na proizvoljnom kompaktnom podskupu od \mathbb{R}^n postoje P -integrabilne funkcije koje dominiraju nad funkcijom H i njenom Jakobijevom matricom ∇H .

Kao prvo, pretpostavka 5.5, zajedno s pretpostavkom 5.1, obezbeđuje da su funkcije h i ∇h dobro definisane i diferencijabilne, kao i da za funkciju ograničenja važi

$$\nabla h(x) = \nabla E[H(x, \xi)] = E[\nabla H(x, \xi)].$$

Dalje, ocene $h_N(x)$ i $\nabla h_N(x)$ su nepristrasne, tj.

$$E[h_N(x)] = h(x) \quad \text{i} \quad E[\nabla h_N(x)] = \nabla h(x).$$

Pored toga, na osnovu jakog zakona velikih brojeva, za proizvoljan kompaktan skup $S \subset \mathbb{R}^n$

$$\lim_{N \rightarrow \infty} \max_{x \in S} \|h_N(x) - h(x)\| = 0 \quad \text{s.s.} \quad (5.24)$$

i

$$\lim_{N \rightarrow \infty} \max_{x \in S} \|\nabla h_N(x) - \nabla h(x)\| = 0 \quad \text{s.s.} \quad (5.25)$$

Nastavljamo s lemapa u kojima pokazujemo da za kaznenu funkciju i njen gradijent s fiksim kaznenim parametrom važi uniformna konvergencija na proizvoljnom kompaktnom podskupu od \mathbb{R}^n . Ista osobina važi i za gradijent mere nedopustivosti.

Lema 5.2. *Neka su zadovoljene pretpostavke 5.1 i 5.5. Tada za proizvoljan neprazan kompaktan skup $S \subset \mathbb{R}^n$ i kazneni parametar $\mu > 0$ važi*

$$\lim_{N \rightarrow \infty} \varphi_N^S = 0 \quad \text{s.s.},$$

gde je

$$\varphi_N^S := \max_{x \in S} |\phi_N(x; \mu) - \phi(x; \mu)|,$$

a

$$\phi(x; \mu) := f(x) + \mu \theta(x)$$

kaznena funkcija originalnog problema (5.1) s kaznenim parametrom μ i merom nedopustivosti

$$\theta(x) := \|h(x)\|^2.$$

Dokaz. Pošto je kazneni parametar fiksan, zapis možemo skratiti koristeći oznake:

$$\phi_N(x) := \phi_N(x; \mu) \quad \text{i} \quad \phi(x) := \phi(x; \mu).$$

Kako je

$$\phi_N(x) - \phi(x) = \mu (\theta_N(x) - \theta(x))$$

i

$$\theta_N(x) - \theta(x) = h_N(x)^T h_N(x) - h(x)^T h(x),$$

razmotrimo normu razlike funkcije ograničenja i njene ocene.

$$\begin{aligned} \|h_N(x) - h(x)\|^2 &= h_N(x)^T h_N(x) - 2h_N(x)^T h(x) + h(x)^T h(x) = \\ &= \theta_N(x) - \theta(x) - 2(h_N(x) - h(x))^T h(x) \end{aligned}$$

Sada je

$$\begin{aligned}\varphi_N^S &= \max_{x \in S} |\mu (\|h_N(x) - h(x)\|^2 + 2(h_N(x) - h(x))^T h(x))| \leq \\ &\leq \mu \max_{x \in S} \|h_N(x) - h(x)\|^2 + 2\mu \max_{x \in S} |(h_N(x) - h(x))^T h(x)|,\end{aligned}$$

pa je na osnovu Koši-Švarcove nejednakosti

$$\varphi_N^S = \mu \max_{x \in S} \|h_N(x) - h(x)\|^2 + 2\mu \max_{x \in S} \|h_N(x) - h(x)\| \|h(x)\|.$$

Kako je S neprazan kompaktan skup, h neprekidna funkcija, μ fiksirana vrednost i važi (5.24), dobijamo da φ_N^S skoro sigurno teži nuli kad $N \rightarrow \infty$. \square

Lema 5.3. *Ako su zadovoljene pretpostavke 5.1 i 5.5, tada za proizvoljan neprazan kompaktan skup $S \subset \mathbb{R}^n$ važi*

$$\lim_{N \rightarrow \infty} \max_{x \in S} \|\nabla \theta_N(x) - \nabla \theta(x)\| = 0 \quad s.s.$$

Dokaz. Krenimo od razlike gradijenata mera nedopustivosti

$$\nabla \theta_N(x) - \nabla \theta(x) = \nabla h_N(x)^T h_N(x) - \nabla h(x)^T h(x).$$

Primetimo da je

$$\begin{aligned}& (\nabla h_N(x) - \nabla h(x))^T (h_N(x) - h(x)) = \\ &= \nabla h_N(x)^T h_N(x) - \nabla h_N(x)^T h(x) - \nabla h(x)^T h_N(x) + \nabla h(x)^T h(x) = \\ &= \nabla \theta_N(x) - \nabla \theta(x) + \nabla h(x)^T h(x) - \\ &\quad - \nabla h_N(x)^T h(x) - \nabla h(x)^T (h_N(x) - h(x)) = \\ &= \nabla \theta_N(x) - \nabla \theta(x) + \\ &\quad + (\nabla h(x) - \nabla h_N(x))^T h(x) - \nabla h(x)^T (h_N(x) - h(x)),\end{aligned}$$

pa, koristeći da je norma proizvoda matrice i vektora manja ili jednaka proizvodu norme matrice i norme vektora, dobijamo

$$\begin{aligned}& \max_{x \in S} \|\nabla \theta_N(x) - \nabla \theta(x)\| \leq \\ &\leq \max_{x \in S} \left(\|(\nabla h_N(x) - \nabla h(x))^T (h_N(x) - h(x))\| + \right. \\ &\quad \left. + \|(\nabla h(x) - \nabla h_N(x))^T h(x)\| + \|\nabla h(x)^T (h_N(x) - h(x))\| \right) \leq \\ &\leq \max_{x \in S} \|\nabla h_N(x) - \nabla h(x)\| \|h_N(x) - h(x)\| + \\ &\quad + \max_{x \in S} \|\nabla h(x) - \nabla h_N(x)\| \|h(x)\| + \max_{x \in S} \|\nabla h(x)\| \|h_N(x) - h(x)\|.\end{aligned}$$

Kompaktnost skupa S i neprekidnost funkcija h i ∇h zajedno s (5.24) i (5.25) impliciraju da $\max_{x \in S} \|\nabla \theta_N(x) - \nabla \theta(x)\|$ skoro sigurno konvergira ka nuli kada $N \rightarrow \infty$. \square

S obzirom na to da je

$$\nabla \phi_N(x; \mu) - \nabla \phi(x; \mu) = 2\mu (\nabla \theta_N(x) - \nabla \theta(x)),$$

uniformna konvergencija gradijenta kaznene funkcije na proizvoljnom kompaktnom podskupu od \mathbb{R}^n sledi na osnovu prethodne leme.

Posledica 5.1. *Neka su zadovoljene pretpostavke 5.1 i 5.5. Tada za proizvoljan neprazan kompaktan skup $S \subset \mathbb{R}^n$ i kazneni parametar $\mu > 0$ važi*

$$\lim_{N \rightarrow \infty} \max_{x \in S} \|\nabla \phi_N(x; \mu) - \nabla \phi(x; \mu)\| = 0 \quad \text{s.s.}$$

Gore pokazane uniformne konvergencije nam obezbeđuju konzistentnost naših ocena, što ćemo i formulirati u lemi.

Lema 5.4. *Neka su zadovoljene pretpostavke 5.1 i 5.5. Tada za proizvoljan neprazan kompaktan skup $S \subset \mathbb{R}^n$ i proizvoljnu funkciju*

$$\psi \in \{h, \nabla h, \nabla \theta, \phi(\cdot; \mu), \nabla \phi(\cdot; \mu)\},$$

sa fiksnim kaznenim parametrom $\mu > 0$, važi

$$\lim_{\substack{x \rightarrow x^* \\ N \rightarrow \infty}} \psi_N(x) = \psi(x^*) \quad \text{s.s.}$$

Dokaz. Pod navedenim pretpostavkama važe: (5.24), (5.25), leme 5.2 i 5.3 i posledica 5.1, te sledi

$$\lim_{N \rightarrow \infty} \max_{x \in S} \|\psi_N(x) - \psi(x)\| = 0 \quad \text{s.s.} \quad (5.26)$$

Kako je

$$\begin{aligned} \|\psi_N(x) - \psi(x^*)\| &\leq \|\psi_N(x) - \psi(x)\| + \|\psi(x) - \psi(x^*)\| \leq \\ &\leq \max_{x \in S} \|\psi_N(x) - \psi(x)\| + \|\psi(x) - \psi(x^*)\|, \end{aligned}$$

puštajući da $x \rightarrow x^*$ i $N \rightarrow \infty$, neprekidnost funkcije ψ i uniformna konvergencija (5.26) nam daju željenu skoro sigurnu konvergenciju. \square

S obzirom na to da naš posmatrani problem ima ograničenja, prikazani algoritam mora da obezbedi dopustivost. Kao i u standardnom kvadratnom kaznenom postupku, to se postiže puštanjem kaznenog parametra ka beskonačnosti. Za algoritam 5.1 možemo da pokažemo da niz kaznenih parametara skoro sigurno konvergira ka beskonačnosti jedino kada se korak K2 izvrši samo konačno mnogo puta. Međutim, to neće biti prepreka za postizanje dopustivosti, što će se videti u poslednjoj teoremi.

Teorema 5.2. *Neka važe pretpostavke 5.1–5.5 i neka je niz iteracija $\{x_k\}_{k \in \mathbb{N}_0}$ generisan algoritmom 5.1 ograničen. Ako postoji konačan prirodan broj q za koji uslov (5.7) nije zadovoljen ni za jedno $k \geq q$, tada generisan niz kaznenih parametara $\{\mu_k\}_{k \in \mathbb{N}_0}$ ima osobinu*

$$\lim_{k \rightarrow \infty} \mu_k = \infty \quad \text{s.s.}$$

Dokaz. Tvrdjenje ćemo pokazati kontrapozicijom. Pretpostavićemo da je niz kaznenih parametara $\{\mu_k\}$ ograničen i razmatraćemo niz dužina koraka $\{\alpha_k\}$. Pošto je dužina koraka pozitivna, slučajevi

$$\liminf_{k \rightarrow \infty} \alpha_k = 0 \quad \text{ i } \quad \liminf_{k \rightarrow \infty} \alpha_k \geq a > 0$$

su komplementarni. Pokazaćemo da oba slučaja dovode do kontradikcije.

Pretpostavimo da je niz $\{\mu_k\}$ ograničen, odnosno da postoje $\bar{q} \geq q$ i $\bar{\mu} > 0$ takvi da je

$$\mu_k = \bar{\mu} \quad \text{ za svako } k \geq \bar{q}.$$

Tada korak K6 algoritma 5.1 implicira

$$dm_k > \frac{\alpha_k}{\mu_k^2} \quad \text{ za svako } k \geq \bar{q}, \quad (5.27)$$

odnosno

$$-g_k^T d_k > \frac{1}{\bar{\mu}^2} > 0 \quad \text{ za svako } k \geq \bar{q}. \quad (5.28)$$

Zbog pretpostavke da je niz iteracija $\{x_k\}$ ograničen, postoji kompaktan skup S takav da je

$$\{x_k\}_{k \in \mathbb{N}_0} \subseteq S.$$

Dalje, na osnovu teoreme 5.1, N_k teži ka beskonačnosti, pa lema 5.2 implicira postojanje niza $\{\varphi_{N_k}^S\}$ za koji važi

$$|\phi_{N_k}(x_j; \bar{\mu}) - \phi(x_j; \bar{\mu})| \leq \varphi_{N_k}^S \quad \text{ za svako } k, j \geq \bar{q} \quad (5.29)$$

i $\varphi_{N_k}^S$ skoro sigurno teži nuli.

U cilju da dođemo do kontradikcije, pretpostavimo sada da je niz dužina koraka $\{\alpha_k\}$ ograničen odozdo pozitivnim brojem, odnosno da postoji $\bar{\alpha} > 0$ takvo da je

$$\alpha_k \geq \bar{\alpha} \quad \text{za svako } k \geq \bar{q}.$$

Stoga (5.27) implicira

$$dm_k \geq \frac{\bar{\alpha}}{\bar{\mu}^2} \quad \text{za svako } k \geq \bar{q},$$

te iz zadovoljenog Armijo uslova u koraku K4 algoritma 5.1 sledi

$$\phi_{N_k}(x_{k+1}; \bar{\mu}) \leq \phi_{N_k}(x_k; \bar{\mu}) - \eta \frac{\bar{\alpha}}{\bar{\mu}^2}.$$

Kako po (5.29) za svako $k \geq \bar{q}$ važi:

$$\phi(x_{k+1}; \bar{\mu}) - \phi_{N_k}(x_{k+1}; \bar{\mu}) \leq \varphi_{N_k}^S \quad \text{i} \quad \phi_{N_k}(x_k; \bar{\mu}) - \phi(x_k; \bar{\mu}) \leq \varphi_{N_k}^S,$$

dobija se

$$\phi(x_{k+1}; \bar{\mu}) - \varphi_{N_k}^S \leq \phi(x_k; \bar{\mu}) + \varphi_{N_k}^S - \eta \frac{\bar{\alpha}}{\bar{\mu}^2},$$

odnosno

$$\phi(x_{k+1}; \bar{\mu}) \leq \phi(x_k; \bar{\mu}) + 2\varphi_{N_k}^S - \frac{\eta \bar{\alpha}}{\bar{\mu}^2} \quad \text{za svako } k \geq \bar{q}.$$

Pošto $\varphi_{N_k}^S \rightarrow 0$ s.s., postoji $\bar{k} \geq \bar{q}$ za koje skoro sigurno važi da je

$$\varphi_{N_k}^S \leq \frac{\eta \bar{\alpha}}{3\bar{\mu}^2} \quad \text{za svako } k \geq \bar{k},$$

a to implicira

$$\phi(x_{k+1}; \bar{\mu}) \leq \phi(x_k; \bar{\mu}) - \frac{\eta \bar{\alpha}}{3\bar{\mu}^2} \quad \text{s.s. za svako } k \geq \bar{k}.$$

Sledi da niz $\{\phi(x_k; \bar{\mu})\}$ skoro sigurno nije ograničen odozdo, ali je to u kontradikciji s pretpostavkom 5.1.

Dakle, nenegativan niz $\{\alpha_k\}$ ne može biti ograničen odozdo pozitivnim brojem, te ostaje da je

$$\lim_{k \in K_1} \alpha_k = 0 \quad (5.30)$$

za neki beskonačan podniz $K_1 \subseteq \mathbb{N} \cap \{\bar{k}, \bar{k} + 1, \dots\}$. Bez gubitka opštosti možemo pretpostaviti da je

$$\alpha_k < 1 \quad \text{za svako } k \in K_1.$$

To znači da za svako $k \in K_1$ postoji dužina koraka $\alpha'_k \leq 1$ za koju je $\alpha_k = \beta \alpha'_k$ i ne važi Armižo uslov, tj. za nju važi

$$\phi_{N_k}(x_k + \alpha'_k d_k; \bar{\mu}) > \phi_{N_k}(x_k; \bar{\mu}) + \eta \alpha'_k g_k^T d_k,$$

odnosno

$$\frac{1}{\alpha'_k} (\phi_{N_k}(x_k + \alpha'_k d_k; \bar{\mu}) - \phi_{N_k}(x_k; \bar{\mu})) > \eta g_k^T d_k.$$

Primenom teoreme o srednjoj vrednosti dobijamo da za svako $k \in K_1$ postoji $t_k \in (0, 1)$ takvo da je

$$\nabla_x^T \phi_{N_k}(x_k + t_k \alpha'_k d_k; \bar{\mu}) d_k - \eta g_k^T d_k > 0. \quad (5.31)$$

Pored toga, ograničenost niza $\{x_k\}$ implicira postojanje $K_2 \subseteq K_1$ i x^* za koje važi

$$\lim_{k \in K_2} x_k = x^*.$$

Kako je još kazneni parametar fiksiran (jednak s $\bar{\mu}$) i $N_k \rightarrow \infty$ (teorema 5.1), lema 5.4 implicira

$$\lim_{k \in K_2} g_k = \lim_{k \in K_2} \nabla_x \phi_{N_k}(x_k; \bar{\mu}) = \nabla_x \phi(x^*; \bar{\mu}) \quad \text{s.s.}$$

Odavde, na osnovu pretpostavke 5.4, odnosno (5.23), sledi da je niz $\{d_k\}_{k \in K_2}$ skoro sigurno ograničen, pa samim tim skoro sigurno postoje $K_3 \subseteq K_2$ i d^* s osobinom

$$\lim_{k \in K_3} d_k = d^*.$$

Zbog osobine da je $\alpha_k = \beta \alpha'_k$, iz (5.30) sledi

$$\lim_{k \in K_3} \alpha'_k = 0.$$

Ta činjenica, zajedno s konvergencijom nizova $\{x_k\}_{k \in K_3}$ i $\{d_k\}_{k \in K_3}$, puštanjem limesa po K_3 u (5.31), na osnovu leme 5.4, daje

$$\nabla_x^T \phi(x^*; \bar{\mu}) d^* - \eta \nabla_x^T \phi(x^*; \bar{\mu}) d^* \geq 0 \quad \text{s.s.}$$

Dalje je

$$\nabla_x^T \phi(x^*; \bar{\mu}) d^* \geq \nabla_x^T \phi(x^*; \bar{\mu}) d^* (1 - \eta) \geq 0 \quad \text{s.s.}$$

jer je $\eta \in (0, 1)$. Međutim, iz (5.28) sledi

$$\nabla_x^T \phi(x^*; \bar{\mu}) d^* \leq -\frac{1}{\bar{\mu}^2} < 0 \quad \text{s.s.,}$$

što daje kontradikciju.

Konačno, s pretpostavkom da je $\{\mu_k\}$ ograničen skoro sigurno dolazimo do kontradikcije, pa možemo da zaključimo da je on skoro sigurno neograničen. Sada činjenica da je niz $\{\mu_k\}_{k \in \mathbb{N}}$ neopadajući povlači tvrđenje teoreme. \square

Sad možemo da formulišemo rezultat o konvergenciji. Primetimo da pod uslovom da (5.7) važi samo konačno mnogo puta, na osnovu prethodne teoreme $\mu_k \rightarrow \infty$ skoro sigurno, pa činjenica da je niz $\{\mu_k\}$ neopadajući implicira da se skoro sigurno generiše beskonačan niz $\{z_t\}$, kroz korak K6 algoritma 5.1. Pokazujemo da u tom slučaju, uz LICQ uslove, svaka tačka nagomilavanja tog niza zadovoljava KKT uslove problema (5.1):

$$\begin{aligned} \nabla_x \mathcal{L}(x, \lambda) = \nabla f(x) + \nabla^T h(x) \lambda &= 0 \\ h(x) &= 0 \end{aligned}$$

za neko $\lambda \in \mathbb{R}^m$, pri čemu je Lagranžova funkcija

$$\mathcal{L}(x, \lambda) = f(x) + \lambda^T h(x).$$

Videćemo da su KKT uslovi zadovoljeni i u slučaju kada (5.7) važi beskonačno mnogo puta.

Teorema 5.3. *Neka su zadovoljene pretpostavke 5.1–5.5 i neka je niz iteracija $\{x_k\}_{k \in \mathbb{N}_0}$ generisan algoritmom 5.1 ograničen. Tada postoji tačka nagomilavanja x^* niza $\{x_k\}$ koja je skoro sigurno stacionarna tačka za θ . Ukoliko je još i LICQ uslov zadovoljen, tada je x^* skoro sigurno KKT tačka originalnog problema (5.1).*

Dokaz. Posmatramo dva slučaja: uslov (5.7) važi beskonačno mnogo puta i uslov (5.7) važi samo konačno mnogo puta.

Prvo pretpostavimo da postoji beskonačan podniz $J_1 \subseteq \mathbb{N}$ takav da za svako $k \in J_1$ važi (5.7), odnosno

$$g_k = 0 \quad \text{i} \quad h_{N_k}(x_k) = 0 \quad \text{za svako } k \in J_1.$$

Sledi da za svako $k \in J_1$ važi

$$0 = g_k = \nabla f(x_k) + 2\mu_k \nabla^T h_{N_k}(x_k) h_{N_k}(x_k) = \nabla f(x_k).$$

Po pretpostavci da je niz $\{x_k\}$ ograničen, postoje podniz $J_2 \subseteq J_1$ i tačka x^* takvi da je

$$\lim_{k \in J_2} x_k = x^*,$$

te zbog neprekidnosti gradijenta funkcije cilja sledi

$$0 = \lim_{k \in J_2} \nabla f(x_k) = \nabla f(x^*).$$

Pored toga, na osnovu teoreme 5.1 niz $\{N_k\}$ teži ka beskonačnosti, pa lema 5.4 implicira

$$0 = \lim_{k \in J_2} h_{N_k}(x_k) = h(x^*) \quad \text{s.s.} \quad (5.32)$$

To znači da je tačka nagomilavanja x^* skoro sigurno dopustiva tačka za problem (5.1) i

$$\nabla_x \mathcal{L}(x^*, 0) = 0 \quad \text{s.s.}, \quad (5.33)$$

ali i

$$\nabla \theta(x^*) = \nabla^T h(x^*) h(x^*) = 0 \quad \text{s.s.}$$

Dakle, dobili smo da je u ovom slučaju x^* skoro sigurno stacionarna tačka za θ , a na osnovu (5.32) i (5.33) i da je skoro sigurno KKT tačka originalnog problema (5.1).

Sada pretpostavimo da uslov (5.7) važi samo konačno mnogo puta. Tada na osnovu teoreme 5.2 sledi

$$\lim_{k \rightarrow \infty} \mu_k = \infty \quad \text{s.s.}, \quad (5.34)$$

te se skoro sigurno generiše beskonačan niz $\{z_t\}_{t \in \mathbb{N}}$, podniz od $\{x_k\}$. Stoga, zbog ograničenosti niza $\{x_k\}$, postoji tačka nagomilavanja x^* niza $\{z_t\}$, a samim tim i niza $\{x_k\}$. Znači, postoji skup indeksa $K_1 \subseteq \mathbb{N}$ s osobinama:

$$\lim_{k \in K_1} x_k = x^*$$

i (po koraku K6 algoritma 5.1)

$$dm_k \leq \frac{\alpha_k}{\mu_k^2} \quad \text{za svako } k \in K_1.$$

Odavde, koristeći definiciju dm_k i posledicu (5.22) pretpostavke 5.4, važi

$$\lambda_{min} \|g_k\|^2 \leq -g_k^T d_k \leq \frac{1}{\mu_k^2} \quad \text{za svako } k \in K_1, \quad (5.35)$$

sa $\lambda_{min} > 0$. Pošto (5.34) implicira da i podniz $\{\mu_k\}_{k \in K_1}$ teži ka beskonačnosti, puštajući limes po K_1 u (5.35) dobijamo

$$\lim_{k \in K_1} \|g_k\| = 0, \quad (5.36)$$

odnosno

$$\lim_{k \in K_1} g_k = 0. \quad (5.37)$$

Pored toga, kako je za svako k

$$g_k = \nabla f(x_k) + \mu_k \nabla \theta_{N_k}(x_k),$$

te

$$\nabla \theta_{N_k}(x_k) = \frac{1}{\mu_k} (g_k - \nabla f(x_k)),$$

sledi

$$\|\nabla \theta_{N_k}(x_k)\| \leq \frac{1}{\mu_k} (\|g_k\| + \|\nabla f(x_k)\|).$$

Dalje, (5.36), neprekidnost norme i funkcije ∇f i činjenica da i niz $\{\mu_k\}_{k \in K_1}$ teži ka beskonačnosti daju

$$\lim_{k \in K_1} \|\nabla \theta_{N_k}(x_k)\| = 0.$$

Kako još i $N_k \rightarrow \infty$ (po teoremi 5.1), lema 5.4 implicira

$$\lim_{k \in K_1} \nabla \theta_{N_k}(x_k) = \nabla \theta(x^*) \quad \text{s.s.,}$$

pa sledi

$$\nabla \theta(x^*) = 0 \quad \text{s.s.} \quad (5.38)$$

Dakle, dokazali smo da je svaka tačka nagomilavanja niza $\{z_t\}$ skoro sigurno stacionarna tačka funkcije θ .

Ako sada važe još i LICQ uslovi, tada matrica $\nabla h(x^*)$ ima pun rang. Pošto je

$$\nabla \theta(x^*) = 2\nabla^T h(x^*)h(x^*),$$

iz (5.38) sledi da je $h(x^*) = 0$ skoro sigurno, te je x^* skoro sigurno dopustiva tačka problema (5.1). Da bismo odredili Lagranžov množitelj za koji x^* zadovoljava KKT uslove, definišimo

$$\lambda_k := 2\mu_k h_{N_k}(x_k).$$

Tada je

$$g_k = \nabla f(x_k) + \nabla^T h_{N_k}(x_k) \lambda_k,$$

odnosno

$$\nabla^T h_{N_k}(x_k) \lambda_k = g_k - \nabla f(x_k). \quad (5.39)$$

Primenivši sada lemu 5.4 na gradijent funkcije ograničenja, dobijamo

$$\lim_{k \in K_1} \nabla h_{N_k}(x_k) = \nabla h(x^*) \quad \text{s.s.}, \quad (5.40)$$

pa sledi da i matrice $\nabla h_{N_k}(x_k)$, za svako dovoljno veliko $k \in K_1$, skoro sigurno imaju pun rang. Pomnoživši s leve strane izraz (5.39) matricom $\nabla h_{N_k}(x_k)$ i izrazivši λ_k , dobijamo da je za svako dovoljno veliko $k \in K_1$

$$\lambda_k = (\nabla h_{N_k}(x_k) \nabla^T h_{N_k}(x_k))^{-1} \nabla h_{N_k}(x_k) (g_k - \nabla f(x_k)).$$

Uzmimo sada limes po K_1 . Na osnovu (5.37) i (5.40) sledi

$$\lim_{k \in K_1} \lambda_k = -(\nabla h(x^*) \nabla^T h(x^*))^{-1} \nabla h(x^*) \nabla f(x^*) \quad \text{s.s.} \quad (5.41)$$

Označavanjem desne strane izraza (5.41) sa λ^* , dobijamo

$$0 = \lim_{k \in K_1} g_k = \nabla f(x^*) + \nabla^T h(x^*) \lambda^* \quad \text{s.s.},$$

što sa skoro sigurnom dopustivošću tačke x^* implicira da je x^* skoro sigurno KKT tačka problema (5.1), što smo i hteli da dokažemo. \square

Sprovedena analiza konvergencije pokazuje primenljivost prikazanog algoritma za rešavanje problema (5.1). Numeričko testiranje, navedeno u poglavlju 6, potvrđuje uspešnost tog algoritma u uštedi pri broju izračunavanja funkcije, poređenjem adaptivne šeme za određivanje veličine uzorka s unapred definisanim heurističkom šemom.

Glava 6

Numerički rezultati

U ovoj glavi ćemo prikazati numeričke rezultate dobijene primenom predstavljenih kaznenih postupaka s promenljivom veličinom uzorka za rešavanje problema stohastičkog programiranja s ograničenjima tipa jednakosti u obliku matematičkog očekivanja (problema SP). Cilj nam je bio da, pored konvergencije navedenih algoritama ka odgovarajućim rešenjima, postignemo da kreirani postupci budu efikasniji (u pogledu manjeg broja izračunavanja funkcija) od uporedivih postupaka. Rezultati numeričkog poređenja potvrđuju dostizanje željenog cilja.

Oba algoritma smo implementirali na istoj test grupi problema. Probleme i detalje o implementaciji navodimo u odeljku 6.1, gde još ilustrujemo kretanja veličine uzorka, njegove donje granice i kaznenog parametra kreiranih posmatranim algoritmima. Potom ćemo u sekciji 6.2 prikazati rezultate poređenja kaznenog VSS postupka za rešavanje SAA reformulacije problema SP (algoritam 4.1) sa SAA metodom i heurističkom procedurom. Poređenje kaznenog VSS postupka za rešavanje samog problema stohastičkog programiranja (algoritam 5.1) i metode zasnovane na heurističkom ažuriranju uzorka ćemo ilustrovati u delu 6.3. Pored toga, u datom odeljku ćemo analizirati uticaj uslova u kojem figuriše mera nedopustivosti, uz uslov s merom optimalnosti, na porast veličine uzorka u kaznenom VSS postupku navedenom u prethodnom poglavlju.

6.1 Implementacija algoritama

Algoritme prikazane u prethodnim poglavljima smo implementirali na test problemima s ograničenjima tipa jednakosti iz Hock i Schittkowski [24]. Kako je za konvergenciju oba prikazana algoritma potrebna pretpostavka da je funkcija cilja odozdo ograničena na određenom skupu, uzeli smo probleme koji

zadovoljavaju tu pretpostavku na \mathbb{R}^n . S obzirom na to da je među tim problemima bilo i problema s beskonačno mnogo rešenja, radi lakšeg poređenja rešenja, testirali smo probleme s jedinstvenim rešenjem: 6, 27, 28, 42, 46-52, 61, 77 i 79. Posmatrani problemi su deterministički, te smo slučajnost uključili u ograničenja tako što smo funkcije ograničenja $c(x)$ iz [24] transformisali u

$$H(x, \xi) = c(\xi x),$$

gde slučajna promenljiva ξ ima Normalnu raspodelu $\mathcal{N}(1, 1)$. Testiranje smo sprovedli implementacijom algoritama u *Matlab* sa po 10 različitih uzoraka za svaki problem. Uzorci su generisani ugrađenom funkcijom *randn*.

Radi uporedivosti, algoritmi koji se porede imaju istu strukturu, jedino se razlikuju u načinu ažuriranja veličine uzorka. Pored toga, za sve njih smo koristili i iste parametre i uzorke (tačnije, uzorci iste veličine su jednaki). Za početne iterativne tačke x_0 smo uzeli tačke iz [24]. Kazneni parametar smo uvećavali s faktorom $\rho = 1.5$, a za početni kazneni parametar smo koristili $\mu_0 = 1$. Parametri za linijsko pretraživanje su standardni, $\beta = 0.5$ i $\eta = 10^{-4}$. Za pravac pretraživanja smo upotreбили BFGS pravac

$$d_k = -B_k g_k$$

uzimajući jediničnu matricu za B_0 . Matrica B_k je aproksimacija inverznog Hესijana $(\nabla_{xx}^2 \phi_{N_k}(x_k; \mu_k))^{-1}$, te je pravilo ažuriranja

$$B_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k} \right) B_k \left(I - \frac{y_k s_k^T}{y_k^T s_k} \right) + \frac{s_k s_k^T}{y_k^T s_k}, \quad (6.1)$$

sa

$$s_k = x_{k+1} - x_k$$

i

$$y_k = \nabla_x \phi_{N_{k+1}}(x_{k+1}; \mu_k) - \nabla_x \phi_{N_k}(x_k; \mu_k). \quad (6.2)$$

Da bismo obezbedili da pravac pretraživanja bude opadajući, primenili smo meru zaštite za očuvavanje pozitivne definitnosti BFGS matrice. Pravilo (6.1) je upotrebljeno ukoliko je važio uslov $y_k^T s_k > b$, sa $b > 0$, a u suprotnom je BFGS matrica ostala nepromenjena, tj.

$$B_{k+1} = B_k \quad \text{za} \quad y_k^T s_k \leq b.$$

Pri implementaciji algoritma 4.1, kao i algoritama s kojim smo ga poredili, koristili smo $b = 10^{-10}$. Prilikom testiranja algoritma 5.1 ta vrednost nije bila

dovoljna da u svakoj iteraciji obezbedi pozitivnu definitnost BFGS matrice, te smo algoritam 5.1 i algoritam s kojim smo ga poredili računali sa $b = 10^{-6}$.

Pretpostavljamo da je izračunavanje vrednosti funkcije ograničenja i gradijenata njenih komponenti dominantan trošak u iterativnom postupku, pa je zato za nas relevantna efikasnost postupaka po pitanju broja izračunavanja vrednosti funkcija, a ne broja iteracija ili utrošenog vremena, kao što je naznačeno u Spall [54]. Stoga postupke poredimo po broju izračunavanja vrednosti funkcija ograničenja H i njihovih Jakobijevih matrica ∇H , označenog sa FEV. Izračunata vrednost $H(x_k, \xi_i)$ vredi jedan FEV, dok vrednost funkcije $\nabla H(x_k, \xi_i)$ brojimo kao n FEV-a.*

Primetimo da se u izrazu (6.2) gradijent kaznene funkcije u novoj iterativnoj tački x_{k+1} određuje uzorkom veličine N_{k+1} , a ne N_k . Razlog za taj izbor je to što se $\nabla_x \phi_{N_k}(x_{k+1}; \mu_k)$ računa na osnovu vrednosti funkcija

$$H(x_{k+1}, \xi_1), \dots, H(x_{k+1}, \xi_{N_k})$$

i

$$\nabla H(x_{k+1}, \xi_1), \dots, \nabla H(x_{k+1}, \xi_{N_k})$$

koje u slučaju kada je $N_{k+1} < N_k$ sadrže vrednosti

$$\nabla H(x_{k+1}, \xi_{N_{k+1}+1}), \dots, \nabla H(x_{k+1}, \xi_{N_k})$$

koje se ne koriste u narednoj iteraciji, a za njihovo izračunavanje je potrošeno $(N_k - N_{k+1})n$ FEV-a. S druge strane, za $\nabla_x \phi_{N_{k+1}}(x_{k+1}; \mu_k)$ se računaju vrednosti

$$H(x_{k+1}, \xi_1), \dots, H(x_{k+1}, \xi_{N_{k+1}})$$

i

$$\nabla H(x_{k+1}, \xi_1), \dots, \nabla H(x_{k+1}, \xi_{N_{k+1}})$$

koje se koriste u izračunavanju vrednosti $\nabla_x \phi_{N_{k+1}}(x_{k+1}; \mu_{k+1})$ potrebne u $(k+1)$ -oj iteraciji i time ne koštaju dodatne FEV.

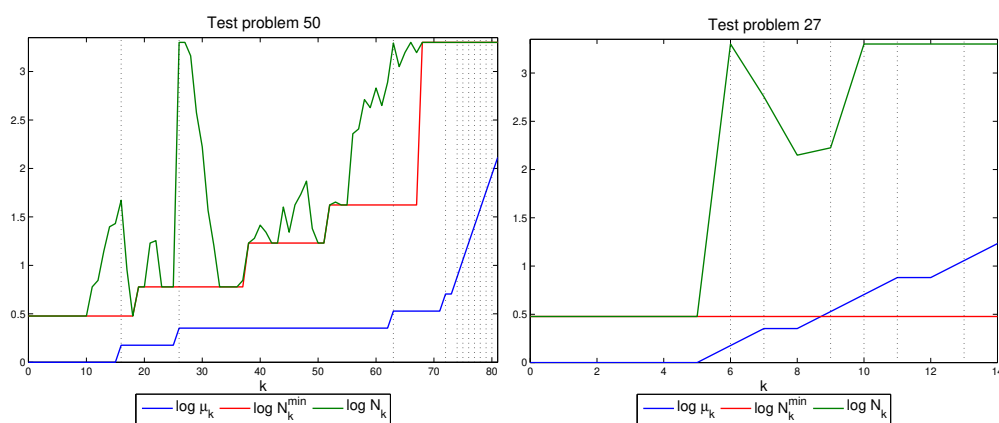
Za oba prikazana algoritma i za sve postupke s kojima je izvršeno njihovo poređenje problem 77 je imao neuspešan ishod i to sa svim uzorcima, tako da smo taj problem isključili iz razmatranja. Dakle, rezultati prikazani u nastavku se odnose na 13 test problema: 6, 27, 28, 42, 46-52, 61 i 79.

*U brojanje FEV-a nismo uključili m , broj komponenti funkcije H , što ne utiče na ishod rezultata poređenja algoritama. Razlog tome je to što bi se uključivanjem m -a dobile m puta veće vrednosti, a način vrednovanja efikasnosti algoritama nije osetljiv na tu promenu.

6.1.1 Ponašanje veličine uzorka, njegove donje granice i kaznenog parametra

Pre nego što pokažemo rezultate poređenja kreiranih kaznenih VSS postupaka ilustrovaćemo ponašanja niza veličina uzoraka $\{N_k\}_{k \in \mathbb{N}_0}$, niza donjih granica veličina uzoraka $\{N_k^{min}\}_{k \in \mathbb{N}_0}$ i niza kaznenih parametara $\{\mu_k\}_{k \in \mathbb{N}_0}$ koji su njima generisani.

Na slici 6.1 je prikazano kretanje veličine uzorka, njegove donje granice i kaznenog parametra dobijenih algoritmom 4.1 primenjenog na test probleme 50 i 27. Može se reći da su to slike tipičnog ponašanja nizova $\{N_k\}$, $\{N_k^{min}\}$ i $\{\mu_k\}$ i za preostale testirane probleme. Grafici za preostale testirane probleme (dobijeni upotrebom istih uzoraka) mogu se videti u narednom poglavlju.

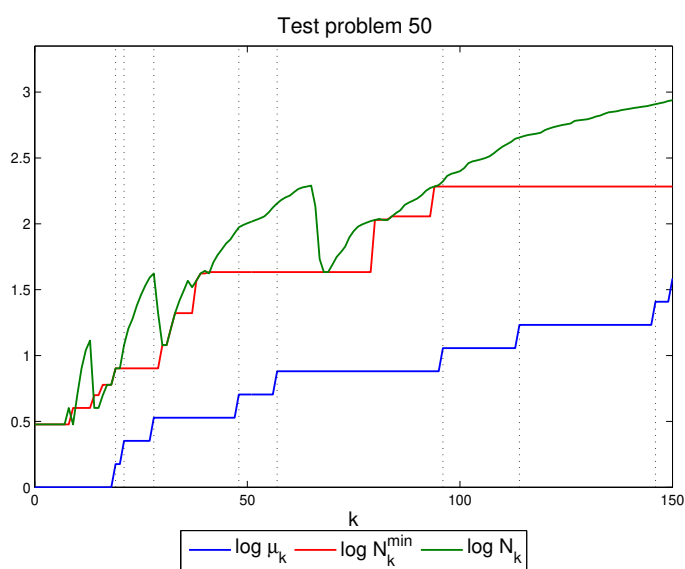


Slika 6.1: Tipično kretanje veličine uzorka N_k , njegove donje granice N_k^{min} i kaznenog parametra μ_k generisanih algoritmom 4.1

Radi boljeg prikaza skokova u vrednosti veličine uzorka, upotrebili smo logaritamsku skalu na y-osi. Možemo da vidimo kako veličina uzorka varira iz iteracije u iteraciju, da raste i opada, s tim da nakon nekog broja iteracija dostiže maksimalnu vrednost i ne menja se više (kao što je i pokazano u lemi 4.2). Vidimo i da su nizovi $\{N_k^{min}\}$ i $\{\mu_k\}$ neopadajući. Za niz $\{N_k^{min}\}$ imamo dva slučaja: praktično ne utiče na N_k (problem 27) i odezbeđuje da N_k postane N_{max} (problem 50), što potvrđuje da nema samo teoretsku ulogu. Istačkane vertikalne linije su pomoćne linije koje označavaju iteracije u kojima se prvi put računa s uvećanim kaznenim parametrom, odnosno to su k -te iteracije s osobinom da je $\mu_{k-1} < \mu_k$. Primitimo da je u ponekim iteracijama s tom osobinom uvećan kazneni parametar implicirao smanjenje veličine uzorka (pro-

blem 50: $k \in \{16, 62\}$), ali to nije pravilo što potvrđuju ostale iteracije s kaznenim parametrom povećanim u prethodnoj iteraciji. Dakle, ne možemo reći da postoji pravilo po kojem povećanje kaznenog parametra utiče na promenu veličine uzorka, čak ni u prvih nekoliko iteracija kada je aproksimacija kaznenom funkcijom obično manje tačna i promena kaznenog parametra može dovesti do većih razlika u vrednosti kaznene funkcije, a samim tim i povećavanju mere optimalnosti.

Što se tiče nizova $\{N_k\}$, $\{N_k^{min}\}$ i $\{\mu_k\}$ generisanih algoritmom 5.1 analogno možemo reći da se njihova tipična kretanja mogu opisati rezultatima dobijenim za test problem 50 pokazanim na slici 6.2. Grafički prikazi za ostale testirane probleme se mogu naći u dodatku.



Slika 6.2: Tipično ponašanje niza veličina uzorka $\{N_k\}$, niza donjih granica veličina uzorka $\{N_k^{min}\}$ i niza kaznenih parametara $\{\mu_k\}$ generisanih algoritmom 5.1

Možemo da vidimo kako i po ovom postupku kazneni parametar i donja granica veličine uzorka ne opadaju, a veličina uzorka varira, raste i opada, iz iteracije u iteraciju, s tim da se sada ne zaustavlja na nekoj vrednosti (ova osobina je formulisana u lemi 5.1). Po ovom postupku niz $\{N_k^{min}\}$ ima ulogu da obezbedi da niz $\{N_k\}$ teži u beskonačno (teorema 5.1). Možemo videti da u nekim iteracijama ($k \in \{9, 14, 15, 16, 41, 82, 83\}$) važi da je $N_k^{min} > N_k$, kao što je i naznačeno u analizi nakon algoritama 5.1 i 5.2 da se to može dogoditi.

Ta analiza je pokazatelj da je i po ovom postupku niz $\{N_k\}$ ograničen odozdo, ali sada s nizom $\{N_k^{min} - 1\}$. Za odnos kaznenog parametra i veličine uzorka možemo zaključiti isto što smo i za algoritam 4.1.

6.2 Kazneni VSS postupak za rešavanje SAA problema

Pri testiranju kaznenog postupka s promenljivom veličinom uzorka za minimizaciju SAA reformulacije problema SP rešavali smo SAA problem

$$\min f(x) \quad \text{tako da je} \quad h_{N_{max}}(x) := \frac{1}{N_{max}} \sum_{i=1}^{N_{max}} H(x, \xi_i) = 0 \quad (6.3)$$

za koji je pun uzorak $\xi_1, \xi_2, \dots, \xi_{N_{max}}$ sa $N_{max} = 2000$ generisan na samom početku iterativnog postupka. To znači da smo svaki od navedenih 13 problema testirali na 10 uzoraka veličine 2000.

Algoritam 4.1 (VSS) smo uporedili sa SAA metodom (SAA) i heurističkom procedurom (HEUR). Po SAA metodi je tokom celog iterativnog postupka računato s punim uzorkom, pa je

$$N_k = N_{max} \quad \text{za svako } k.$$

Kod heurističke procedure je unapred zadata šema za povećavanje veličine uzorka do maksimalne. Na osnovu radova [18,30,40], uzeli smo pravilo

$$N_{k+1} = \lceil \min\{1.1N_k, N_{max}\} \rceil, \quad k = 0, 1, 2, \dots,$$

s tim da je $N_0 = 3$. Za VSS je uzeto da je minimalna veličina uzorka $N_{min} = 3$ i time je $N_0 = 3$ kao i za HEUR. Ostalo je još da navedemo da je u algoritmu 4.1 parametar $v_1 = 1/\sqrt{N_{max}}$ i da smo za meru preciznosti $\varepsilon_\delta^{N_k}$ koristili

$$\varepsilon_\delta^{N_k}(x_k) := \frac{1.96 \hat{\sigma}_{N_k}(x_k)}{\sqrt{N_k}},$$

s uzoračkom varijansom

$$\hat{\sigma}_{N_k}^2(x_k) := \frac{1}{N_k - 1} \sum_{i=1}^{N_k} \|H(x_k, \xi_i) - h_{N_k}(x_k)\|^2. \quad (6.4)$$

Algoritam 4.1 je kreiran tako da generiše beskonačno mnogo iteracija, ali nam je praktično bio potreban kriterijum zaustavljanja. Kako nam je cilj bio

da dođemo do rešenja SAA problema (6.3), za kriterijum zaustavljanja smo koristili uslov

$$\left\| (\nabla_x \phi_{N_{max}}(x_k; \mu_k), h_{N_{max}}(x_k))^T \right\| \leq 10^{-1}.$$

To znači da se algoritam uspešno završio jedino kada je $N_k = N_{max}$ i x_k je približno KKT tačka SAA problema (6.3). Ukoliko se tokom testiranja došlo do 10^8 FEV-a, ali ne i do tačke koja zadovoljava izlazni kriterijum, smatramo da se algoritam neuspešno završio.

Rezultate poređenja smo prikazali pomoću takozvanog profila performansi (*performance profile*) uvedenog u radu Dolan i Moré [14]. Suština tog prikaza je da se za svaki postupak koji se poredi za svaki testiran problem odredi količnik performansi (*performance ratio*) koji predstavlja odnos mere performansi posmatranog postupka prema najboljem rezultatu poređenih postupaka na posmatranom problemu. Kumulativna funkcija raspodele količnika performansi posmatranog postupka predstavlja profil performansi tog postupka. Da preciziramo, imamo skup \mathcal{P} od 130 test problema (13 problema sa po 10 različitih uzoraka), skup \mathcal{S} od tri postupka: VSS, SAA i HEUR i FEV kao meru performanse postupka na konkretnom test problemu (FEV_p^s je broj FEV-a za test problem p testiran postupkom s). Količnik performansi postupka s na problemu p je

$$r_p^s := \frac{FEV_p^s}{\min\{FEV_p^s : s \in \mathcal{S}\}}.$$

U slučaju da se postupak neuspešno završio, uzeli smo da je količnik performansi znatno veći broj od ostalih količnika, $r_p^s = 10^{15}$.[†] Empirijska verovatnoća postupka s da je njegov količnik performansi ne veći od realnog koeficijenta τ je

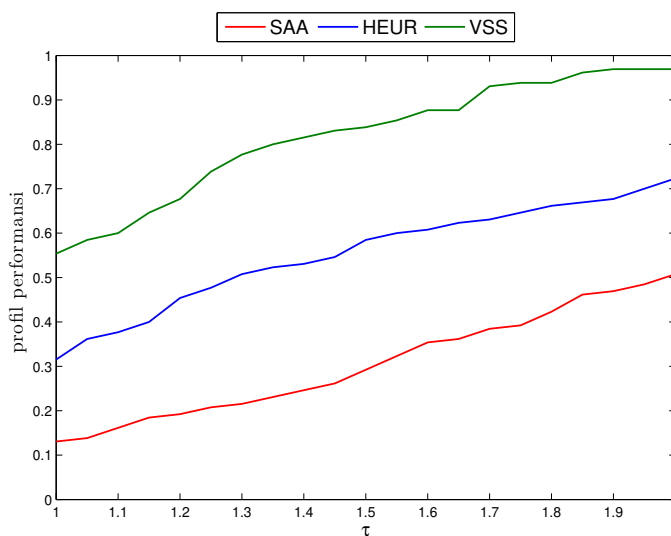
$$p^s(\tau) := P_E \{r_p^s \leq \tau : p \in \mathcal{P}\},$$

tako da funkcija p^s predstavlja profil performansi postupka s na skupu testiranih problema \mathcal{P} . Primetimo da $p^s(1)$ predstavlja verovatnoću da je postupak s bio najefikasniji među postupcima iz skupa \mathcal{S} .

Rezultati poređenja algoritma 4.1, heurističke procedure i SAA metode su predstavljeni na slici 6.3. Kao što vidimo, verovatnoće pobeđivanja (rezultati za $\tau = 1$) postupaka VSS, HEUR i SAA su redom 0.55, 0.32 i 0.13. Razlike u verovatnoćama kada su mere performansi postupaka dvostruko veće od mere performanse najboljeg postupka su otprilike istog reda – verovatnoće

[†]U radu Dolan i Moré [14] je pokazano da izbor parametra koji se koristi za neuspele postupke ne utiče na procenu performansi.

za $\tau = 2$ su 0.97, 0.72 i 0.51 redom za VSS, HEUR i SAA. Otuda možemo da kažemo da su razlike u broju izračunavanja vrednosti funkcija upoređenim postupcima značajne. Pored toga, možemo da zaključimo da postupci s promenljivom veličinom uzorka daju značajno bolje rezultate od SAA metode, te se upotrebom manjih uzoraka kada smo udaljeni od rešenja značajno štedi na broju izračunavanja vrednosti funkcija. Nadalje, krive profila performansi pokazuju znatno bolju efikasnost algoritma 4.1 u odnosu na heurističku šemu, odnosno da je adaptivni postupak u kojem se uzorak smanji kada se udaljimo od rešenja efikasniji od postupka sa unapred definisanim pravilom povećavanja veličine uzorka. Moramo istaći da je moguće naći heurističku šemu koja će dati bolje rezultate od adaptivne šeme, ali je za to potrebno imati informacije o posmatranom problemu ili sprovesti preliminarna testiranja radi utvrđivanja najpogodnijeg stepena povećavanja uzorka. Još bismo napomenuli da sprovedeni zaključci važe na posmatranom skupu problema, ali smatramo da je taj skup, sačinjen od problema s linearnim ograničenjima i problema s nelinearnim ograničenjima, relevantan.



Slika 6.3: Profili performansi kaznenog VSS postupka (algoritam 4.1), heurističke procedure i SAA metode (p^{VSS} , p^{HEUR} i p^{SAA})

6.3 Kazneni VSS postupak za rešavanje problema stohastičkog programiranja

Kazneni postupak s promenljivom veličinom uzorka formulisan u algoritmu 5.1 je kreiran tako da generiše beskonačan niz iteracija $\{x_k\}_{k \in \mathbb{N}_0}$ koji nas skoro sigurno dovodi do KKT tačke problema SP

$$\min f(x) \quad \text{tako da je} \quad h(x) := E[H(x, \xi)] = 0. \quad (6.5)$$

Niz veličina uzoraka $\{N_k\}_{k \in \mathbb{N}_0}$ koji se generiše njime, pod datim uslovima, teži ka beskonačnosti (teorema 5.1). Stoga je ovaj postupak neuporediv sa SAA metodom, ali je uporediv s heurističkim postupkom koji koristi beskonačan niz veličina uzoraka. Tako kazneni VSS definisan algoritmom 5.1 (VSS) poredimo s heurističkom šemom (HEUR) u kojoj se veličina uzorka menja, slično kao u prethodnom testiranju, po pravilu

$$N_{k+1} = \lceil \min\{1.1N_k\} \rceil, \quad k = 0, 1, 2, \dots$$

Za početnu veličinu uzorka smo uzeli $N_0 = 3$ za oba postupka.

Preostalo je još da navedemo funkcije e i γ za VSS postupak. Za meru preciznosti SAA aproksimacije u k -toj iteraciji algoritma 5.1 upotreбили smo meru intervala poverenja ocene $h_{N_k}(x_k)$ jer uzima u obzir informacije o tački x_k i korišćena je u mnogim radovima, [5,29–32,35,43]. Dakle, mera preciznosti je definisana funkcijom

$$e(x_k; N_k) = \frac{1.96\hat{\sigma}_{N_k}(x_k)}{\sqrt{N_k}}, \quad (6.6)$$

pri čemu je $\hat{\sigma}_N^2(x)$ uzoračka varijansa računata sa (6.4). Udeo mere preciznosti pri definisanju zadovoljavajuće mere opadanja mere dopustivosti određen je funkcijom

$$\gamma(N_k) = \exp(-1/N_k),$$

koja zadovoljava pretpostavku 5.3. Mera preciznosti (6.6) ne zadovoljava pretpostavku 5.2 pošto generalno nije ograničena odozdo pozitivnim brojem, ali to nije pokvarilo očekivane rezultate.

U ovom poređenju, određivanje adekvatnog kriterijuma za zaustavljanje algoritama predstavlja izazov pošto je pitanje za koju veličinu uzorka, odnosno meru preciznosti, imamo dovoljno precizno rešenje. S obzirom na to da je za probleme s linearnim ograničenjima problem SP ekvivalentan problemu iz Hock i Schittkowski [24] iz kojeg je formiran, za te probleme imamo optimalna

rešenja u [24]. Za probleme s nelinearnim ograničenjima smo odredili analitički izraz za matematičko očekivanje u funkciji ograničenja, te smo problem s tim izrazom u ograničenjima približno rešili primenom ugrađene naredbe *fmincon* u *Matlab*-u. Računali smo s velikom preciznošću, pa smo dobili rešenja za koja su KKT uslovi zadovoljeni s preciznošću najviše 10^{-5} . Dakle, za testirane probleme imamo optimalna ili skoro optimalna rešenja. Označimo ih sa x^* . Ta rešenja smo koristili za merenje efikasnosti postupaka, posmatrajući iterativne tačke dobijene do zaustavljanja algoritama zbog dostizanja 10^6 FEV-a.

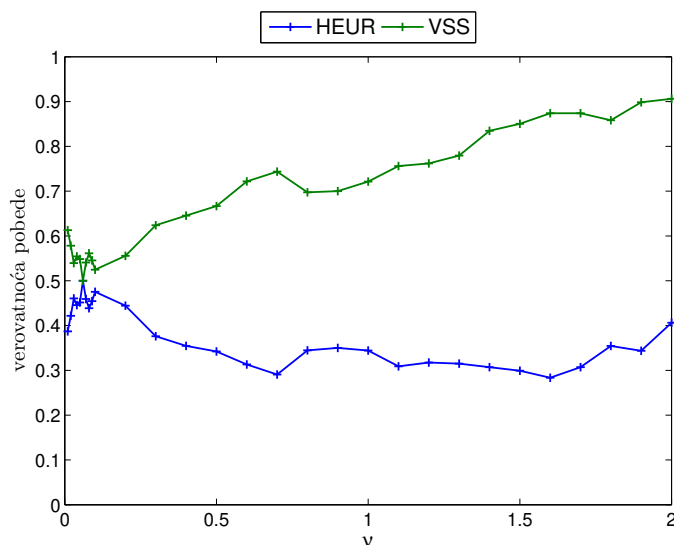
Razmatrali smo različite nivoe preciznosti v aproksimacije rešenja x^* i utvrđivali koji postupak je pobedio, odnosno s manjim brojem izračunavanja vrednosti funkcija došao do rešenja odgovarajuće tačnosti. Ukoliko im je bio potreban isti broj izračunavanja, smatramo da su oba postupka pobedila. Tako smo za postupke VSS i HEUR za svaki testirani problem (ukupno 130 problema, tj. 13 test problema sa po 10 različitih uzoraka) utvrdili koliki mu je broj FEV-a, FEV_k^{VSS} i FEV_l^{HEUR} , bio potreban da dostigne v -optimalno rešenje, x_k^{VSS} i x_l^{HEUR} , za različite vrednosti v . Time smo došli do empirijske verovatnoće da VSS postupak nije lošiji od HEUR postupka, odnosno do verovatnoće pobeđe VSS postupka nad postupkom HEUR

$$p_v^{\text{VSS}} := P_E \left\{ \min_k \{ FEV_k^{\text{VSS}} : \|x_k^{\text{VSS}} - x^*\| \leq v \} \leq \min_l \{ FEV_l^{\text{HEUR}} : \|x_l^{\text{HEUR}} - x^*\| \leq v \} \right\},$$

kao i empirijske verovatnoće pobeđe HEUR postupka

$$p_v^{\text{HEUR}} := P_E \left\{ \min_l \{ FEV_l^{\text{HEUR}} : \|x_l^{\text{HEUR}} - x^*\| \leq v \} \leq \min_k \{ FEV_k^{\text{VSS}} : \|x_k^{\text{VSS}} - x^*\| \leq v \} \right\}.$$

Verovatnoće pobeđa posmatranih postupaka, razmatrajući nivoe preciznosti $v \in (0, 2]$, su prikazane na slici 6.4. Možemo da vidimo da je predstavljeni kazneni VSS postupak efikasniji od heurističke procedure. Šta više, kad su u pitanju manje precizne aproksimacije rešenja problema SP, tada je prednost još izraženija. Dakle, možemo da zaključimo da posmatrani skup testiranih problema ukazuje na prednost korigovanja veličine uzorka na osnovu tekućih informacija u iterativnom postupku u odnosu na određivanje veličine uzorka po unapred određenoj šemi.



Slika 6.4: Empirijska verovatnoća pobeđe p_v^{VSS} i p_v^{HEUR} za različite nivoe tačnosti v

6.3.1 Analiza rasta veličine uzorka

Sada ćemo ilustrovati kako uslov

$$\|h_N(x_k)\| < \frac{N_k}{N(N - N_k)} e(x_k; N). \quad (6.7)$$

iz koraka K1 3) algoritma 5.2, odnosno količnik $N_k / (N(N - N_k))$ u njemu, regulišu stepen rasta veličine uzorka sa N_k na N_{k+1} i kako utiču na broj izračunavanja funkcija, tj. na FEV. Razmotrili smo uticaj količnika:

$$\frac{N_k}{N}, \quad \frac{N_k}{N^2} \quad \text{i} \quad \frac{N_k}{N(N - N_k)}.$$

Pored toga, ispitali smo i ishod algoritma 5.2 bez uslova (6.7). Tako smo napravili prikaz niza veličina uzoraka i vrednosti FEV-a po iteracijama dobijene algoritmom 5.1 koristeći algoritam 5.2 u kojem je korak K1 3) definisan sa:

Ako je $dm_k < e(x_k; N_k)$, tada je $N = N_k + 1$.

Dokle god je

test

stavljati da je $N = N + 1$.

Uzeti da je $N_{k+1} = N$,

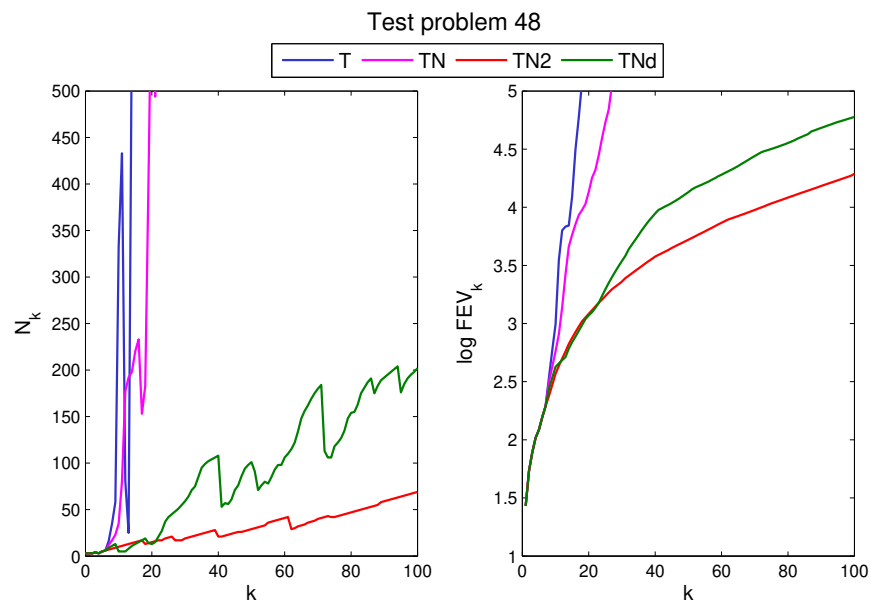
pri čemu je *test* definisan sa:

$$\mathbf{T} \quad dm_k < \frac{N_k}{N} e(x_k; N)$$

$$\mathbf{TN} \quad dm_k < \frac{N_k}{N} e(x_k; N) \quad \text{i} \quad \|h_N(x_k)\| < \frac{N_k}{N} e(x_k; N)$$

$$\mathbf{TN2} \quad dm_k < \frac{N_k}{N} e(x_k; N) \quad \text{i} \quad \|h_N(x_k)\| < \frac{N_k}{N^2} e(x_k; N)$$

$$\mathbf{TNd} \quad dm_k < \frac{N_k}{N} e(x_k; N) \quad \text{i} \quad \|h_N(x_k)\| < \frac{N_k}{N(N - N_k)} e(x_k; N).$$

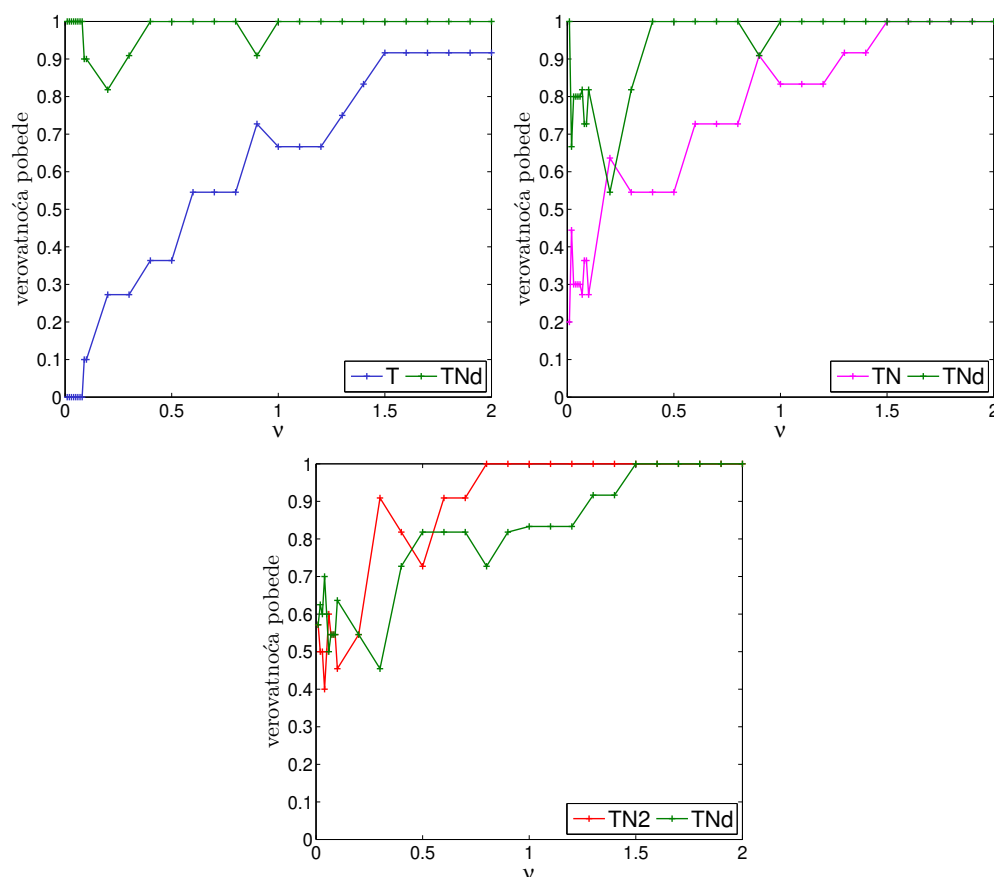


Slika 6.5: Rast veličine uzorka i broja izračunavanja funkcija u zavisnosti od uslova *test* u koraku K1 3) algoritma 5.2

Na slici 6.5 su ilustrovane vrednosti veličina uzorka N_k i odgovarajućih FEV-a za prvih 100 iteracija test problema 48 (za jedan uzorak). Možemo da vidimo da verzija koja ne uzima u obzir promene u funkciji ograničenja s porastom veličine uzorka (T) veoma brzo dostiže veliki broj FEV-a. Nadalje, količnik u uslovu (6.7) ima značajan uticaj. Ukoliko količnik nije dovoljno restriktivan, kao N_k/N , veličina uzorka će takođe brzo porasti i time će dovesti

do velikih troškova po pitanju broja izračunavanja vrednosti funkcija. Količnik N_k/N^2 ima značajan uticaj na uštedu broja FEV-a, ali je prejak jer dovodi do povećavanja veličine uzorka samo za 1 ili 2. Količnik $N_k/(N(N - N_k))$ nije previše restriktivan niti previše slab, tako da se njime postiže da veličina uzorka slobodno varira i ne postaje prevelika.

Slični rezultati su dobijeni i za preostale probleme računane s istim uzorkom. Njihovi grafici se mogu videti u dodatku.



Slika 6.6: Empirijska verovatnoća pobeđe verzije TNd u odnosu na verziju T, TN, odnosno TN2, upoređena s empirijskom verovatnoćom pobeđivanja navedenih verzija nad verzijom TNd, na skupu od 13 testiranih problema s istim uzorkom, za različite nivoe tačnosti v

Rezultati prikazani na slici 6.6 pokazuju da su verzijama T i TN dobijene veličine uzorka nepotrebno velike jer taj porast dovodi te postupke ka

traženoj aproksimaciji rešenja problema SP s velikim brojem izračunavanja vrednosti funkcija. Naime, analognim poređenjem kao što su upoređeni postupci VSS i HEUR u uvodnom delu odeljka 6.3, upoređena je verzija TNd (što je u stvari algoritam 5.2) sa preostale tri verzije, ali na skupu od 13 testiranih problema s istim uzorkom za koji su dobijeni rezultati na slici 6.5. Tako je empirijska verovatnoća pobeđe verzije $r \in \{T, TN, TN2, TNd\}$ nad verzijom $s \in \{T, TN, TN2, TNd\}$ definisana sa

$$p_v^{r-s} := P_E \left\{ \min_k \{FEV_k^r : \|x_k^r - x^*\| \leq v\} \leq \min_l \{FEV_l^s : \|x_l^s - x^*\| \leq v\} \right\}.$$

Grafici empirijskih verovatnoća p_v^{TNd-T} i p_v^{T-TNd} , kao i p_v^{TNd-TN} i p_v^{TN-TNd} , pokazuju da je, razmatranjem rešenja nivoa tačnosti $v \in (0, 2]$, verzija TNd znatno efikasnija od verzija T i TN. Primetimo da poređenje verzija T i TNd posebno ističe značaj uslova (6.7) kada se traže aproksimacije rešenja velike preciznosti.

Na slici 6.6 je prikazan i odnos verzija TNd i TN2. Možemo reći da su za polovinu test primera obe verzije podjednako efikasne. Kada su u pitanju manje precizne aproksimacije rešenja, tada postoji mala prednost verzije s količnikom N_k/N^2 , ali se ta prednost gubi kada se traže preciznije aproksimacije rešenja problema SP. S obzirom na to da taj količnik može izuzetno jako da ograničava povećanje veličine uzorka, verzija TN2 smanjuje prednost adaptivnog ažuriranja veličine uzorka.

Dakle, prethodna analiza sugerise da je po pitanju uštede u broju izračunavanja vrednosti funkcija značajno da se prati mera nedopustivosti dok se povećava uzorak radi utvrđivanja kada više nema potrebe za daljim proširivanjem uzorka.

Glava 7

Zaključak

U disertaciji su posmatrani problemi minimizacije s ograničenjima tipa jednakosti koja su u obliku matematičkog očekivanja. Prikazana su dva postupka za njihovo rešavanje, sa dva različita pristupa. U oba slučaja smo primenili stohastičku optimizaciju, radeći sa uzoračkim očekivanjem kao aproksimacijom matematičkog očekivanja, i iskoristili prednosti postupaka s promenljivom veličinom uzoraka zasnovanih na adaptivnom ažuriranju veličine uzorka. Adaptivno ažuriranje podrazumeva određivanje veličine uzorka na osnovu informacija o greškama nastalim zbog aproksimacije matematičkog očekivanja i aproksimacije optimalne tačke. Ograničenja i funkciju cilja smo povezali kaznenom funkcijom, te nam se iterativni postupci zasnivaju na linijskom pretraživanju primenjenom na kvadratnu kaznenu funkciju.

U prvom pristupu rešavali smo SAA reformulaciju originalnog problema. U tom slučaju je uzorak SAA aproksimacije (pun uzorak) zadat pre iterativnog postupka, te obezbeđuje determinističku analizu konvergencije kreiranog algoritma. Analiza algoritma je pokazala da niz veličina uzoraka generisan njime varira po iteracijama, ali dostiže veličinu punog uzorka i nakon konačnog broja iteracija se uzorak ustali na punom uzorku. Pošto se prikazani algoritam zasniva na kaznenom postupku, njime se generiše i niz kaznenih parametara. Niz kaznenih parametara se takođe adaptivno ažurira. U pitanju je neopadajući niz za koji smo pokazali da teži ka beskonačnosti. Pokazane osobine niza veličine uzoraka i niza kaznenih parametara, pod standardnim pretpostavkama, obezbeđuju da se navedenim algoritmom dobija niz iteracija koji sadrži podniz čija je tačka nagomilavanja KKT tačka SAA reformulacije.

Drugi pristup je rešavanje originalnog problema, te je analiza konvergencije stohastička. Predstavljenim algoritmom se generiše niz iteracija koji sadrži podniz čija je tačka nagomilavanja, pod standardnim pretpostavkama za sto-

hastičku optimizaciju, skoro sigurno KKT tačka originalnog problema. Ova osobina važi zahvaljujući tome da se navedenim algoritmom generiše niz veličina uzoraka koji teži ka beskonačnosti i neopadajući niz kaznenih parametara koji skoro sigurno teži ka beskonačnosti.

Sprovedeno numeričko testiranje kreiranih algoritama je pokazalo njihovu prednost u poređenju sa postupcima u kojima je ažuriranje veličine uzorka zasnovano na unapred definisanoj šemi. Prednost se ogleda u rešavanju posmatranih problema s manjim brojem izračunavanja funkcija. Algoritam za rešavanje aproksimacije originalnog problema je upoređen s postupkom u kojem se u svakoj iteraciji računa s punim uzorkom, kao i s algoritmom koji koristi heurističku šemu za povećavanje veličine uzorka do postizanja punog uzorka. Kako ne postoji maksimalna veličina uzorka za algoritam za rešavanje originalnog problema, taj algoritam je upoređen s postupkom zasnovanim na heurističkoj šemi za beskonačno povećavanje veličine uzorka. Dakle, skup testiranih problema nas je doveo do zaključka da je za rešavanje problema stohastičkog programiranja s ograničenjima tipa jednakosti u kojima je evaluacija funkcije skupa isplativo koristiti predložene algoritme s adaptivnim ažuriranjem veličine uzorka zasnovanim na informacijama o greškama nastalim zbog aproksimacije matematičkog očekivanja i aproksimacije optimalne tačke, dobijenim tokom iterativnog postupka.

Glava 8

Dodatak

Razmatrani test problemi:

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, \quad H: \mathbb{R}^n \times \mathbb{R}^p \rightarrow \mathbb{R}^m, \quad \xi \sim \mathcal{N}(1, 1), \quad x = (x_1, x_2, \dots, x_n)^T$$

problem 6

$$\begin{aligned} f(x) &= (1 - x_1)^2 \\ H(x, \xi) &= 10(\xi x_2 - (\xi x_1)^2) \end{aligned}$$

problem 27

$$\begin{aligned} f(x) &= 0.01(x_1 - 1)^2 + (x_2 - x_1^2)^2 \\ H(x, \xi) &= \xi x_1 + (\xi x_3)^2 + 1 \end{aligned}$$

problem 28

$$\begin{aligned} f(x) &= (x_1 + x_2)^2 + (x_2 + x_3)^2 \\ H(x, \xi) &= \xi x_1 + 2\xi x_2 + 3\xi x_3 - 1 \end{aligned}$$

problem 42

$$\begin{aligned} f(x) &= (x_1 - 1)^2 + (x_2 - 2)^2 + (x_3 - 3)^2 + (x_4 - 4)^2 \\ H(x, \xi) &= (\xi x_1 - 2, \\ &\quad (\xi x_3)^2 + (\xi x_4)^2 - 2) \end{aligned}$$

problem 46

$$\begin{aligned} f(x) &= (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6 \\ H(x, \xi) &= (\xi^3 x_1^2 x_4 + \sin(\xi x_4 - \xi x_5) - 1, \\ &\quad \xi x_2 + \xi^6 x_3^4 x_4^2 - 2) \end{aligned}$$

problem 47

$$\begin{aligned} f(x) &= (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^4 + (x_4 - x_5)^4 \\ H(x, \xi) &= (\xi x_1 + (\xi x_2)^2 + (\xi x_3)^3 - 3, \\ &\quad \xi x_2 - (\xi x_3)^2 + \xi x_4 - 1, \\ &\quad \xi^2 x_1 x_5 - 1) \end{aligned}$$

problem 48

$$\begin{aligned} f(x) &= (x_1 - 1)^2 + (x_2 - x_3)^2 + (x_4 - x_5)^2 \\ H(x, \xi) &= (\xi x_1 + \xi x_2 + \xi x_3 + \xi x_4 + \xi x_5 - 5, \\ &\quad \xi x_3 - 2\xi(x_4 + x_5) + 3) \end{aligned}$$

problem 49

$$\begin{aligned} f(x) &= (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6 \\ H(x, \xi) &= (\xi x_1 + \xi x_2 + \xi x_3 + 4\xi x_4 - 7, \\ &\quad \xi x_3 + 5\xi x_5 - 6) \end{aligned}$$

problem 50

$$\begin{aligned} f(x) &= (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^4 + (x_4 - x_5)^4 \\ H(x, \xi) &= (\xi x_1 + 2\xi x_2 + 3\xi x_3 - 6, \\ &\quad \xi x_2 + 2\xi x_3 + 3\xi x_4 - 6, \\ &\quad \xi x_3 + 2\xi x_4 + 3\xi x_5 - 6) \end{aligned}$$

problem 51

$$\begin{aligned} f(x) &= (x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2 \\ H(x, \xi) &= (\xi x_1 + 3\xi x_2 - 4, \\ &\quad \xi x_3 + \xi x_4 - 2\xi x_5, \\ &\quad \xi x_2 - \xi x_5) \end{aligned}$$

problem 52

$$\begin{aligned} f(x) &= (4x_1 - x_2)^2 + (x_2 + x_3 - 2)^2 + (x_4 - 1)^2 + (x_5 - 1)^2 \\ H(x, \xi) &= (\xi x_1 + 3\xi x_2, \\ &\quad \xi x_3 + \xi x_4 - 2\xi x_5, \\ &\quad \xi x_2 - \xi x_5) \end{aligned}$$

problem 61

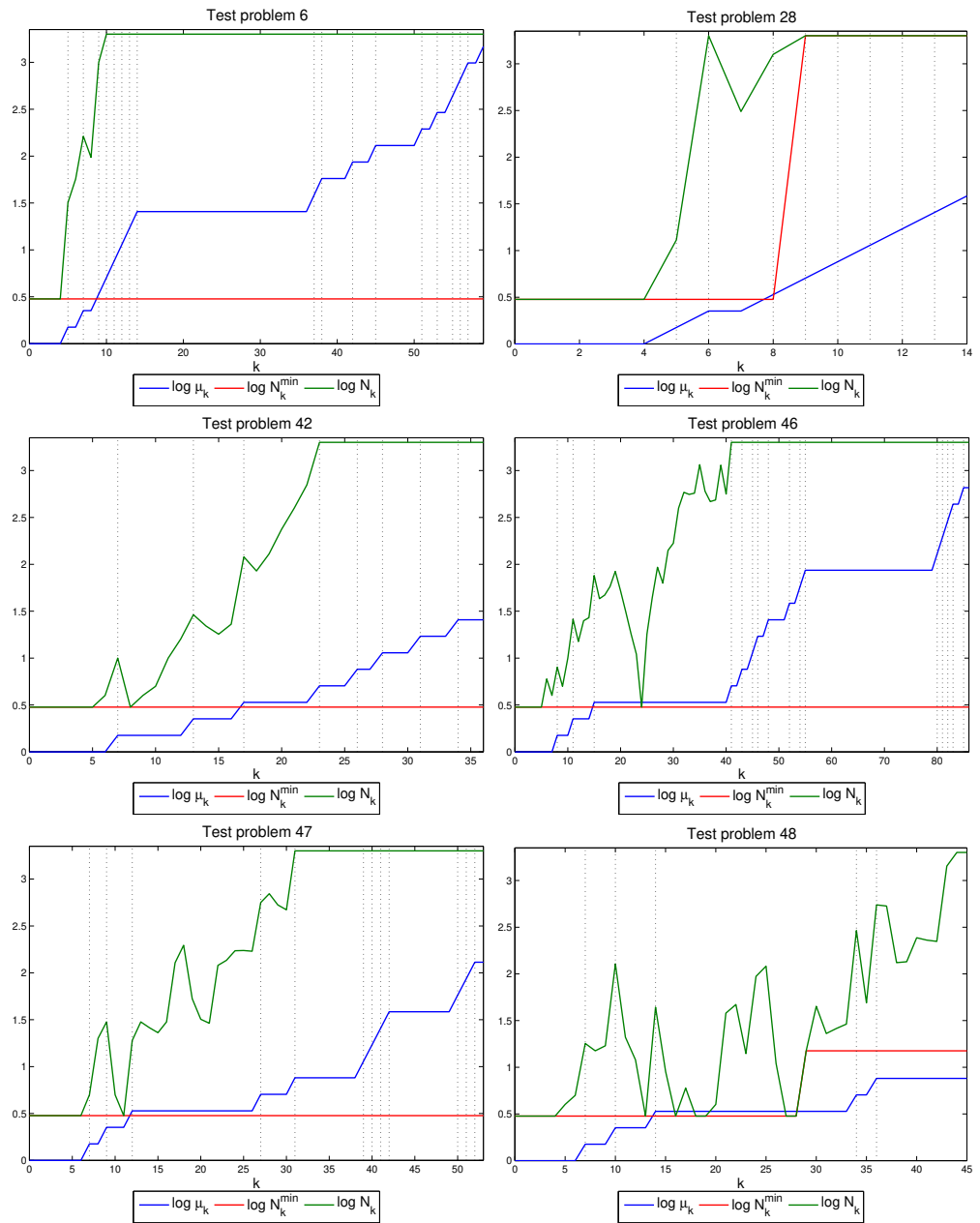
$$\begin{aligned} f(x) &= 4x_1^2 + 2x_2^2 + 2x_3^2 - 33x_1 + 16x_2 - 24x_3 \\ H(x, \xi) &= (3\xi x_1 - 2(\xi x_2)^2 - 7, \\ &\quad 4\xi x_1 - (\xi x_3)^2 - 11) \end{aligned}$$

problem 79

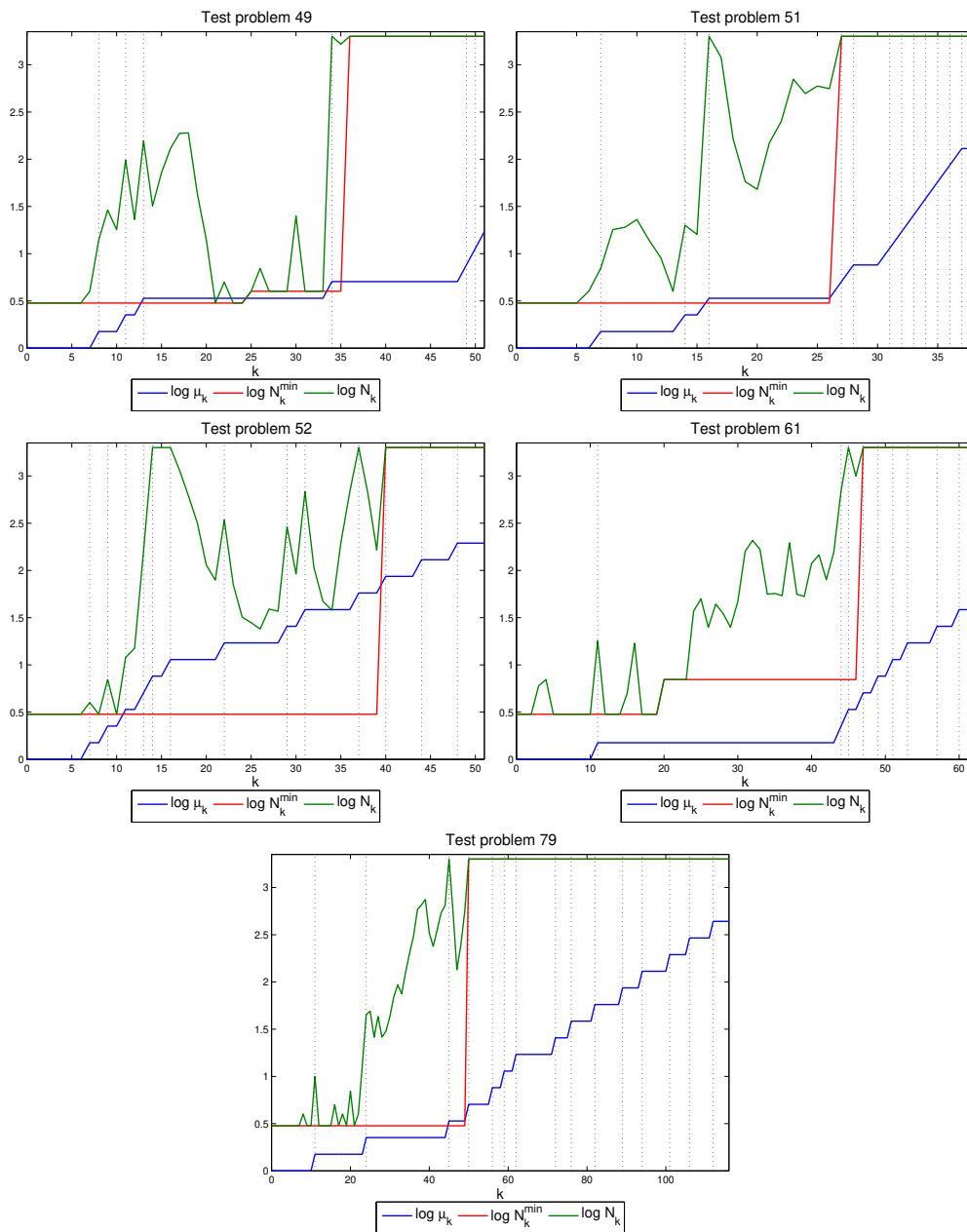
$$\begin{aligned} f(x) &= (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^4 + (x_4 - x_5)^4 \\ H(x, \xi) &= (\xi x_1 + (\xi x_2)^2 + (\xi x_3)^3 - 2 - 3\sqrt{2}, \\ &\quad \xi x_2 - (\xi x_3)^2 + \xi x_4 + 2 - 2\sqrt{2}, \\ &\quad \xi^2 x_1 x_5 - 2) \end{aligned}$$

problem 77

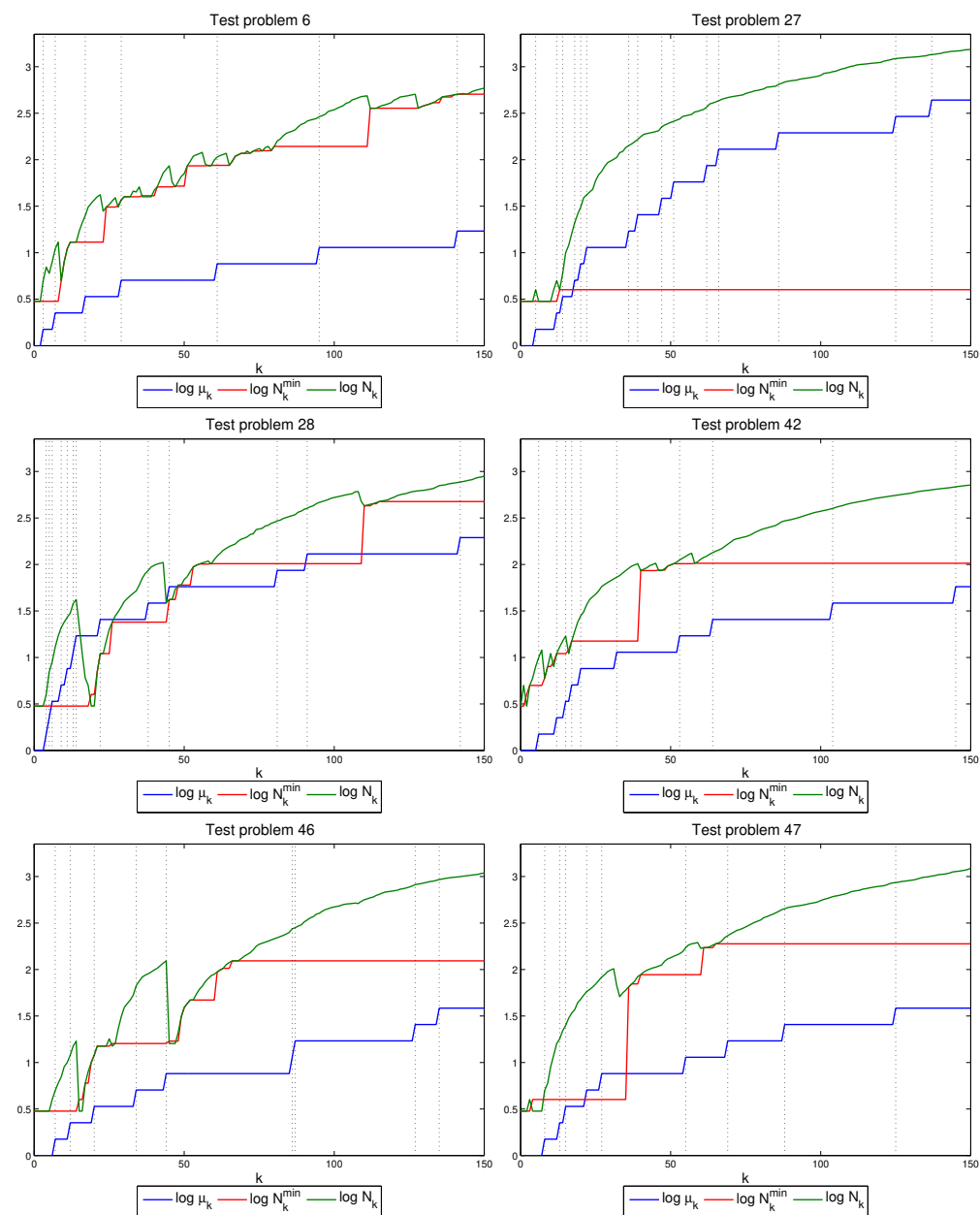
$$\begin{aligned} f(x) &= (x_1 - 1)^2 + (x_1 - x_2)^2 + (x_3 - 1)^2 + (x_4 - 1)^4 + (x_5 - 1)^6 \\ H(x, \xi) &= (\xi^3 x_1^2 x_4 + \sin(\xi x_4 - \xi x_5) - 2\sqrt{2}, \\ &\quad \xi x_2 + \xi^6 x_3^4 x_4^2 - 8 - \sqrt{2}) \end{aligned}$$



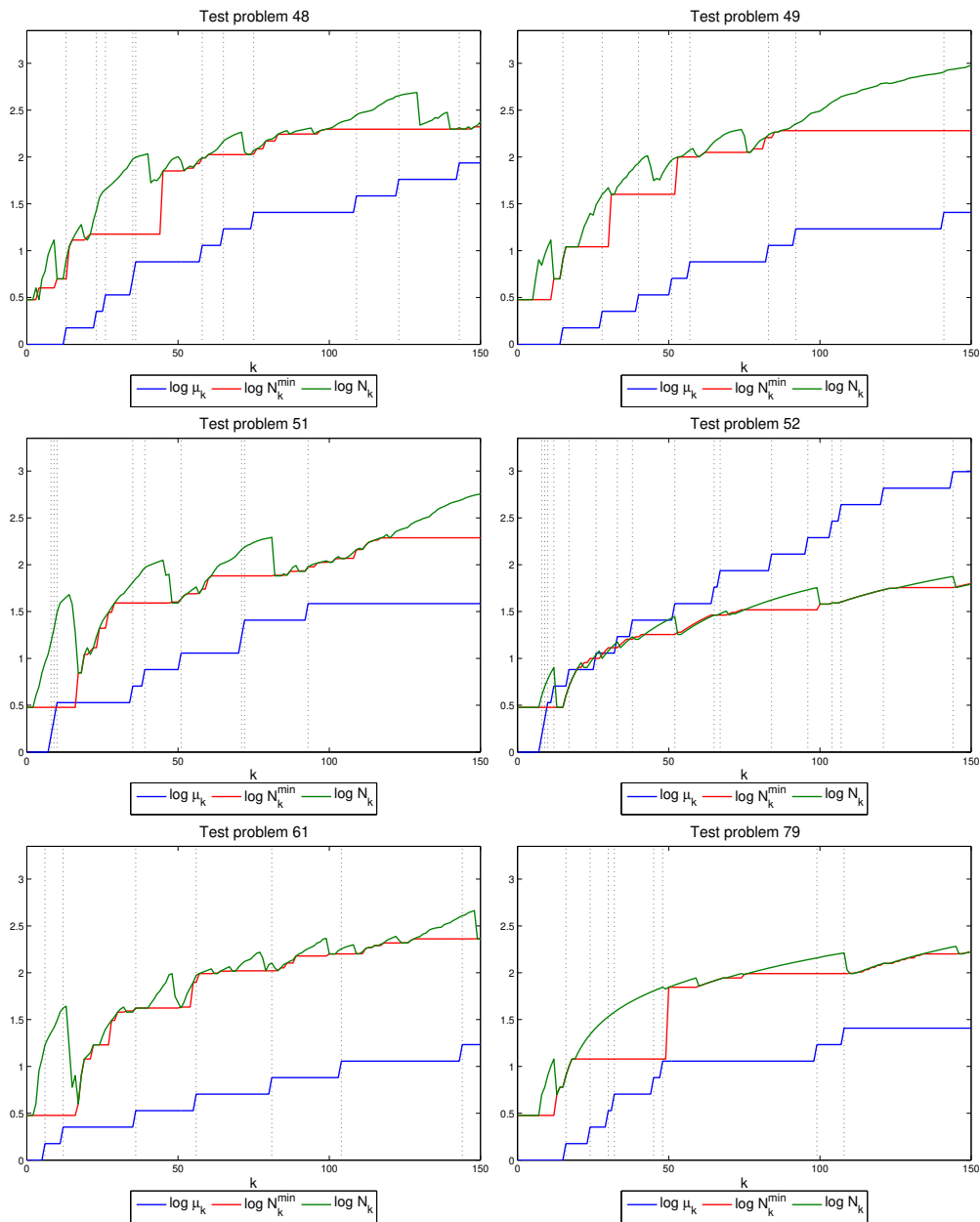
Slika 8.1: Tipično kretanje veličine uzorka N_k , njegove donje granice N_k^{\min} i kaznenog parametra μ_k generisanih algoritmom 4.1



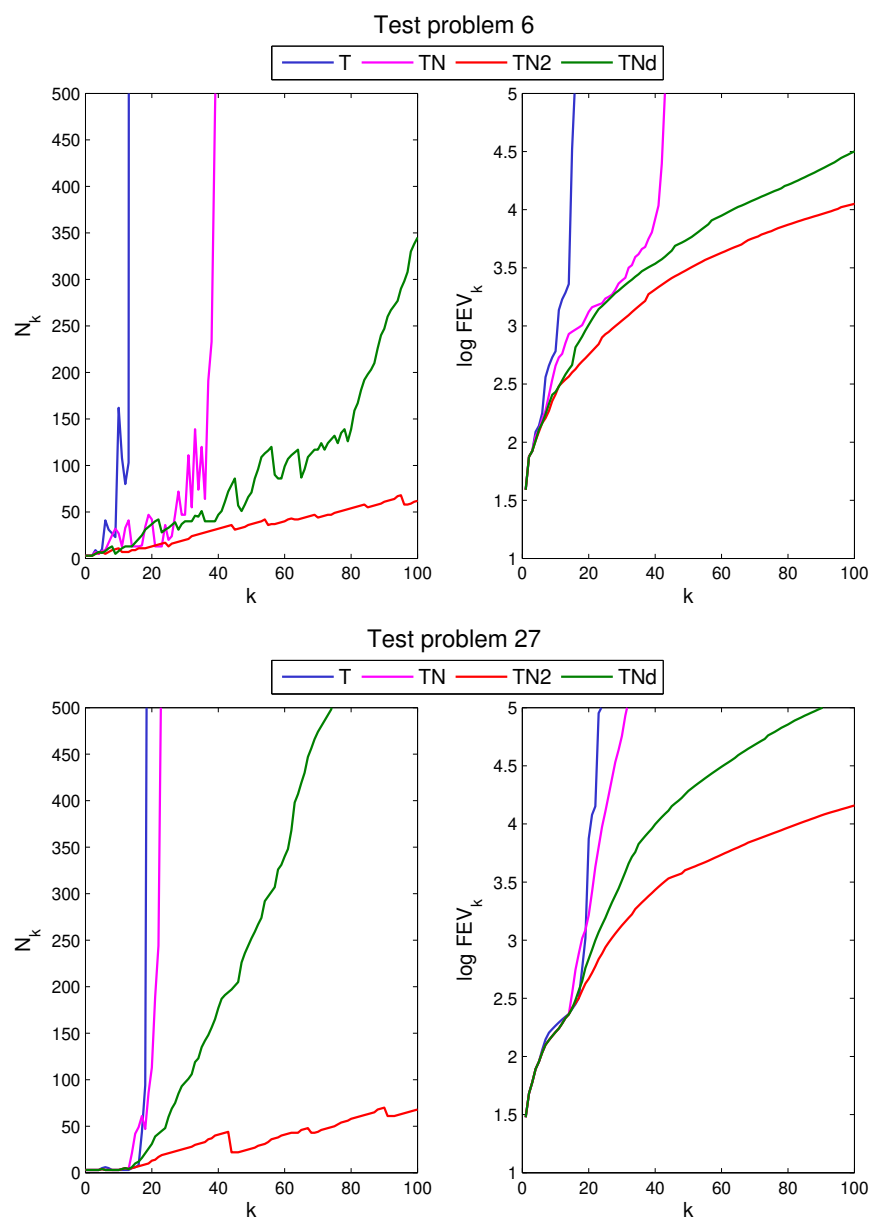
Slika 8.2: Tipično kretanje veličine uzorka N_k , njegove donje granice N_k^{min} i kaznenog parametra μ_k generisanih algoritmom 4.1



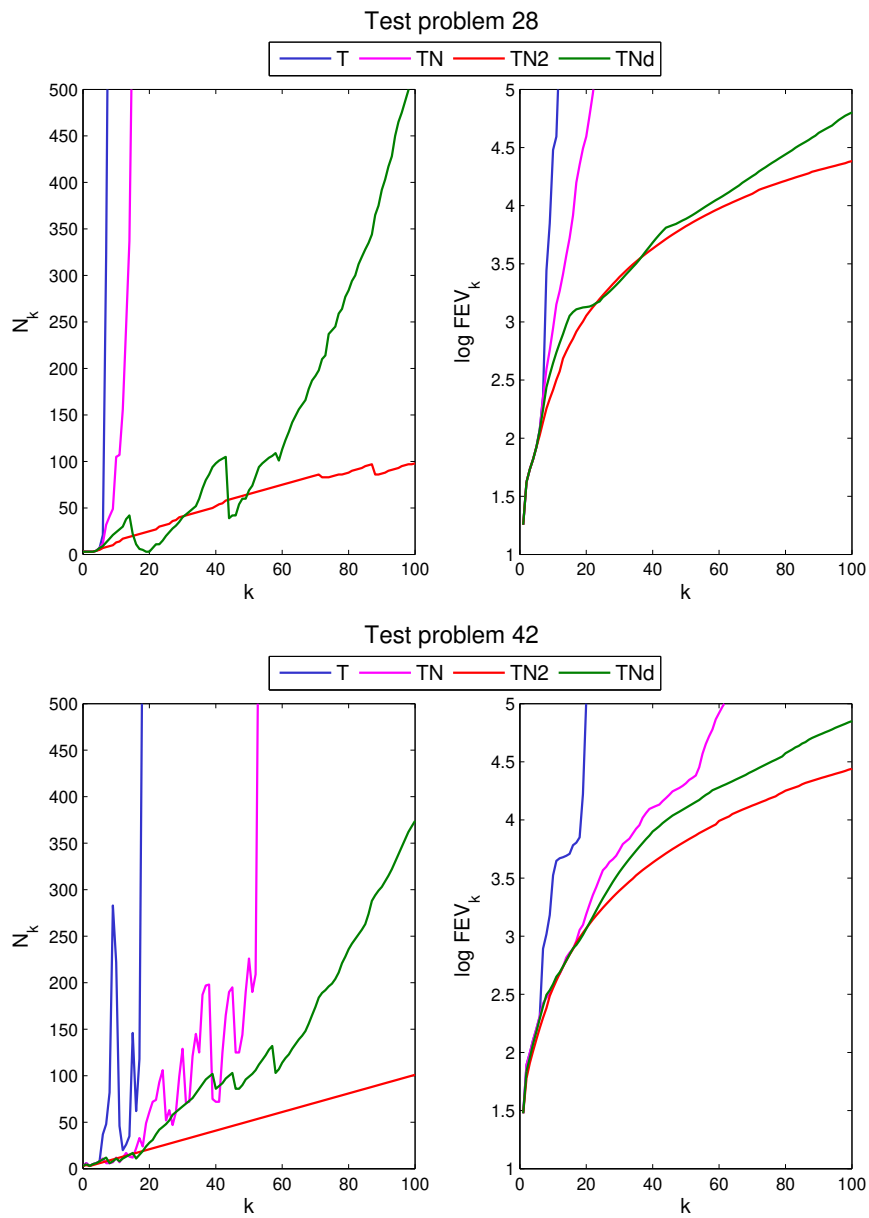
Slika 8.3: Tipično ponašanje veličine uzorka N_k , donje granice veličine uzorka N_k^{\min} i kaznenog parametra μ_k generisanih algoritmom 5.1



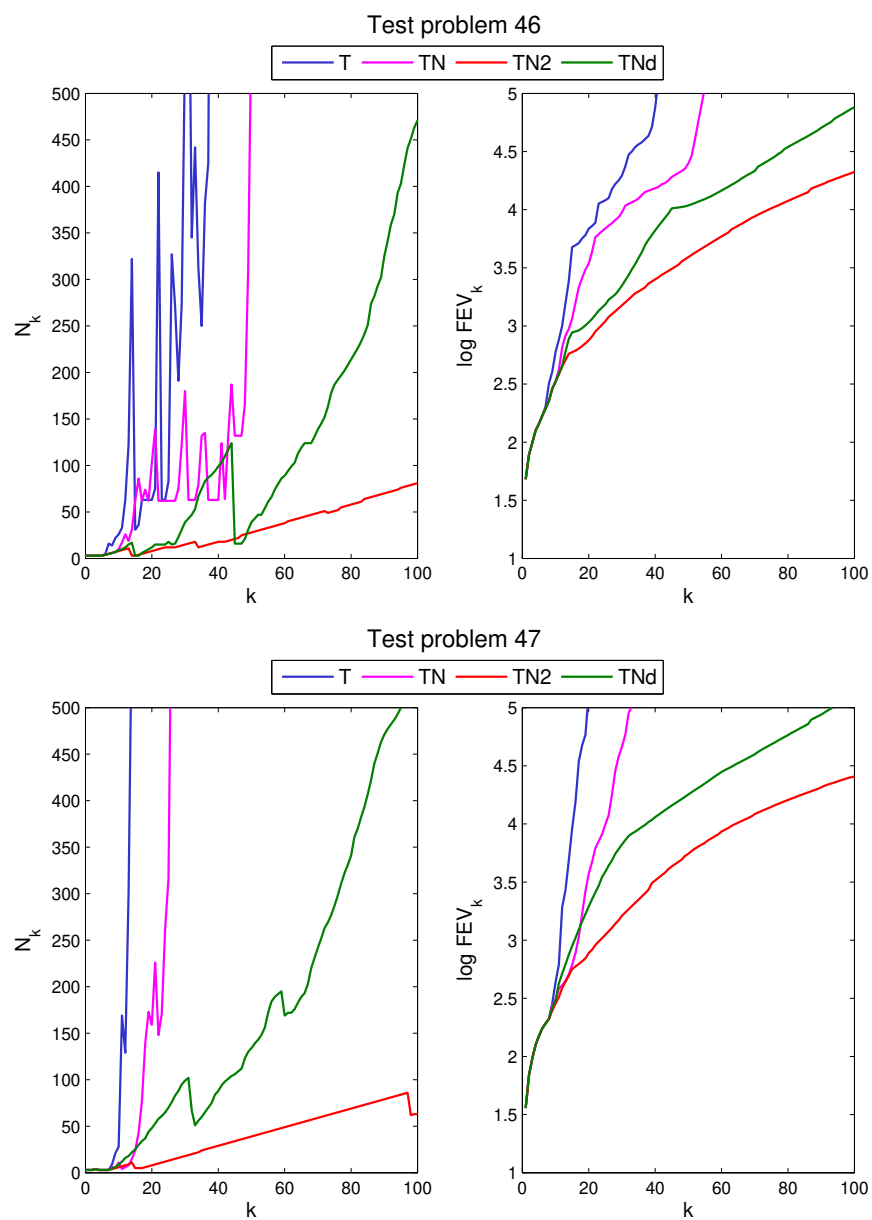
Slika 8.4: Tipično ponašanje veličine uzorka N_k , donje granice veličine uzorka N_k^{\min} i kaznenog parametra μ_k generisanih algoritmom 5.1



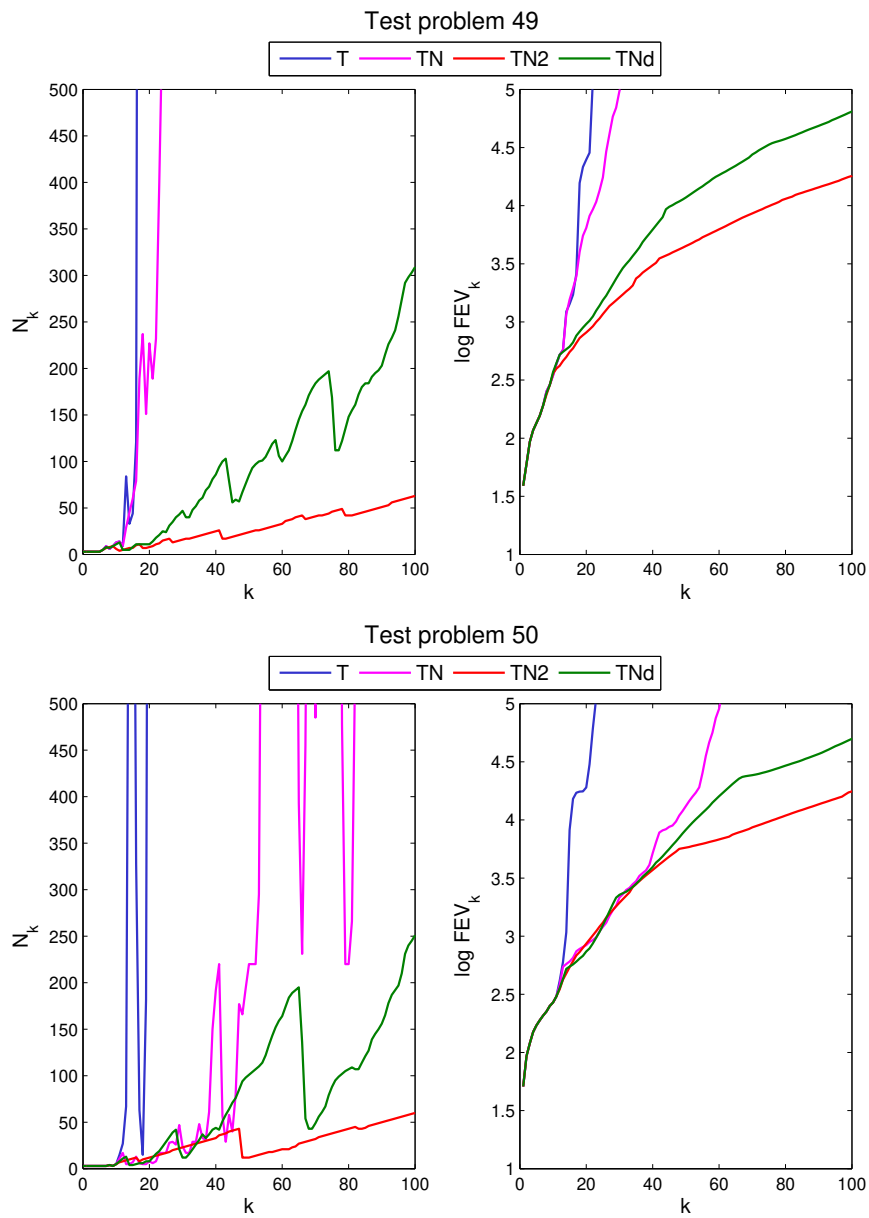
Slika 8.5: Rast veličine uzorka i broja izračunavanja funkcija u zavisnosti od uslova u koraku K1 3) algoritma 5.2



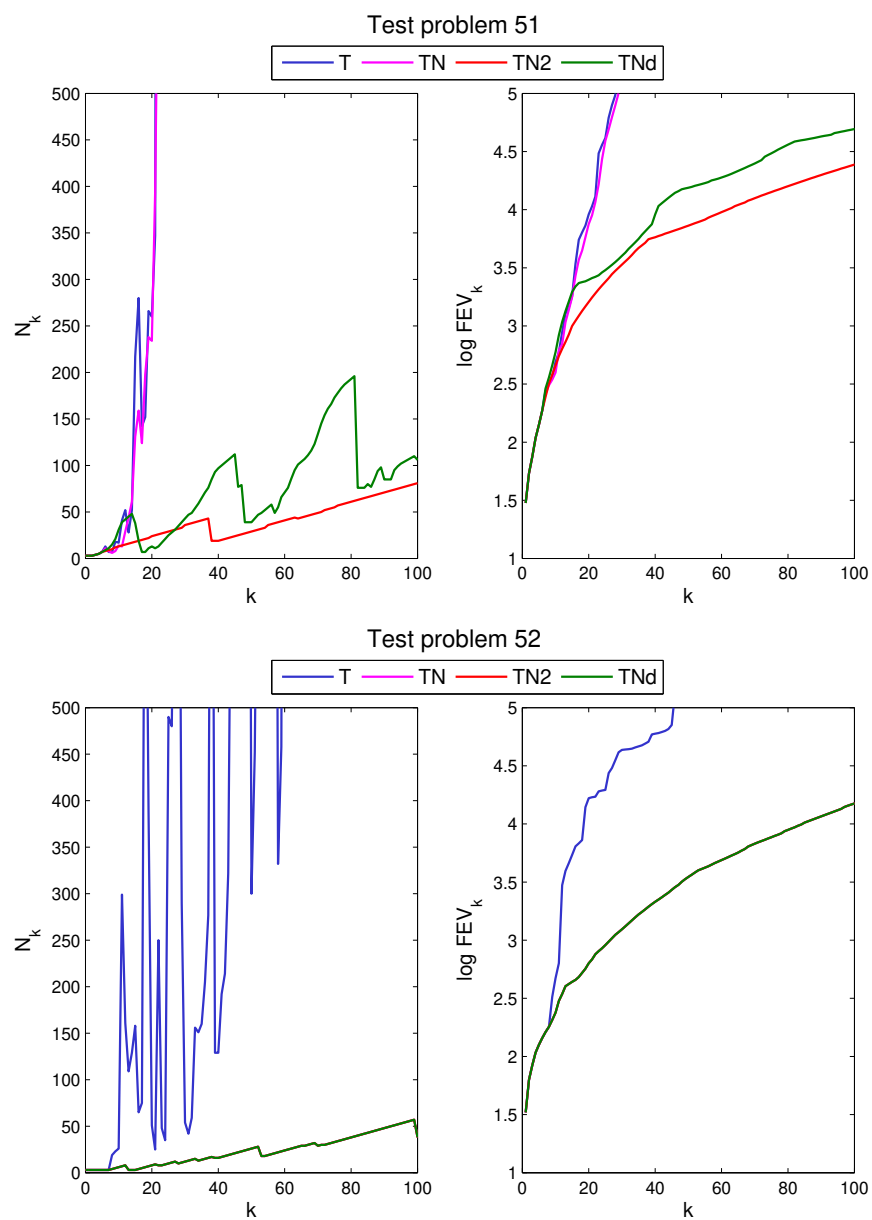
Slika 8.6: Rast veličine uzorka i broja izračunavanja funkcija u zavisnosti od uslova u koraku K1 3) algoritma 5.2



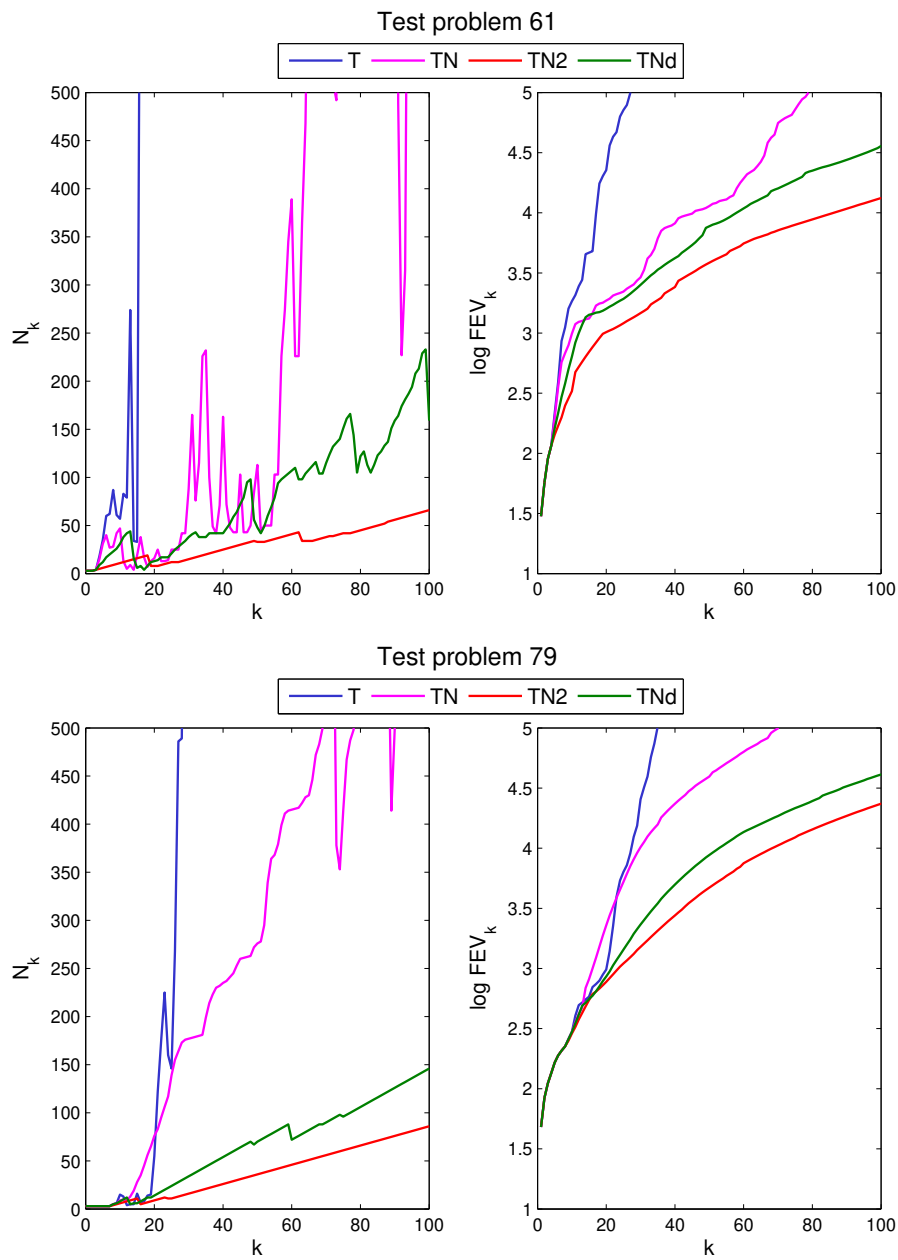
Slika 8.7: Rast veličine uzorka i broja izračunavanja funkcija u zavisnosti od uslova u koraku K1 3) algoritma 5.2



Slika 8.8: Rast veličine uzorka i broja izračunavanja funkcija u zavisnosti od uslova u koraku K1 3) algoritma 5.2



Slika 8.9: Rast veličine uzorka i broja izračunavanja funkcija u zavisnosti od uslova u koraku K1 3) algoritma 5.2



Slika 8.10: Rast veličine uzorka i broja izračunavanja funkcija u zavisnosti od uslova u koraku K1 3) algoritma 5.2

Literatura

- [1] D. Adnađević, Z. Kadelburg: Matematička analiza I, Nauka – Studentski trg, Beograd, 1995.
- [2] D. Adnađević, Z. Kadelburg: Matematička analiza II, Nauka – Studentski trg, Beograd, 1994.
- [3] M. Bartholomew-Biggs, Nonlinear Optimization with Financial Applications, Kluwer Academic Publishers, Boston, 2005.
- [4] M. Bartholomew-Biggs, Nonlinear Optimization with Engineering Applications, Springer, New York, 2008.
- [5] F. Bastin: Trust-Region Algorithms for Nonlinear Stochastic Programming and Mixed Logit Models, PhD thesis, University of Namur, Belgium, 2004.
- [6] F. Bastin, C. Cirillo, P. L. Toint: An adaptive Monte Carlo algorithm for computing mixed logit estimators, Computational Management Science 3(1), 2006, 55-79
- [7] F. Bastin, C. Cirillo, P. L. Toint: Convergence theory for nonconvex stochastic programming with an application to mixed logit, Mathematical Programming 108(2-3), 2006, 207-234
- [8] M. S. Bazaraa, H. D. Sherali, C. M. Shetty: Nonlinear Programming: Theory and Algorithms, Wiley-Interscience, New York, 2006.
- [9] D. P. Bertsekas: Nonlinear Programming, Athena Scientific, Belmont, Massachusetts, 1999.
- [10] J. R. Birge, F. Louveaux: Introduction to Stochastic Programming, Springer, 2011.

-
- [11] L. Dai, C. H. Chen, J. R. Birge: Convergence properties of two stage stochastic programming, *J. Optim. Theory Appl.* 106, 2000, 489-509
- [12] G. Deng, M. C. Ferris: Variable-number sample path optimization, *Mathematical Programming* 117(12), 2009, 81-109
- [13] M. A. Diniz-Ehrhardt, J.M. Martínez, M. Raydan: A derivative-free non-monotone line-search technique for unconstrained optimization, *Journal of Computational and Applied Mathematics* 219(2), 2008, 383-397
- [14] E. D. Dolan, J. J. Mor: Benchmarking optimization software with performance profiles, *Mathematical Programming* 91(2), 2002, 201-213
- [15] Y. Ermoliev, R. J.-B. Wets (eds.): *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, Berlin, 1988.
- [16] R. Fletcher: *Practical Methods of Optimization*, John Wiley & Sons, Chichester, 2000.
- [17] R. Fletcher, S. Leyffer: Nonlinear programming without a penalty function, *Mathematical Programming* 91, 2002, 239-269
- [18] M. P. Friedlander, M. Schmidt: Hybrid deterministic-stochastic methods for data fitting, *SIAM Journal on Scientific and Statistical Computing* 34(3), 2012, 1380-1405
- [19] M. C. Fu, S. D. Hill: Optimization of Discrete Event Systems via Simultaneous Perturbation Stochastic Approximation, *IIE Transactions* 29, 1997, 233-243
- [20] A. Gaivoronski: Implementation of Stochastic Quasigradient Methods, in: Y. Ermoliev, R. J.-B. Wets (eds.): *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, Berlin, 1988, 313-351
- [21] P. E. Gill, W. Murray, M. H. Wright: *Practical optimization*, Academic Press, London, 1997.
- [22] L. Grippo, F. Lampariello, S. Lucidi: A Nonmonotone Line Search Technique for Newton's Method, *SIAM Journal on Numerical Analysis* 23(4), 1986, 707-716
- [23] O. Hadžić, S. Pilipović: *Uvod u funkcionalnu analizu*, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, 1996.

-
- [24] W. Hock, K. Schittkowski: Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems 187, Springer, 1981.
- [25] T. Homem-de-Mello: Variable-Sample Methods for Stochastic Optimization, ACM Transactions on Modeling and Computer Simulation 13(2), 2003, 108-133
- [26] T. Homem-de-Mello: On rates of convergence for stochastic optimization problems under nonindependent and identically distributed sampling, SIAM Journal on Optimization 19(2), 2008, 524-551
- [27] P. Kall, S. W. Wallace: Stochastic Programming, John Wiley & Sons, Chichester, 1994.
- [28] A. Kleywegt, A. Shapiro, T. Homem-de-Mello: The sample average approximation method for stochastic discrete optimization, SIAM Journal on Optimization 12(2), 2001, 479-502
- [29] N. Krejic, N. Krklec: Line search methods with variable sample size for unconstrained optimization, Journal of Computational and Applied Mathematics 245, 2013, 213-231
- [30] N. Krejic, N. Krklec Jerinkic: Nonmonotone line search methods with variable sample size, Numerical Algorithms 68(4), 2015, 711-739
- [31] N. Krejić, N. Krklec Jerinkić: Spectral projected gradient method for stochastic optimization, Journal of Global Optimization (2018). <https://doi.org/10.1007/s10898-018-0682-6>
- [32] N. Krejić, N. Krklec Jerinkić, A. Rožnjik: Variable sample size method for equality constrained optimization problems, Optimization Letters 12(3), 2017, 485-497
- [33] G. Lan, Z. Zhou: Algorithms for stochastic optimization with expectation constraints, arXiv:1604.03887v3 [math.OC] 29 May 2017
- [34] P.-Y. Li, Z.-F. He, G.-H. Lin: Sampling average approximation method for a class of stochastic Nash equilibrium problems, Optimization Methods and Software 28(4), 2013, 785-795

-
- [35] J. Linderoth, A. Shapiro, S. Wright: The empirical behavior of sampling methods for stochastic programming, *Annals of Operations Research* 142, 2006, 215-241
- [36] W.-K. Mak, D. P. Morton, R. K. Wood: Monte Carlo bounding techniques for determining solution quality in stochastic programs, *Operations Research Letters* 24, 1999, 47-56
- [37] B. L. Nelson, J. Swann, D. Goldsman, W. Song: Simple Procedures for Selecting the Best Simulated System When the Number of Alternatives is Large, *Operations Research* 49(6), 2001, 950-963
- [38] A. Nemirovski, A. Juditsky, G. Lan, A. Shapiro: Robust stochastic approximation approach to stochastic programming, *SIAM Journal on Optimization* 19(4), 2009, 1574-1609
- [39] J. Nocedal, S. J. Wright: *Numerical Optimization*, Springer Series in Operations Research, Springer, New York, 2006.
- [40] R. Pasupathy: On choosing parameters in retrospective-approximation algorithms for stochastic root finding and simulation optimization, *Operations Research* 58(4), 2010, 889-901
- [41] S. Pilipović, D. Seleši: *Mera i integral: fundamenti teorije verovatnoće*, Zavod za udžbenike, Beograd, 2012.
- [42] E. L. Plambeck, B. Fu, S. M. Robinson, R. Suri: Sample-path optimization of convex stochastic performance functions, *Mathematical Programming* 75, 1996, 137-176
- [43] E. Polak, J. O. Royset: Efficient sample sizes in stochastic nonlinear programming, *Journal of Computational and Applied Mathematics* 217(2), 2008, 301-310
- [44] D. Rajter-Ćirić: *Verovatnoća*, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, 2013.
- [45] H. Robbins, S. Monro: A Stochastic Approximation Method, *The Annals of Mathematical Statistics* 22(3), 1951, 400-407
- [46] R. T. Rockafellar, R. J.-B. Wets: Scenarios and Policy Aggregation in Optimization under Uncertainty, *Mathematics of Operations Research* 16(1), 1991, 119-147

-
- [47] J. O. Royset: Optimality functions in stochastic programming, *Mathematical Programming* 135, 2012, 293-321
- [48] J. O. Royset, R. Szechtman: Optimal Budget Allocation for Sample Average Approximation, *Operations Research* 61(3), 2013, 777-790
- [49] R. Y. Rubinstein, A. Shapiro: Optimization of static simulation models by the score function method, *Mathematics and Computers in Simulation* 32, 1990, 373-392
- [50] P. Sadegh: Constrained Optimization via Stochastic Approximation with a Simultaneous Perturbation Gradient Approximation, *Automatica* 33(5), 1997, 889-892
- [51] Santoso, T., Ahmed, S., Goetschalckx, M., Shapiro, A.: A stochastic programming approach for supply chain network design under uncertainty, *European Journal of Operational Research* 167, 2005, 96-115
- [52] A. Shapiro: Monte Carlo Sampling Methods, in: *Stochastic programming, Handbook in Operations Research and Management Science* 10, Elsevier, 2003, 353-425
- [53] A. Shapiro, D. Dentcheva, A. Ruszczyński: *Lectures on Stochastic Programming: Modeling and Theory*, MPS-SIAM, Philadelphia, 2009.
- [54] J. C. Spall: *Introduction to Stochastic Search and Optimization: Estimation, Simulation and Control*, Wiley-Interscience, New Jersey, 2003.
- [55] Z. Stojaković, D. Herceg: *Numeričke metode linearne algebre*, Iro *Građevinska knjiga*, Beograd, 1988.
- [56] S. Uryasiev: Adaptive Stochastic Quasigradient Methods, in: Y. Ermoliev, R. J.-B. Wets (eds.): *Numerical Techniques for Stochastic Optimization*, Springer-Verlag, Berlin, 1988, 373-384
- [57] A. Wächter, L. T. Biegler: Line search filter methods for nonlinear programming: Motivation and global convergence, *SIAM Journal on Optimization* 16, 2005, 1-31
- [58] A. Wächter, L. T. Biegler: Line search filter methods for nonlinear programming: Local convergence, *SIAM Journal on Optimization* 16, 2005, 32-48

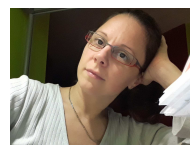
-
- [59] A. Wächter, L. T. Biegler: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Mathematical Programming* 106, 2006, 25-57
- [60] S. W. Wallace, W. T. Ziemba (eds.): *Applications of Stochastic Programming*, MPS-SIAM, Philadelphia, 2005.
- [61] W. Wang, S. Ahmed: Sample average approximation of expected value constrained stochastic programs, *Operations Research Letters* 36(5), 2008, 515-519
- [62] X. Wang, S. Ma, Y. Yuan: Penalty Methods with Stochastic Approximation for Stochastic Nonlinear Programming, *Mathematics of computation* 86(306), 2017, 1793-1820
- [63] S. S. Wilks: *Mathematical Statistics*, John Wiley & Sons, Inc., New York, 1962.
- [64] D. H. Wolpert, W. G. Macready: No Free Lunch Theorems for Optimization, *IEEE Transactions On Evolutionary Computation* 1(1), 1997, 67-82
- [65] H. Xu: Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *Journal of Mathematical Analysis and Applications* 368, 2010, 692-710

Biografija

Rođena sam 16. oktobra 1979. godine u Subotici. U rodnom gradu sam završila osnovnu školu i gimnaziju. Studije sam upisala 1998. godine u Novom Sadu na Prirodno-matematičkom fakultetu na odseku matematika, smeru diplomirani matematičar - matematika finansija. Diplomirala sam 2002. godine s radom na temu „Obveznice - matematički aspekt” i time završila osnovne studije s prosečnom ocenom 9,77. Na nivou Univerziteta u Novom Sadu dobila sam „Mileva Marić - Ajnštajn” nagradu za najbolji diplomski rad i izuzetnu nagradu za izrađen temat sa temom „Približno određivanje ravnotežne cene”.

Magistarske studije, smer primenjena matematika, sam upisala na istom fakultetu. Predviđene ispite sam položila s prosečnom ocenom 9,78, a 2008. godine sam odbranila magistarsku tezu pod naslovom „VaR kao mera rizika u optimizaciji portfolia”. Dobijeni rezultat je objavljen u međunarodnom časopisu. Godine 2012. sam upisala doktorske studije na Prirodno-matematičkom fakultetu u Novom Sadu. Ispite sam položila s prosečnom ocenom 9,42. Deo originalnog naučnog doprinosa iz disertacije je publikovan u međunarodnom časopisu. Koautor sam na još nekim naučnim radovima i učestvovala sam na nekoliko međunarodnih i nacionalnih konferencija.

Krajem 2004. godine zaposlila sam se na Građevinskom fakultetu u Subotici. Kao asistent pripravnik, a potom kao asistent, držala sam vežbe iz matematičkih predmeta. Koautor sam zbirke zadataka za predmet Matematika 1 i rečnika „Mađarsko-srpsko-engleski matematički rečnik”.



Novi Sad, jun 2018.

Andrea Rožnjik

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: Andrea Rožnjik

AU

Mentor: Dr Nataša Krklec Jerinkić

MN

Naslov rada: Optimizacija problema sa stohastičkim ograničenjima tipa jednakosti – kazneni metodi sa promenljivom veličinom uzorka

MR

Jezik publikacije: Srpski (latinica)

JP

Jezik izvoda: s/e

JI

Zemlja publikovanja: Republika Srbija

ZP

Ue geografsko područje: Vojvodina

UGP

Godina: 2018

GO

Izdavač: Autorski reprint

IZ

Mesto i adresa: Novi Sad, Departman za matematiku i informatiku, PMF, Trg
Dositeja Obradovića 4

MA

Fizički opis rada: (8,138,65,0,0,16,0)

(broj poglavlja, strana, lit. citata, tabela, slika, grafika, priloga)

FO

Naučna oblast: Matematika

NO

Naučna disciplina: Numerička matematika

ND

Ključne reči: problem stohastičkog programiranja, ograničenja tipa jednakosti, SAA aproksimacija, metodi s promenljivom veličinom uzorka, kazneni postupci

PO

UDK:

Čuva se: Biblioteka Departmana za matematiku i informatiku, Novi Sad

ČU

Važna napomena:

VN

Izvod: U disertaciji je razmatran problem stohastičkog programiranja s ograničenjima tipa jednakosti, odnosno problem minimizacije s ograničenjima koja su u obliku matematičkog očekivanja. Za rešavanje posmatranog problema kreirana su dva iterativna postupka u kojima se u svakoj iteraciji računa s uzoračkim očekivanjem kao aproksimacijom matematičkog očekivanja. Oba postupka koriste prednosti postupaka s promenljivom veličinom uzorka zasnovanih na adaptivnom ažuriranju veličine uzorka. To znači da se veličina uzorka određuje na osnovu informacija u tekućoj iteraciji. Konkretno, tekuće informacije o preciznosti aproksimacije očekivanja i tačnosti aproksimacije rešenja problema definišu veličinu uzorka za narednu iteraciju. Oba iterativna postupka su zasnovana na linijskom pretraživanju, a kako je u pitanju problem s ograničenjima, i na kvadratnom kaznenom postupku prilagođenom stohastičkom okruženju. Postupci su zasnovani na istim idejama, ali s različitim pristupom.

Po prvom pristupu postupak je kreiran za rešavanje SAA reformulacije problema stohastičkog programiranja, dakle za rešavanje aproksimacije originalnog problema. To znači da je uzorak definisan pre iterativnog postupka, pa je analiza konvergencije algoritma deterministička. Pokazano je da se, pod standardnim pretpostavkama, navedenim algoritmom dobija podniz iteracija čija je tačka nagomilavanja KKT tačka SAA reformulacije.

Po drugom pristupu je formiran algoritam za rešavanje samog problema stohastičkog programiranja, te je analiza konvergencije stohastička. Predstavljenim algoritmom se generiše podniz iteracija čija je tačka nagomilavanja, pod standardnim pretpostavkama za stohastičku optimizaciju, skoro sigurno KKT tačka originalnog problema.

Predloženi algoritmi su implementirani na istim test problemima. Rezultati numeričkog testiranja prikazuju njihovu efikasnost u rešavanju posmatranih

problema u poređenju s postupcima u kojima je ažuriranje veličine uzorka zasnovano na unapred definisanoj šemi. Za meru efikasnosti je upotrebljen broj izračunavanja funkcija. Dakle, na osnovu rezultata dobijenih na skupu testiranih problema može se zaključiti da se adaptivnim ažuriranjem veličine uzorka može uštedeti u broju evaluacija funkcija kada su u pitanju i problemi s ograničenjima.

Kako je posmatrani problem deterministički, a formulisani postupci su stohastički, prva tri poglavlja disertacije sadrže osnovne pojmove determinističke i stohastičke optimizacije, ali i kratak pregled definicija i teorema iz drugih oblasti potrebnih za lakše praćenje analize originalnih rezultata. Nastavak disertacije čini prikaz formiranih algoritama, analiza njihove konvergencije i numerička implementacija.

IZ

Datum prihvatanja teme od strane NN veća: 23. novembar 2017.

DP

Datum odbrane:

DO

Članovi komisije:

KO

Predsednik: dr Nataša Krejić, redovni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sad

Član: dr Sanja Rapajić, vanredni profesor, Prirodno-matematički fakultet, Univerzitet u Novom Sad

Član: dr Zoran Ovcin, docent, Fakultet tehničkih nauka, Univerzitet u Novom Sad

Član: dr Nataša Krklec Jerinkić, docent, Prirodno-matematički fakultet, Univerzitet u Novom Sad

UNIVERSITY OF NOVI SAD
FAKULTY 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 thesis

CC

Author: Andrea Rožnjik

AU

Mentor: Dr. Nataša Krklec Jerinkić

MN

Title: Optimization of problems with stochastic equality constraints – penalty variable sample size methods

TI

Language of text: Serbian

LT

Language of abstract: Serbian/English

LA

Country of publication: Republic of Serbia

CP

Locality of publication: Vojvodina

LP

Publication year: 2018

PY

Publisher: Author's reprint

PU

Publ. place: Novi Sad, Faculty of Sciences, Trg Dositeja Obradovića 4

PP

Physical description: (8/138/65/0/0/16/0)

(chapters/pages/literature/tables/pictures/graphics/appendices)

PD

Scientific field: Mathematics

SF

Scientific discipline: Numerical mathematics

SD

Key words: stochastic programming problem, equality constraints, sample average approximation, variable sample size methods, penalty methods

SKW

UC:

Holding data: Library of the Department of Mathematics and Informatics, Novi Sad

HD

Note:

N

Abstract: Stochastic programming problem with equality constraints is considered within thesis. More precisely, the problem is minimization problem with constraints in the form of mathematical expectation. We proposed two iterative methods for solving considered problem. Both procedures, in each iteration, use a sample average function instead of the mathematical expectation function, and employ the advantages of the variable sample size method based on adaptive updating the sample size. That means, the sample size is determined at every iteration using information from the current iteration. Concretely, the current precision of the approximation of expectation and the quality of the approximation of solution determine the sample size for the next iteration. Both iterative procedures are based on the line search technique as well as on the quadratic penalty method adapted to stochastic environment, since the considered problem has constraints. Procedures relies on same ideas, but the approach is different.

By first approach, the algorithm is created for solving an SAA reformulation of the stochastic programming problem, i.e., for solving the approximation of the original problem. That means the sample size is determined before the iterative procedure, so the convergence analyses is deterministic. We show that, under the standard assumptions, the proposed algorithm generates a subsequence which accumulation point is the KKT point of the SAA problem.

Algorithm formed by the second approach is for solving the stochastic programming problem, and therefore the convergence analyses is stochastic. It generates a subsequence with accumulation point that is almost surely the KKT point of the original problem, under the standard assumptions for stochastic optimization.

The presented algorithms are implemented on the same test problems. Results of numerical testing show their efficiency in solving the considered problems comparing to the procedures which use a priori defined update scheme

for sample size. The number of function evaluations is used as measure of efficiency. Results of the set of tested problems suggest that it is possible to make smaller number of function evaluations by adaptive sample size scheduling in the case of constrained problems, too.

Since the considered problem is deterministic, but the formed procedures are stochastic, the first three chapters of thesis contain basic notations of deterministic and stochastic optimization, as well as a short sight of definitions and theorems from another fields necessary for easier tracking the original results analysis. The rest of thesis consists of the presented algorithms, their convergence analysis and numerical implementation.

AB

Accepted by the Scientific Board on: November 23, 2017

ASB

Defended:

DE

Thesis defend board:

DB

President: Nataša Krejić, PhD, Full Professor, Faculty of Sciences, University of Novi Sad

Member: Sanja Rapajić, PhD, Associate Professor, Faculty of Sciences, University of Novi Sad

Member: Zoran Ovcin, PhD, Assistant Professor, Faculty of Technical Sciences, University of Novi Sad

Member: Nataša Krklec Jerinkić, PhD, Assistant Professor, Faculty of Sciences, University of Novi Sad